

MCP/55

MONASH UNIVERSITY  
THESIS ACCEPTED IN SATISFACTION OF THE  
REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

ON... 3 August 2001

Sec. Ph.D. and Scholarships Committee

Under the copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing for the purposes of research, criticism or review. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

## Amendments

1. Chapter 2, p. 69.

Line 2: replace "and below the range" with "and above the range".

Line 5: replace "+ 1" with "-1"

Line 7: replace "-1" with "+ 1"

2. Chapter 5, Section 5.2.1 Synthesis and Oxidation, p. 155.

A disulphide exchange system was not used for oxidation of acyclic permutant

6. A similar profile of products was obtained when the reaction was repeated in the presence of low MW thiol species (oxidised and reduced glutathione, 0.3 mM and 3 mM, respectively) (N Daly personal communication).

3. Chapter 5, Section 5.2.2 Purification.

P. 155, Para 3: Add to second sentence to read " Components were separated and collected at a flow rate of 1.0 ml/min, using a slow gradient elution over the region of interest, beginning with 10% eluent B and ending with 50% eluent B in sixty minutes".

P. 156, Para 3: Add to second sentence to read " Linear gradient elution, with a flow rate of 1.0 ml/min, beginning with 0% eluent B and ending with 100% eluent B in thirty minutes, was used".

**STRUCTURAL STUDIES AND MODELLING OF NOVEL MACROCYCLIC  
PLANT PEPTIDES WITH MULTIPLE DISULPHIDE BONDS**

by

Anita Claire Koltay

This thesis is submitted in fulfilment of  
the requirements for the degree of  
Doctor of Philosophy

Department of Medicinal Chemistry  
Victorian College of Pharmacy  
MOANSH UNIVERSITY

February 2001

## TABLE OF CONTENTS

<b>Abstract</b>	IX
<b>Declaration</b>	XI
<b>Acknowledgements</b>	XII
<b>Lists of abbreviations and definitions</b>	XIII
1. Abbreviations	XIII
2. Symbols	XIV
3. Nomenclature adopted for description of peptide structures	XV
4. Illustrations of $K_{3,3}$ and $K_5$ subgraphs	XV
5. Three and one letter codes for the common amino acids	XVI
<b>Chapter 1 - Introduction to macrocyclic peptides</b>	1
1.1 Macrocyclic plant polypeptides	2
1.1.1 Kalata	2
1.1.2 Viola peptides and a fractionation protocol for the isolation of plant polypeptides	5
1.1.3 Cyclopsychotride	10
1.1.4 Circulins	11
1.1.5 Synthesis of macrocyclic polypeptides	12
1.1.6 Antimicrobial activity	16
1.1.7 Expansion and comparison of the macrocyclic polypeptide database	17
1.2 Cystine knot motifs	22
1.2.1 Disulphide bonds in nature	22



1.2.2 Structural superfamilies	25
1.2.2.1 Growth factor cystine knot	28
1.2.2.2 Inhibitor cystine knot	30
1.3 Peptide folding	35
1.3.1 Folding of peptides and proteins with disulphide bonds	35
1.3.2 Chemistry of disulphide bond formation	36
1.3.3 The folding pathway	38
1.3.4 Folding of carboxypeptidase potato inhibitor	41
1.3.5 Isomers of $\omega$ -conotoxin GVIA	42
1.3.6 Folding <i>in vivo</i>	43
1.4 Molecular topology in proteins and polypeptides	45
1.4.1 Achiral, planar proteins	46
1.4.2 Simple, chiral, nonplanar proteins	47
1.4.3 Knots and links in proteins	50
1.4.4 Molecular topology in the cystine knot family	51
1.4.5 Kinemages	53
 Chapter 2 - 2D Nuclear magnetic resonance techniques and peptide structure determination	 55
2.1 Introduction	56
2.2 2D NMR spectroscopy	56
2.2.1 The 2D experiment	56
2.2.2 Homonuclear pulse sequences	57
2.2.3 Heteronuclear pulse sequences	61
2.2.4 Solvent suppression	63
2.2.5 Assignment strategy	65
2.3 Secondary structure determination by NMR spectroscopy	68
2.3.1 The chemical shift	68
2.3.2 Coupling constants	71

2.3.3 Exchanging amide protons	72
2.3.4 Temperature coefficients	73
2.3.5 Interresidual NOEs	74
2.4 3D Structure calculations	75
2.4.1 Experimental restraints	75
2.4.2 Dynamical simulated annealing	77
2.4.3 Analysis of structural quality	79
2.5 Peptide purification	82
2.5.1 Separation mechanism	82
2.5.2 Chromatographic conditions	84
2.6 Applications	87
 <b>Chapter 3 - Topological considerations for cyclic knotted peptides</b>	 <b>88</b>
3.1 Introduction	89
3.2 Strategic structural analysis of cyclic cystine knot peptides	89
3.2.1 Linear segments	90
3.2.2 Acyclic permutants	92
3.3 The study of native acyclic permutants	96
3.4 Identification of the natural synthetic pathway	98
3.5 Topological isomerism of kalata-B1	108
3.6 Summary	109
 <b>Chapter 4 - Studies of the solution structure of kalata-B1</b>	 <b>113</b>
4.1 Introduction	114
4.2 Materials and Methods	114
4.2.1 Materials	115
4.2.2 NMR experiments	115
4.2.3 Structural restraints	117
4.2.4 Structure calculations	117

4.3 Results and discussion	118
4.3.1 Proton resonances	118
4.3.1.1 CSI values from alpha proton chemical shifts	119
4.3.1.2 Effect of TFE concentration on amide proton chemical shifts	121
4.3.1.3 Effect of temperature on amide proton chemical shifts	125
4.3.2 $^{13}\text{C}$ Chemical shifts and prediction of secondary structure	125
4.3.3 Determination of <i>cis/trans</i> proline isomerisation	136
4.3.4 Three dimensional structure of kalata-B1 in the presence of TFE	140
4.3.4.1 Quality of structures	140
4.3.4.2 Comparison of the structures of kalata-B1 in different solvents	140
4.3.4.2.1 Overall 3D fold	142
4.3.4.2.2 Residual deviations	142
4.3.4.2.3 Elements of secondary structure	144
4.3.4.2.4 Distortion in space	146
4.3.4.2.5 Hydrophobic surface area	147
4.4 Summary	149
 Chapter 5 - Studies of acyclic permutant 6	 152
5.1 Introduction	153
5.2 Materials and Methods	153
5.2.1 Synthesis and oxidation	155
5.2.2 Purification	155
5.2.3 Purity	156
5.2.4 NMR Spectroscopy	157
5.2.4.1 Sample preparation	157

5.2.4.2 NMR Experiments	157
5.3 Results and discussion	159
5.3.1 Separation and purity	159
5.3.1.1 Oxidation	159
5.3.1.2 Separation	159
5.3.1.3 Purity	161
5.3.2 Assignment of proton chemical shifts and detection of isomers	164
5.3.3 Proton chemical shifts and structural similarity	168
5.3.3.1 Information from alpha protons	168
5.3.3.2 Information from amide protons	173
5.3.4 Secondary structure identification from CSI values	175
5.3.4.1 CSI Values of isomers and comparison of their structures	176
5.3.4.2 Comparison of kalata-B1 with fragment isomer structures	178
5.3.5 $^3J_{\text{NH}-\alpha\text{H}}$ Coupling constants	179
5.3.6 Nuclear Overhauser enhancements	181
5.3.6.1 Short and medium range NOE connectivities	181
5.3.6.2 Medium and long range NOE connectivities	183
5.3.7 <i>Cis/trans</i> proline isomerism	187
5.3.7.1 P14	187
5.3.7.2 P21	188
5.4 Summary	189
 Chapter 6 - Structure determination of circulin A	 193
6.1 Introduction	194
6.2 Materials and Methods	194
6.2.1 NMR experiments	195
6.2.2 Structure calculations	196

6.3 Results and discussion	198
6.3.1 Proton resonance assignments	198
6.3.2 Secondary structure	200
6.3.3 Three dimensional structure	206
6.3.3.1 Quality of structures	206
6.3.3.2 Three dimensional arrangement	208
6.3.3.3 Analysis of structure	210
6.3.3.4 Charge and hydrophobic distribution	213
6.4 Comparison of circulin A and kalata-B1	215
6.5 Discussion of biological activity	217
6.6 Summary	219
 <b>Chapter 7 - Structure determination of circulin B</b>	 221
7.1 Introduction	222
7.2 Materials and Methods	223
7.2.1 NMR experiments	223
7.2.2 Structure calculations	225
7.3 Results and discussion	227
7.3.1 Proton resonance assignments	227
7.3.2 Secondary structure	227
7.3.3 Three dimensional structure	235
7.3.3.1 Quality of structures	235
7.3.3.2 Three dimensional arrangement	237
7.4 Comparison of circulin topology	241
7.4.1 Conserved features	241
7.4.2 Topological differences in relation to activity	245
7.5 Summary	252
 <b>Chapter 8 - Conclusions</b>	 255

<b>Publications and Communications</b>	<b>261</b>
<b>Appendices</b>	<b>263</b>
Appendix I	264
Appendix II	268
Appendix III	270
Appendix IV	277
Appendix V	284
Appendix VI	285
Appendix VII	296
<b>References</b>	<b>297</b>

## ABSTRACT

Structural studies on an emerging family of macrocyclic polypeptides extracted from plants have been conducted. These polypeptides are the largest naturally occurring cyclic polypeptides. They are comprised of around thirty amino acids and possess three disulphide bonds connected to form the inhibitor cystine knot structure. These peptides exhibit a range of diverse activities, eliciting considerable pharmacological interest.

A strategy for the study of structure/activity relationships of the first discovered macrocycle, kalata-B1, was proposed. A set of fragments of kalata-B1, named acyclic permutants, for which selected sections of residues were separately omitted, was designed. These fragments are not cyclic, however, the three disulphide bonds are retained. The acyclic permutants may be then individually analysed for structure and activity. The proposition is a novel approach in the study of structure/activity relationships.

In an attempt to describe folding of macrocyclic peptides and their acyclic permutants, 'kinemage' computer graphics was used. Animations were created for acyclic permutants in which the backbones were unthreaded to produce 2D planar structures. These were compared to the 2D planar structure of a non-cyclic naturally occurring inhibitor cystine knot peptide. The comparison gave a topological basis to suggest that the starting residue for natural biosynthesis is likely to occur in the disordered region of kalata-B1. Furthermore, the animations highlight the stabilising effect of the cystine knot on the structure.

Another animated kinemage was created to illustrate the topological isomerism of kalata-B1. In the animation, the 3D structure was transformed to the 2D reduced graph. This reduced graph shows that kalata-B1 is not topologically simple, nonplanar and has D chirality.

Using homonuclear and heteronuclear 2D NMR spectroscopy and simulated annealing calculations, the structure of kalata-B1 was investigated in trifluoroethanol (TFE). Changes in structure are subtle in the presence of

TFE, indicating the stable nature of cyclic cystine knots. There is net increase in the hydrophobic surface area, concomitant with loss of secondary structure definition and development of a tighter sulphur core. Full assignments of  $^{13}\text{C}$  chemical shifts were obtained in TFE.  $^{13}\text{C}$  data were not useful in predicting location of secondary structure and changes in TFE.  $^{13}\text{C}$  data were reliable in determination of *cis/trans* proline conformation and the unusual *cis* arrangement for one of three prolines is found.

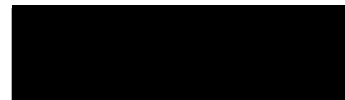
Structural studies using 2D NMR spectroscopy were applied to a key acyclic permutant. The linear precursor was oxidised and a mixture of species was obtained. The main fraction was isolated and found to consist of two distinct isomers, resulting from *cis/trans* isomerisation of a proline. The NMR data indicate that the isomers are similar in structure, but differ in conformation from the parent. The disulphide connectivity was postulated. It was deduced that the native disulphide, Cys(2)-Cys(5), is maintained, however, the other two disulphides are interchanged.

The solution structures of two new macrocyclic peptides with anti-HIV activity, circulin A and B were determined through 2D NMR and simulated annealing calculations. The fold, shape and distribution of hydrophobic residues are similar to that of kalata-B1. However, there are some differences. The structures and surfaces of the circulins were compared in detail. Although the circulins are closely related, their activities are remarkably different. Several structural differences were observed between circulin A and circulin B, which may explain the activity discontinuities. However, the cystine knot core is maintained. The findings illustrate the strategic importance of the cyclic cystine knot scaffold in retaining a stable structure that with minor residual modifications result in substantial and specific activity changes. The work for this thesis has contributed to and increased knowledge of the structures of cyclic cystine knot peptides



## DECLARATION

I, Anita Claire Koltay declare that the material contained in this thesis has not been published elsewhere, except where due reference has been made in the text, that no other persons work has been used without due acknowledgement, and that this thesis is not and has not been used for the award of any other degree or diploma in any university.

A solid black rectangular box used to redact the signature of Anita Claire Koltay.

Anita Claire Koltay

February 2001

## ACKNOWLEDGEMENTS

I would especially like to thank Professor David Craik for providing this tremendous opportunity and for his unending patience and support in the prevailing difficult circumstances throughout my Ph.D. studies. I am deeply grateful to Dr. Jenny Wilson for her heartfelt care and assistance to ensure that I succeed and also to Dr. Peter Baron for his kind provision of software so that I was able to continue working from home in my time of need.

To my many valued colleagues; Murray, Wendy, Lidia, Norelle, Paul, Jackie, Sharon, Michael, Jeff, David, Victoria, Brendan, Neville, Kathy, I will always remember with fondness your friendship, encouragement and assistance.

I would also like to thank Dr. Diane Alewood (Centre for Drug Design and Development, QLD, Australia) for synthesising kalata fragments, Professor Ray Norton (CSIRO, Biomolecular Research Institute, Vic, Australia) for access to NMR facilities, Dr. Norelle Daly (Centre for Drug Design and Development, QLD, Australia) for the recording of NMR data at NMR facilities in QLD and Kirk Gustafson (National Cancer Institute, Maryland, U.S.A.) for the extraction and purification of circulins.

Finally, a very special thank-you to my mother for her love, support and encouragement and to my family and friends for their continued love, understanding and friendship.

## LIST OF ABBREVIATIONS AND DEFINITIONS

### 1. Abbreviations

1D, 2D, 3D	one dimensional, two dimensional, three dimensional
CAN	acetonitrile
AOase	ascorbate oxide enzyme
BPTI	bovine pancreatic trypsin inhibitor
CD	circular dichroism
CCK	cyclic cystine knot
COSY	correlated spectroscopy
CMTI-I	<i>Curcubita maxima</i> trypsin inhibitor
CPI	carboxypeptidase potato inhibitor (same as PCI)
CSI	chemical shift index
DDS	2,2-dimethyl-2-silapentane-5-sulfonate
DQF	double quantum filtered
E.COSY	exclusive correlation spectroscopy
EETI-II	<i>Ecballium elaterium</i> trypsin inhibitor
FAB-MS	fast atom bombardment mass spectrometry
FID	free induction decay
FT	Fourier transformation
hCG	human chorionic gonadotrophin
HiPIP	high potential iron protein
HIV	human immunodeficiency virus
hLF	human lactoferrin
HMQC	heteronuclear multiple quantum coherence
HPLC	high performance liquid chromatography
LB	line broadening
MADH	methyl-amine dehydrogenase
MD	molecular dynamics
NDP	Norrie disease protein

NGF	nerve growth factor
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser and exchange spectroscopy
PCI	potato carboxypeptidase inhibitor (same as CPI)
PDGF-BB	platelet-derived growth factor
ppm	parts per million
RMSD	root mean square deviation
RNase A	bovine pancreatic ribonuclease A
RP	reverse phase
SA	simulated annealing
TEA	triethylamine
TEAP	triethylammonium phosphate
TFA	trifluoroacetic acid
TFE	trifluoroethanol
TGF $\beta$ 2	transforming growth factor
TOCSY	total correlation spectroscopy
TPPI	time proportional phase incrementation
TV	<i>Thiobacillus versutus</i>
UV	ultra-violet
X-PLOR	molecular dynamics structure refinement software program

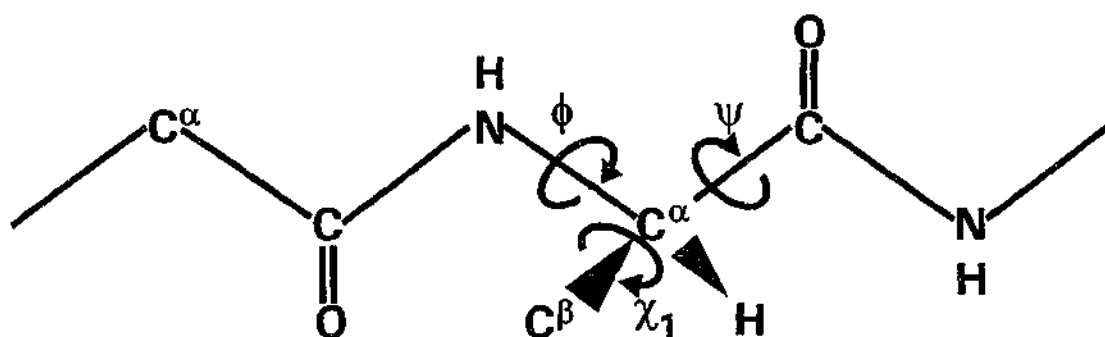
## 2. Symbols

Å	Angstrom
$\delta$	chemical shift
°C	degrees Celcius
$^{13}\text{C}$	carbon-13
$^1\text{H}$	proton
J	scalar coupling constant
K	degrees Kelvin

### 3. Nomenclature adopted for the description of protein and peptide structures

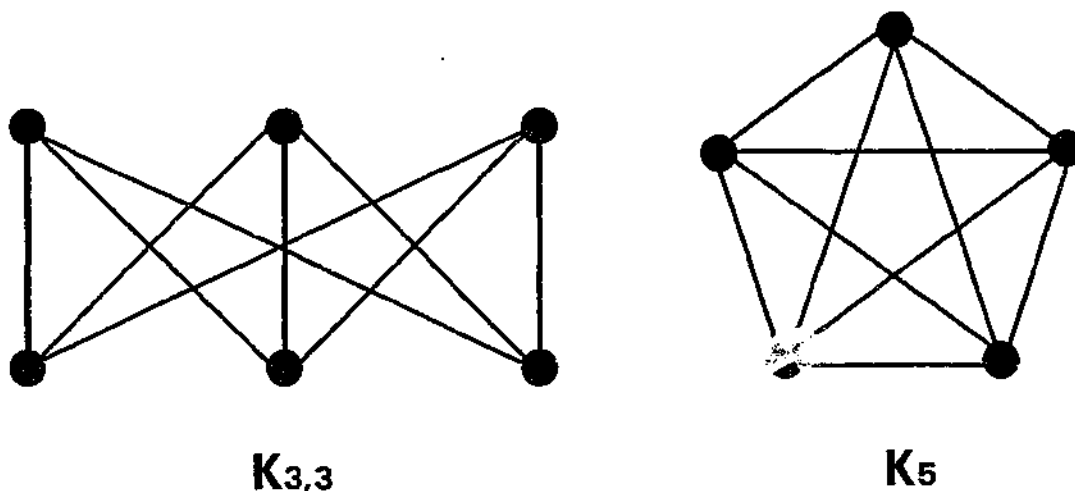
The conventions used to describe the structural characteristics and have been adopted in this thesis are as follows:

#### Definition of bond angles in the peptide or protein structure



### 4. Illustrations of $K_{3,3}$ and $K_5$ subgraphs

The standard presentations of  $K_{3,3}$  and  $K_5$  graphs (Walba, 1985) are shown below. Dots represent vertices and lines, the edges.



5. Three letter and one letter code for the common amino acids

Amino Acid Residue	Three Letter Code	One Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

**CHAPTER ONE**

**INTRODUCTION TO MACROCYCLIC**

**PEPTIDES**

## 1.1 Macrocyclic Plant Polypeptides

The work for this thesis is concerned with a novel and recently discovered class of macrocyclic polypeptides. These peptides are derived from plants and are the largest known, naturally occurring, cyclic peptides, where the peptide backbone chain is cyclised through amide bonds. In addition to being cyclic, they are characterised by exceptional stability and a well defined structure, consist of around 30 amino acids and possess three disulphide bonds. There is much to learn regarding these peptides, which originate from diverse global locations. The peptides share a significant level of sequence homology, with around 35% in common residues.

As yet, their function in plants is not understood. However, a variety of activities of pharmacological significance have been associated with these macrocycles. These activities, together with their unusual structural and physical properties, render the study of these macrocycles of interest with regard to potential for drug design, as well as adding to our knowledge of biochemistry. Certainly, the determination and study of the 3D topology are a crucial step in the overall understanding of activity. Using 2D-NMR spectroscopic techniques in combination with molecular simulations, the work for this thesis aims to elucidate the 3D structures of the members of this intriguing family. This first section of the thesis commences with a presentation on the known information on these macrocycles.

### 1.1.1 Kalata

Kalata-B1 is the most well studied of the macrocycles and is an important part of structural studies in this thesis. Kalata first appeared in the literature in 1970 (Gran, 1970). Kalata is derived from a species of Rubiaceae called *Oldenlandia affinis* DC. The plant grows wild all over tropical Africa, from the Ivory Coast to Transvaal of South Africa, and is also found in south India and Madagascar. The interest in the plant was instigated by its widespread use by a local tribe of the Kasai Province of Zaire to accelerate



the child birthing process, in an effort to make delivery as smooth as possible. Kalata is also used for the same purpose by a tribe in Central Africa Republic (Gran, 1972) and by the Zulus against shortness of breath in asthma and in heart disease.

The plant is collected in the rainy season and dried. The dried plant is boiled for an hour, filtered and the resulting green fluids used as the birthing potion. A mouthful is taken by the expectant mother as birthing contractions commence and is then slowly sipped throughout the birth. Consumption of the potion seems to give good uterine contractions and a considerably shorter delivery time. Unfortunately, there is a high rate of complications in the form of cervical spasms, which necessitate extraction or caesarean section. This led to a detailed study of the plant to isolate the active component(s) and to characterise the activity (Gran, 1970; Gran, 1972; Sletten and Gran, 1973; Gran a-d, 1973).

The active components were found to be remarkably stable. They have excellent thermoresistance, remaining active even after exposure to 200°C and are highly resistive to enzymatic degradation. They are soluble in water and lower alcohols, but not in non-polar solvents. Two active components were isolated with one component, kalata-B1, being the main component and the other component, being insufficient to conduct many tests. Kalata-B1 was found to consist of 29 amino acids and to be a cysteine rich peptide, with three disulphide bridges. Oxidation with performic acid destroyed *in vitro* activity. The extracts of kalata are active on rat, rabbit and human uterii in isolation, giving slow, strong, progressive contractions and increased muscle tone. However, the same dose required to induce contractions *in vitro*, results in increased blood flow, ventricular fibrillation and finally death in rats and respiratory arrest in rabbits.

Detailed studies of the primary and topological structure of kalata-B1 were reported in 1995 (Saether *et al.*, 1995). The peptide was confirmed to be cyclic and consist of 29 amino acids, rich in cysteine (6), threonine (5) and glycine (5). The sequence is shown in Figure 1.1.1.

SWPVCTRNGLPVCGETCVGGTCNTPGCTC

Figure 1.1.1. Sequence of kalata-B1 (Saether *et al.*, 1995).

The 3D structure of the kalata-B1 was determined by distance restrained simulated annealing calculations, which used distance restraints derived from two dimensional nuclear magnetic resonance spectroscopy. Two approaches were applied in order to determine the connectivity of the three disulphides. The first of these involved the calculation of a family of structures based on NOE constraints only with omission of any disulphide bonds. The family was then closely viewed to observe the positions of the sulphur atoms with the expectation that these would infer the disulphide connections. The second approach calculated fifteen families, where each family was based on one of the fifteen possible disulphide connectivity patterns. The families were then analysed to determine if a particular family indicated the most likely connection. Both approaches resulted in the same global fold for kalata-B1 as well as suggesting the mostly likely disulphide connection pattern.

The mostly likely connections seem to be Cys5-Cys22, Cys13-Cys27, and Cys17-Cys29 (numbering according to Saether *et al.*, 1995). The fold may be described as a cyclic peptide backbone folded back onto itself to form two  $\beta$ -sheets on top of each other, rather like a fixed open pocket. The  $\beta$ -sheets are held in place by the disulphide bonds diagonally across opposed  $\beta$ -strands. The Cys5-Cys22 disulphide threads through the eight amino acid loop formed by the other two disulphide bonds to form a knot. Thus kalata-B1 is comprised of  $\beta$ -strands, connected by tight turns. One section however, diverges from the rest of the fold giving a longer and less structured loop. It is thought that this section might be more flexible and therefore, may be effective for adaptation of kalata-B1 to a binding site. The sequence of kalata-B1 in combination with the fold creates an exposed surface, along which most of the hydrophobic residues are concentrated. The two ends of the surface have two diametrically opposed and opposite charged residues. It

is speculated that the charged residues could be involved in initial recognition prior to binding of the hydrophobic surface to a receptor site. The sulphurs of the cysteines are located in the tight central core, acting like glue, so that kalata-B1 is a highly compact structure. A diagram illustrating the topology of kalata-B1 is shown in Figure 1.1.2.

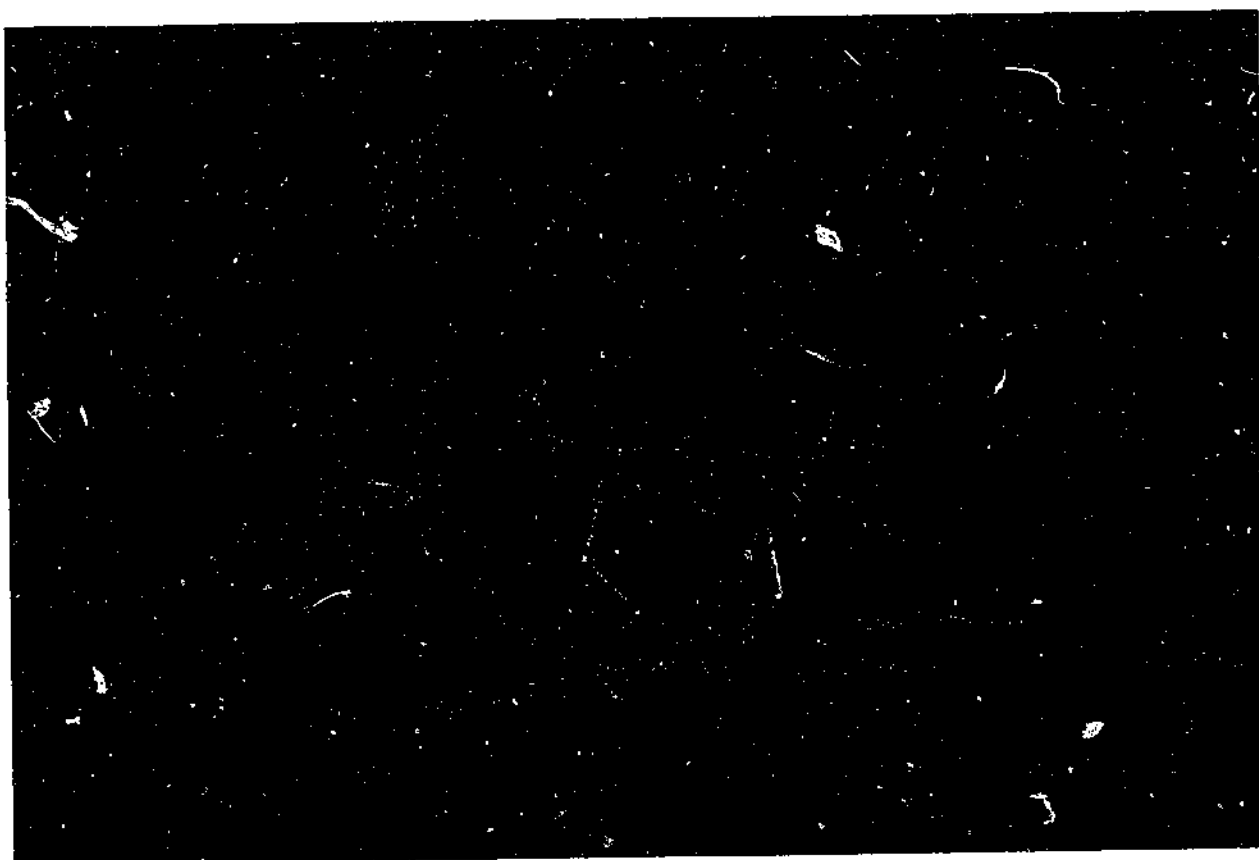
Although the results of the pharmacological studies are a little discouraging in terms of possible clinical application, the study of kalata-B1 could be valuable as a model compound for the design of drugs which act selectively on muscle. Kalata-B1 is definitely interesting in terms of its high stability, unusual well defined structure and uterotonic activity. Additionally, kalata-B1 may well be a precursor to the understanding of other macrocyclic peptides of potentially useful activity. Consequently, an understanding of activity as it relates to structure is of importance. The work for this thesis is directed towards this understanding.

#### 1.1.2 Viola Peptides and a Fractionation Protocol for the Isolation of Plant Polypeptides

Discovery of a peptide, viola peptide-I, almost identical in sequence to kalata-B1 was published in 1993 (Schöpke *et al.*, 1993). The peptide was accidentally found in a search for saponins with hemolytic activity in pansy type plants. Saponins were not present, however, a peptide rich in cysteine, threonine and glycine was extracted and isolated from the dry roots and leaves of a wild central European field pansy, *Viola arvensis* Murray. Viola peptide-I, was determined to have the sequence shown in Figure 1.1.3.

ETCVGGTCNTPGCSCSRPVCTXNGLPVCG

Figure 1.1.3. Sequence of viola peptide-I (Schöpke *et al.*, 1993). X is an unknown residue.



**Figure 1.1.2.** Topology of kalata-B1. The backbone is shown by the red ribbon and the disulphide bonds are shown in yellow in ball and stick form.

Like kalata-B1, viola peptide-I has 29 amino acids and has at least 90% sequence homology with kalata-B1. Difficulties were experienced with sequencing and inferences of a cyclic nature for viola peptide-I and the presence of an unknown number of disulphides are indicated in the original literature report (Schöpke *et al.*, 1993).

More recently, a second peptide from *Viola arvensis* Murray, that is virtually identical to kalata-B1, with only one different residue, was reported (Claeson *et al.*, 1998). The peptide was extracted from the aerial parts of the plant and was found to be cyclic. The sequence of the peptide, named varv peptide A, is given in Figure 1.1.4. Varv peptide A is virtually identical to viola peptide-I.

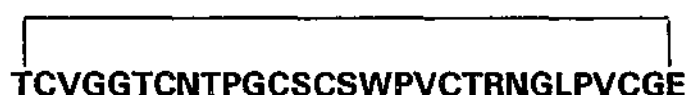


Figure 1.1.4. Sequence of varv peptide A (Claeson *et al.*, 1998).

The extraction of varv peptide A is important, as it formed part of validation of a fractionation protocol for the isolation of polypeptides from plant biomass (Claeson *et al.*, 1998). The protocol had been developed specifically for the purpose of extraction and purification of plant peptides, which to date is not well documented. Plant materials consist of a complex matrix, from which an array of components not found in animal material, such as photosynthetic pigments, polysaccharides, tannins, metabolites, all complicate the isolation of plant peptides. The protocol is designed to remove these components, permitting isolation of plant polypeptides. This fractionation protocol is given in Figure 1.1.5 and could be extremely useful for the isolation of other macrocyclic plant peptides.

With modifications to the fractionation protocol, seven more macrocyclic peptides, varv peptides B-H, from the aerial parts of *Viola arvensis* were discovered in addition to the main peptide, varv peptide A (Göransson *et al.*, 1999). By taking advantage of an adsorption mechanism of chromatography generated through the use of Sephadex LH-20, as well as

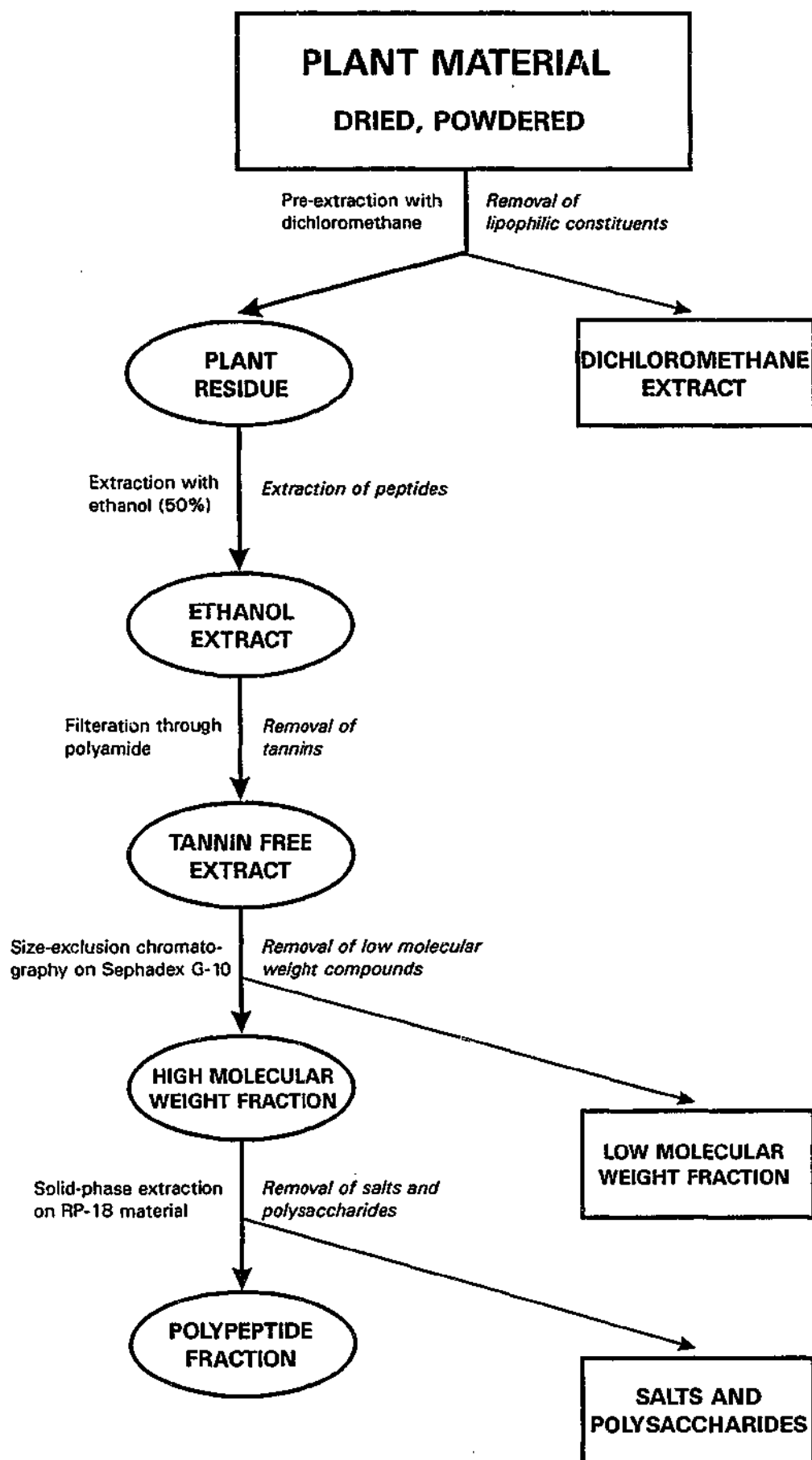


Figure 1.1.5. Fractionation protocol for the isolation of polypeptides from plant biomass (Claeson *et al.*, 1998).

different solvent conditions in the purification procedure, two separate cocktails of varv peptides were produced. Reverse phase HPLC on these two fractions enabled the eight different peptides to be isolated and purified. The sequences of varv peptides A-H are closely related. They consist of 29-30 amino acid residues and overall have a sequence homology of 63%. The sequences of varv peptides B-H are shown in Figure 1.1.6.



Figure 1.1.6. Sequences of varv peptide B-H (Göransson *et al.*, 1999).

The finding of so many similar macrocyclic peptides in one plant suggests that these peptides could be variants for a common purpose, necessary for plant survival. It has been proposed (Göransson *et al.*, 1999)

that the combination of the similar peptides allows for an optimal effect on as yet an unknown biological target.

### 1.1.3 Cyclopsychotride

Cyclopsychotride A appeared in the literature 1994 (Witherup *et al.*, 1994). This cyclic peptide consists of 31 amino acid residues and was extracted with an organic solvent from the tropical plant *Psychotria longipes*. Some 1400 tropical trees and shrubs belong to the genus *Psychotria* and many contain substances used by indigenous tribes for their psychopharmacological effects. Cyclopsychotride A has no such usage, however, it was found to selectively inhibit neurotensin binding to its receptors. It also stimulated intracellular  $\text{Ca}^{2+}$ . The action of cyclopsychotride A could not be blocked by known neurotensin antagonists, suggesting that the peptide acts through different receptors. The sequence of cyclopsychotride A is given in Figure 1.1.7.



Figure 1.1.7. Sequence of Cyclopsychotride A (Witherup *et al.*, 1993).

The  $^1\text{H}$  NMR spectrum indicates similarity with kalata-B1, in that cyclopsychotride A is a highly structured, compact peptide with resonances unlike those of unstructured peptides, a large number of slow exchange amide protons (20) and large vicinal coupling constants. Circular dichroism (CD) data shows the presence of the following secondary structure:  $\beta$ -sheet, 29%;  $\beta$ -turn, 23%; helix, 14%. The authors (Witherup *et al.*, 1994) are continuing with studies to discover other activities, decipher the mechanism of activity and determine the fold of the peptide. The results of these studies will be interesting to compare to kalata-B1.



#### 1.1.4 Circulins

In 1994, two very similar cyclic peptides with three disulphide bonds, circulin A with 30 amino acids and circulin B with 31 amino acids were reported (Gustafson *et al.*, 1994). The peptides were isolated from the crude extracts of the tropical tree, *Chassalia parvifolia* of the Iringa Region of Tanzania. The sequences of the peptides are given in Figure 1.1.8. The cysteine spacings are similar to those in kalata-B1 and in cyclopsychotride A. Furthermore, they also share some sequence similarity with human B-cell antigen CD22.

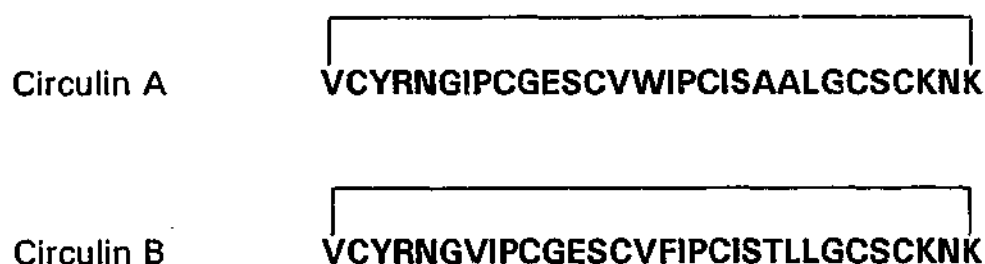


Figure 1.1.8. Sequences of circulin A and circulin B (Gustafson *et al.*, 1994).

The naming of these peptides is an unfortunate duplication of the names, which have also been given to two antibiotics isolated from the culture fluid of *Bacillus circulans* Q19 in 1949 (Hayashi *et al.*, 1968). The antibiotics, circulin A and B, are cyclic decapeptides (Fujikawa *et al.*, 1965; Hayashi *et al.*, 1968), as opposed to the macrocyclic circulins (Gustafson *et al.*, 1994), which are much larger and the subject of discussion in this section and of work for this thesis.

The macrocyclic circulin peptides are of considerable interest as they are active against several diverse strains of human immunodeficiency virus (HIV) (Gustafson *et al.*, 1994). Fast atom bombardment mass spectrometry (FAB-MS) data indicated the presence of three disulphide bonds, which are crucial for the anti-HIV properties, as reduction of the bonds results in a loss of cytoprotective activity for both circulins.

Determination of the specific disulphide connectivities of the circulins is difficult, as they are resistant to enzymatic proteolysis, and acid hydrolysis results in a complex mixture of cleaved products. However, the disulphide connectivities were deciphered by application of peptide recognition software to identify the fragments produced by FAB-MS (Derua *et al.*, 1996). The most likely disulphide connectivity is identical to kalata-B1 for both circulin peptides.

Knowledge of the 3D structure of the circulins would provide important information for the characterisation of the complex activities of the circulins and may assist in the development of pharmacological compounds. Through communication with the authors of initial report on circulins (Gustafson *et al.*, 1994), circulin samples were acquired for this study. A major part of the work for this thesis involves the determination of the structures of the circulins.

#### **1.1.5 Synthesis of Macrocyclic Polypeptides**

It is becoming increasingly apparent that the macrocyclic plant peptides are more abundant and important in the plant kingdom than originally thought. For greater understanding of their biological role and pharmaceutical significance, larger quantities of these peptides will be required for detailed work programs to proceed. Methods to prepare such quantities will be in demand. In light of this need, the synthesis of these peptides has been examined.

The multiple disulphide bonds, the cyclic nature and highly constrained, complicated knot motif present a substantial degree of difficulty for synthesis of these peptides. However, a biomimetic strategy was recently developed (Tam and Lu, 1998; Tam *et al.*, 1999a) that is comparatively simple and direct, avoiding an elaborate scheme involving multitiered protecting groups. The method is a three part strategy, starting with the stepwise solid-phase synthesis of the unprotected peptide precursor as a COOH-thioester, followed by thiazolidine cyclisation for the formation of the end-to-end cyclic peptide in an

unprotected form and completed with a two-step disulphide method to form the disulphide bonds.

The general strategy is based on a convergence of chemical proteolysis and aminolysis of peptide bonds through acyl transfer and it has been successfully applied in the synthetic preparation of circulin B and cyclopsychotride A. The diagram of the scheme for circulin B is depicted in Figure 1.1.9. In addition to this method, alternative synthetic methodology has been developed for the production of kalata-B1 (Daly *et al.*, 1999).

The methodology commences with the synthesis of linear precursors and relies on determined hydrophobic conditions for optimum procurement of native kalata-B1. Two different strategies were investigated. In the first strategy, the cysteine residues of the linear precursor are oxidised prior to cyclisation, while in the second strategy, the linear precursor is cyclised and then, the cysteines are oxidised. The schematic strategies are shown in Figure 1.1.10.

It was established (Daly *et al.*, 1999) that the oxidation conditions have a significant effect on the yield of the native fold. It seems that a hydrophobic environment of 50% (v/v) 2-propanol, which probably stabilises surface exposed hydrophobic residues, is necessary to obtaining native kalata-B1, using the first strategy, and greatly enhances the yield of the native form, using the second strategy. The oxidation results, under the different solvent combinations, have implications for the folding of kalata-B1, in that cyclisation of the linear precursor encourages the formation of native disulphide bonds, suggesting that burial of cysteine residues govern the folding process.

Cyclisation is also found to be critical to hemolytic activity, as uncyclised, folded kalata-B1 is inactive relative to the entire peptide, which is mildly active. This raises interest in the activities of macrocyclic polypeptides. Until recently, the information on the spectrum of general activities has been limited.

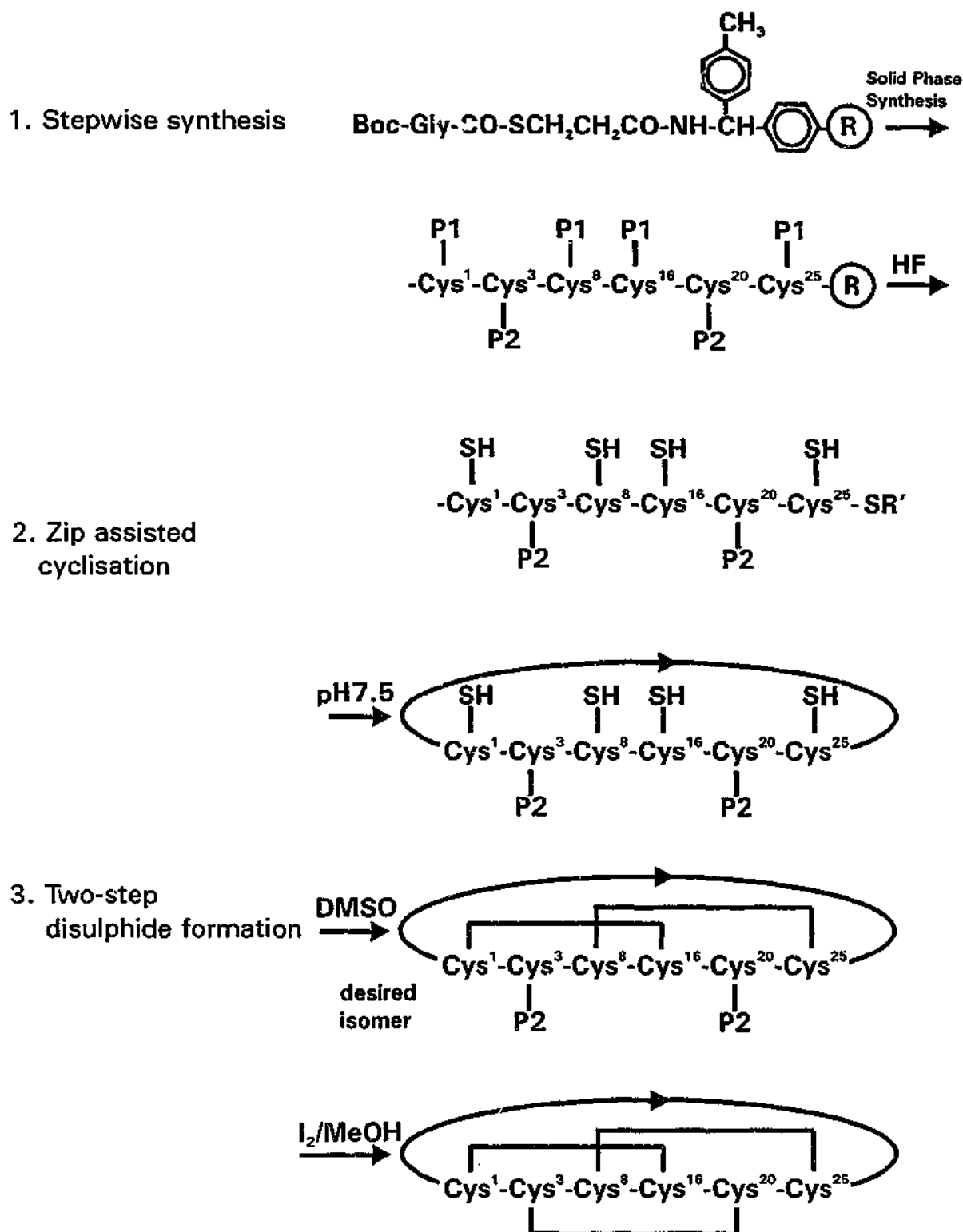


Figure 1.1.9. Scheme for synthesis of circulin B (Tam and Lu, 1998). Protected peptide is obtained by stepwise synthesis (step 1) and the cyclic peptide from the zip-assisted cyclisation (step 2). The disulphide pairs are obtained from the two-step approach (step 3). Four protecting groups (P1 = MeBz) were cleaved by HF and two disulphide bonds are formed by DMSO oxidation to give three disulphide isomers. The desired isomer is treated with I<sub>2</sub> to remove the second set protecting groups (P2 = AcM) to form the third disulphide pair. R' = SCH<sub>2</sub>CH<sub>2</sub>CONH<sub>2</sub>; R = resin. Cysteine positions correspond to those of circulin B. [HF = hydrogen fluoride; DMSO = dimethyl sulphoxide; I<sub>2</sub> = iodine; MeOH = methanol; MeBz = 4-methylbenzyl; AcM = acetamidomethyl]

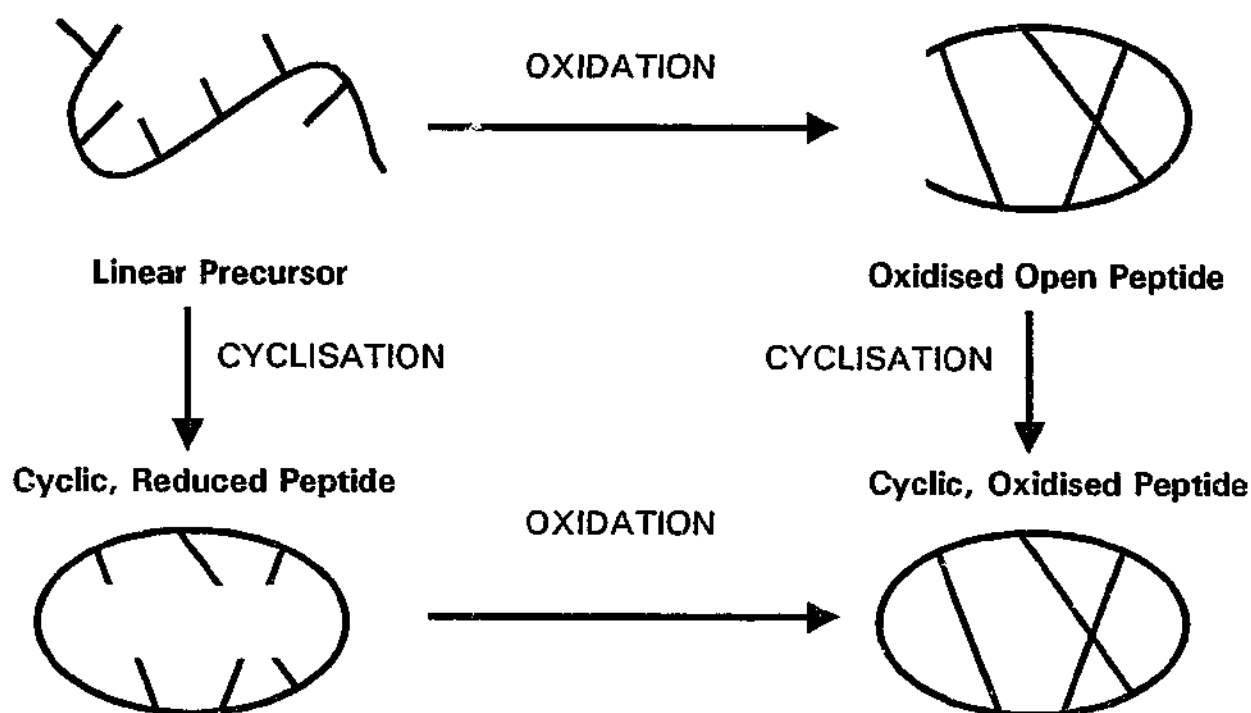


Figure 1.1.10. Scheme showing the two different strategies applied in the synthesis of kalata-B1 (Daly *et al.*, 1999). In the first strategy, the linear precursor is oxidised prior to cyclisation. In the second strategy, the linear precursor is cyclised prior to oxidation.

Activity	Organism	Kalata-B1 MIC, $\mu$ M	Circulin A MIC, $\mu$ M	Circulin B MIC, $\mu$ M	Cyclopsy- chotride MIC, $\mu$ M
Bacterial, Gram Negative	<i>Escherichia coli</i>	> 500	> 500	0.41	1.55
	<i>Pseudomonas aeruginosa</i>	> 500	> 500	25.5 <sup>a</sup>	13.5 <sup>a</sup>
	<i>Proteus vulgaris</i>	> 500	54.6	6.8	13.2
	<i>Klebsiella oxytoca</i>	54.8	> 500	8.2 <sup>a</sup>	5.8 <sup>a</sup>
Bacterial, Gram Positive	<i>Staphylococcus aureus</i>	0.26	0.19	13.5	39.0
	<i>Micrococcus luteus</i>	40.4	> 500	> 500	48.0
Fungal	<i>Candida albicans</i>	> 500	> 500	> 500	> 500
	<i>Candida kefir</i>	21.4	18.6	29.0	14.0 <sup>a</sup>
	<i>Candida tropicalis</i>	> 500	19.4	> 500	56.5
Hemolytic	Human erythrocytes	1510 <sup>b</sup>	1020 <sup>b</sup>	550 <sup>b</sup>	405 <sup>b</sup>
Cytotoxic	Mouse fibroblasts	-	-	820 <sup>c</sup>	1850 <sup>c</sup>

a Activity is apparent in salt concentration.

b concentration at which there is 50% red blood cell hemolysis.

c concentration at which there is 50% cell growth inhibition.

Table 1.1.1. Antimicrobial, hemolytic and cytotoxic activity of kalata-B1, circulin A and B and cyclopsychotride A in low salt assays (Tam *et al.*, 1999b).

### 1.1.6 Antimicrobial Activity

The presence of cationic charge and hydrophobic residues on the structural surface of kalata-B1 is reminiscent of functional characteristics of antimicrobial cationic peptides (Hancock *et al.*, 1995). These observations were the impetus behind a recent extensive investigation of the antimicrobial activity of kalata-B1, cyclopsychotride A, circulins A and B (Tam *et al.*, 1999b). The inhibitory activity against nine strains of well known microbes, including Gram negative and Gram positive bacteria and fungi, along with hemolytic activity and cytotoxic activity were measured. The results are presented in Table 1.1.1.

There are notable differences in activity profiles between the macrocycles. Kalata-B1 is active against the two Gram positive bacteria, especially *S. aureus*, and is inactive against the Gram negative bacteria, except *K. oxytoca* against which, it is moderately active. Kalata-B1 is moderately active against one fungus, *C. kefyr*. Circulin A is similar in profile to kalata-B1. Circulin A is potent against *S. aureus*, although inactive against the other Gram positive bacterium tested, and is inactive against the Gram negative bacteria, except for having moderate activity against *Pr. vulgaris*. Circulin A is moderately active against two fungi, *C. kefyr* and *C. tropicalis*.

Circulin B and cyclopsychotride A both have broader activity ranges. Both are active against all the Gram negative bacteria and are particularly potent against *E. coli*. They are moderately active against the Gram positive bacterium, *S. aureus*, and cyclopsychotride A, is also moderately active against the other Gram positive bacterium tested, *M. luteus*. Both are moderately active against the fungus, *C. kefyr*, and cyclopsychotride A, is also active against the fungus, *C. tropicalis*.

None of the peptides were active against the fungus *C. albicans*. All the peptides are mildly hemolytic and circulin B and cyclopsychotride A are cytotoxic. Activities were also measured in a high salt environment for each peptide and most activities were removed, indicating that there is an electrostatic component involved in the processes of peptide-microbe

interactions. Further information on activity was obtained by activity measurements on various single deleted and scrambled disulphide peptide analogs, as well as blocked/modified cationic arginine analogs.

With regard to one deleted disulphide, the activity profile is generally maintained. However, in some cases there is a two to three fold decrease in activity and on occasion, there is increased activity. For analogs with scrambled disulphides, again there is no significant loss of activity and in some situations, there is improved activity.

Considering the analogs with changes to cationic charge, blocking the single cationic residue, arginine in kalata-B1, resulted in severe reduction in activity. However, modification to the arginine residue in circulin A, which is one of three cationic charged residues, resulted in minimal activity loss against bacteria, but a two and a half fold decrease in activity against the fungi. The study of analogs is by no means complete. Nonetheless, it infers that deletion of single disulphides and scrambling disulphide bonds are less critical to activity than cyclisation. However, the cationic nature of the peptide could be important in activity.

In summary, the antimicrobial study has substantially contributed to the clinical activity profiles of four macrocyclic polypeptides. It is both curious and exciting that the four different plant peptides exhibit selective antimicrobial activity, thus proposing a foundation for the serious potential for the development of a template for novel antibiotics.

#### **1.1.7 Expansion and Comparison of the Macrocyclic Polypeptide Database**

Recently, the database of macrocyclic peptides extracted from Rubiaceae and Violaceae plant species has dramatically expanded to currently include a total of 37 different sequences. Four new circulin peptides, circulin C, D, E and F were extracted from *Chassalia parvifolia* and were characterised (Gustafson *et al.*, 2000). All four peptides inhibit the cytopathic effect of *in vitro* HIV-1 infection. These peptide sequences are highly homologous with circulins A and B. However, circulins D and E carry an

overall neutral charge, due to a different charge distribution from the other circulins, which carry an overall +2 charge.

Another four macrocyclic peptides were isolated from the bark material of the tropical plant in Peru, *Leonia cymosa*, of the Violaceae plant family (Hallock *et al.*, 2000). Cycloviolins A, C and D share a high degree of sequence homology with circulins A and B and cyclopsychotride A and similarly carry an overall +2 charge, due to identical charge location and identity in the sequence. Cycloviolin B, however, appears distinct. Although, there is considerable sequence homology with the other cycloviolins, cycloviolin B has the shortest sequence, lacks a proline, carries an overall neutral charge and has a unique TSSQ sequence.

The database of macrocyclic peptides was doubled with addition of sixteen peptides extracted from *Viola hederaceae*, *Viola odorata* and *Oldenlandia affinis* (Craik *et al.*, 1999). These peptides included cycloviolacin O1 to O12 from *Viola odorata*, cycloviolacin H1 from *Viola hederaceae* and kalata B2 to B5 and kalata S from *Oldenlandia affinis*. In this publication, the presence of certain sequences in multiple species is reported for the first time. Kalata-B1 was isolated from both Rubiaceae and Violaceae plants and cycloviolacin O12 is identical to varv peptide E as is kalata S to varv peptide A. The 3D structure of cycloviolacin O1 was determined and is very similar to that of kalata-B1.

The expansion of the database permits more detailed comparison of sequences and identification of trends (Craik *et al.*, 1999). Two subfamilies have been broadly classified on the basis of backbone cyclisation. In one subfamily, the backbone forms a bracelet-like closed loop, comprised of two sides. In the second family, which contains kalata-B1, the backbone is found to contain a twist in the cyclic sequence, due to a *cis* peptide bond between Trp2 and Pro3. In this context, the backbone can be likened to a molecular Moebius strip, in which there is a single continuous side.

Although, the *cis*-Pro conformation has been identified for only kalata-B1 (chapter 4), the presence of proline preceded by tryptophan is a recurrent sequential feature in a number of sequences. At this stage, it is assumed that



other sequences containing this segment in the identical location also feature a twist in the backbone. The 37 sequences can be initially sorted adopting this classification based on backbone cyclisation.

A comparison of the amino acid sequences for the macrocyclic plant peptides is given in Figure 1.1.11 (Craig *et al.*, 1999; Craig *et al.*, 2001). The sequences are aligned according to the cystine positions. As the sequences are cyclic, the first cystine is arbitrary and is chosen with reference to the first appearance of the alignment of kalata-B1 with non cyclic peptides, with comparable topology isolated from other biological sources (Pallaghy *et al.*, 1994). The sequence alignment of Figure 1.1.11 clearly shows the sequence homology of these macrocyclic plant peptides, which could be indicative of common biochemical ancestry.

Using the cysteine residues for demarcation, six loops can be identified, as shown in Figure 1.1.11. The general formula of macrocyclic peptides, according to this six loop structure, may be represented as  $C(1)X_3C(2)X_4C(3)X_{4-7}C(4)X_1C(5)X_{4-5}C(6)_{5-7}$ , where X is any non cysteine residue and the subscript refers to the general number of intervening residues between the cysteines. The loops exhibit certain trends across both families.

It is interesting to note that many of the hydrophobic residues occur in similar positions for all the sequences. Similarly, the charged residues also appear in similar locations. There is a conserved negatively charged Glu residue in loop 1 and charged residues seem to congregate in loops 5 and 6. Loops 1 and 4 are absolutely conserved, forming an embedded ring (chapter 1.2). Loop 6 is also highly conserved. There are also patterns that characterise each subfamily. The sequences of the bracelet family often carry an overall +2 charge and some are neutral, while the sequences of the Moebius strip family carry either an overall negative or neutral charge. Most variations between the subfamilies are manifest in loops 2, 3 and 5.

No doubt, in time, as more new sequences are extricated, further classifications and patterns are likely. It is becoming increasingly apparent that the macrocyclic peptides may be a prevalent plant component. Much work is required in the discovery and full characterisation of the sequences,

Cysteine residue number		1	2	3	4	5	6						
BRACELET FAMILY													
cycloviolacin O1		A	E	S	V	Y	I	P					
cycloviolacin O2		G	E	S	V	W	I	P					
cycloviolacin O3		G	E	S	V	W	I	P					
cycloviolacin O4		G	E	S	V	W	I	P					
cycloviolacin O5		G	E	S	V	W	I	P					
cycloviolacin O6		G	E	S	V	W	I	P					
cycloviolacin O7		G	E	S	V	W	I	P					
cycloviolacin O8		G	E	S	V	W	I	P					
cycloviolacin O9		G	E	S	V	W	I	P					
cycloviolacin O10		G	E	S	V	Y	I	P					
cycloviolacin O11		G	E	S	V	W	I	P					
cycloviolacin H1		G	E	S	V	Y	I	P					
kalata-B5		G	E	S	V	Y	I	P					
circulin A		G	E	S	V	W	I	P					
circulin B		G	E	S	V	F	I	P					
cyclopsychotride A		G	E	S	V	F	I	P					
cycloviololin A		G	E	S	V	F	I	P					
cycloviololin B		G	E	S	Y	V	L	P					
cycloviololin C		G	E	S	V	F	I	P					
cycloviololin D		G	E	S	V	F	I	P					
circulin C		G	E	S	V	F	I	P					
circulin D		G	E	S	V	W	I	P					
circulin E		G	E	S	V	W	I	P					
circulin F		G	E	S	V	W	I	P					
MOEBIUS FAMILY													
viola peptide-I	V	G	E	T	V	G	G	T					
kalata B1	V	G	E	T	V	G	G	T					
kalata B2	V	G	E	T	F	G	G	T					
kalata B3	T	G	E	T	F	G	G	T					
kalata B4	V	G	E	T	V	G	G	T					
varv peptide A /kalata S	V	G	E	T	V	G	G	T					
varv peptide B	V	G	E	T	F	G	G	T					
varv peptide C	I	G	E	T	V	G	G	T					
varv peptide D	I	G	E	T	V	G	G	S					
varv peptide E/cycloviolacin O12	I	G	E	T	V	G	G	T					
varv peptide F	I	G	E	T	T	L	G	T					
varv peptide G	V	G	E	T	F	G	G	T					
varv peptide H	V	G	E	T	F	G	G	T					
		LOOP 1		LOOP 2		LOOP 3		LOOP 4		LOOP 5		LOOP 6	

**Figure 1.1.11.** Sequence alignment of macrocyclic peptides. Conserved residues, across all sequences, are given in bold. Hydrophobic residues are shown in green, polar residues in light blue, positively charged residues in dark blue, negatively charged residues in red and cysteines in yellow. Disulphide connectivities and the six loops are given at the bottom of the table.

as well as in the understanding of bioactive functions. This thesis is a study into the structural nature of this emerging family of macrocyclic plant peptides.

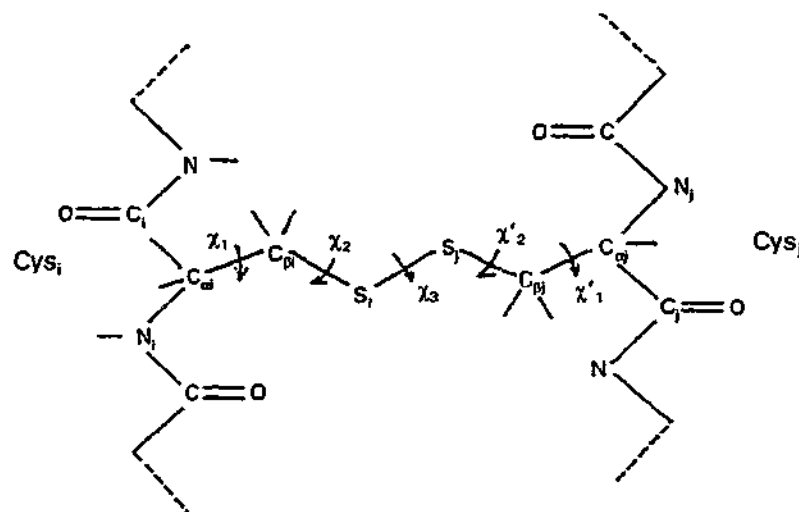
## 1.2 Cystine Knot Motifs

Multiple disulphide bonds and cystine knot motifs are important features of the polypeptides studied for this thesis. This section describes these features in greater detail, with the aim of developing a perspective of their significance in nature, as is currently understood.

### 1.2.1 Disulphide Bonds in Nature

The cysteine amino acid is unique in its ability to form disulphide bonds with other cysteines present in proteins and polypeptides. Through this ability to cross link the polypeptide backbone, this amino acid permits the formation of a wide variety of topological structures, to produce molecules of biological importance such as enzymes, protease inhibitors, plasma proteins, neurotoxins, hormones. Concomitant with the property of disulphides to confer topological structure, the presence of disulphides imposes a limit on the conformational space occupied by the backbones of peptides and proteins (Benham and Jafri, 1993). In fact, no other molecular interaction places such stringent structural requirements upon proteins and peptides as the disulphide bond.

Some of the structural limitations include the disulphide bond length, which is typically found to be  $2.05 \pm 0.03 \text{ \AA}$ , the central torsion angle about the disulphide bond, which is approximately  $90^\circ$  and the angle between the sulphur and the  $\beta$ -carbon of each cystine, which is approximately  $103^\circ$  (Steudel, 1995; Creighton, 1988). The main angles that define the disulphide bond arrangement are given in Figure 1.2.1. Another imposition of disulphide bonding in small peptides is that the intervening loop between the cysteine residues involves a  $\beta$ -turn potential (Thornton, 1981). This is a consequence of disulphide bond formation, which requires a  $180^\circ$  reversal of the polypeptide chain in order for the bond to form. This can lead to production



**Figure 1.2.1.** Atom and angle nomenclature that define the disulphide bond arrangement (Thornton, 1981).

of complicated arrangements in 3D space, particularly, where multiple disulphide bonds are involved, which is a frequent occurrence.

Disulphide bonds are most commonly found in extracellular proteins and are typically crucial for full activity (Thornton, 1981). Compared to the reduced -SH state, which is inherently more bulky, the disulphide allows for stronger van der Waals contacts and a decrease in the side chain volume, leading to a more compact structure that reduces conformational fluctuations that would otherwise occur. Thus, disulphides tend to be buried within the overall structure, remaining fairly inaccessible to solvents. The main role of disulphides appears to be restriction of mobility of certain segments, stabilisation of the structure and maintenance of the overall fold in an effort to preserve activity (Thornton, 1981; Klaus *et al.*, 1993). So important is the disulphide bond, that from the perspective of evolution, cyst(e)lines are often absolutely conserved rendering this amino acid as the second most conserved amino acid after tryptophan (Thornton, 1981).

Following on from cyst(e)line conservation are observations of general patterns for disulphide topology (Thornton, 1981). For instance, there seems to be a strong tendency to form shorter connections, known as local

disulphides, which are defined as connections extending over fewer than 45 residues. In fact, some 49% of disulphides were found to extend over fewer than 24 residues, with the most common disulphide connection extending over 10 to 14 residues. This preference for shorter connections illustrates the domain structure of many biological molecules. Both local and non-local disulphide bonds have been individually associated with certain secondary structural elements. All the disulphide bonds in the polypeptides examined for this thesis fall into the category of being local. For local disulphides, the secondary structural patterns involving the half-cystines, in order of decreasing prevalence, are coil-coil (cc), c- $\beta\beta$ -c where two antiparallel strands separated by a reverse turn are held together by a disulphide, coil-helix ( $\alpha$ c), strand helix ( $\alpha\beta$ ), strand-strand ( $\beta\beta$ ). These patterns, which are illustrated in Figure 1.2.2 indicate a preference for half-cystines to be involved with coil regions.

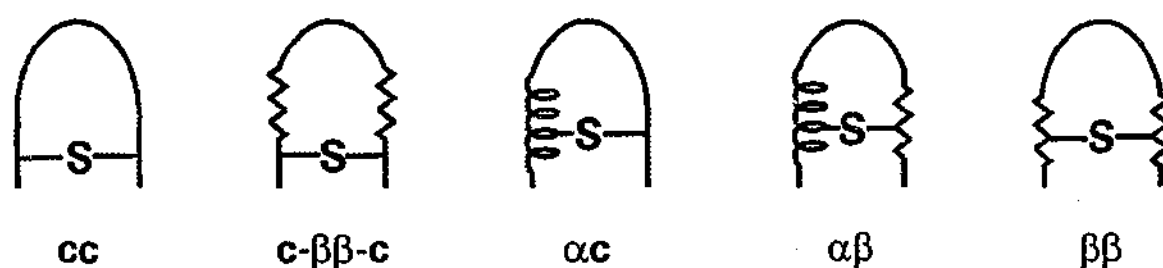


Figure 1.2.2. Representations of structural elements associated with local disulphide bonds where  $\alpha$  = helix,  $\beta$  = strand, c = coil (Thornton, 1981).

With regard to disulphide conformation, there appears to be an equal distribution of left- ( $\chi_3 = -90^\circ$ ) and right-handed ( $\chi_3 = 90^\circ$ ) structures (Richardson, 1981). The left-handed spiral structure, with angles of approximately  $\chi_1 = -60^\circ$ ,  $\chi_2 = -90^\circ$ ,  $\chi_3 = -90^\circ$ ,  $\chi_2' = -90^\circ$ ,  $\chi_1' = -60^\circ$ , has an average  $C_\alpha$ - $C_{\alpha'}$  of 6.1 Å and is the predominant conformation for left-handed structures. Right-handed disulphide structures are more variable in conformation. The majority of these adopt the right-handed hook structure with angles of approximately  $\chi_1 = -60^\circ$ ,  $\chi_2 = 120^\circ$ ,  $\chi_3 = 90^\circ$ ,  $\chi_2' = -50^\circ$ ,  $\chi_1' = -60^\circ$ .

and an average  $C_{\alpha}$ - $C_{\alpha'}$  of 5.2 Å, which is shorter than that for the left-handed spiral structure, indicating that these two major conformations are not mirror images. Examples of the major conformations are shown in Figure 1.2.3.

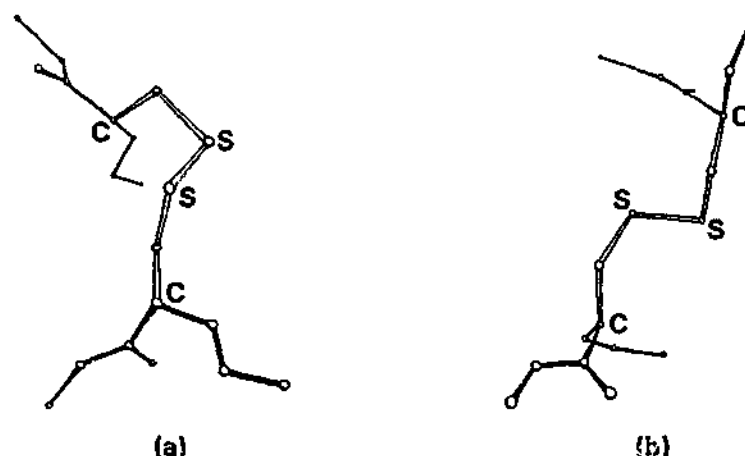


Figure 1.2.3. Examples of the two major disulphide conformations (a) left-handed spiral (b) right-handed hook (Richardson, 1981).

In summary, the covalent disulphide bond broadens the complexity of topological structure in nature and interestingly, is not the product of chance factors, but directed processes.

### 1.2.2 Structural Superfamilies

Over the last few years, it has become increasingly apparent that there is a limited number of topologically different protein folds in nature (Chothia, 1992). With the use of x-ray crystallography and Nuclear Magnetic Resonance (NMR) techniques, the rate of elucidation of 3D structures of proteins and polypeptides has increased dramatically. It has become evident that in some cases, proteins with less than 20% sequence identity can have very similar secondary and tertiary structure (McDonald and Hendrickson, 1993). Excellent superpositions of invariant residues in topologically equivalent positions of proteins and polypeptides from widely varying sources

are often obtained. This is strongly suggestive that the similar structures evolved from a common ancestor. Evidently, only a few key residues are necessary to conserve the overall 3D structure and hence the study of tertiary folds could lead to important findings for evolutionary history.

Based on these observations, similar structures can be grouped into structural superfamilies. The members of these families are not necessarily related in regional function, which would confer a different classification. The members can also originate from completely different sources. The emergence of structural superfamilies may be illustrated with polypeptide growth factors.

Many polypeptide growth factors are known. They are involved in various stages of cell growth and survival in different regions of the human body. However, growth factors may also be classified according to their structure, as shown in the Table 1.2.1. It is expected that as the structures of more growth factors are determined, the structural superfamily sizes will increase. Of special interest to the work for this thesis is the fourth entry in the column of superfamilies, the cystine knot.

The cystine knot motif was clearly identified for a set of polypeptide growth factors in 1993 (McDonald and Hendrickson, 1993; Murray-Rust *et al.*, 1993). The knot motif is characterised by a certain disulphide bonding pattern, involving six half-cystines. More specifically, if the half-cystines are numbered from 1 to 6 in order of appearance in the sequence, the bonding pattern is Cys(1)-Cys(4), Cys(2)-Cys(5), Cys(3)-Cys(6). Such a disulphide bonding pattern also occurs for another set of polypeptides consisting of a neurotoxin from a cone snail, plant enzyme inhibitors and a cardiotoxic plant component. However, there are significant topological differences between the two polypeptide sets, which are distinguished as the growth factor cystine knot, including the growth factors and the inhibitor cystine knot, including the inhibitors and toxins (Isaacs, 1995). The two sets are in fact topologically distinct and it is the latter of these, the inhibitor cystine knot that is the concern of this thesis. However, it is worthwhile from the point of



Superfamily	Members
$\beta$ -Trefoil	FGF IL-1 $\alpha$ , IL-1 $\beta$ Int-2 Keratinocyte growth factor
Four-Helix Bundle (Two Crossovers)	Growth hormone IL-2 G-CSF CNTF
EGF-like	EGF TGF $\alpha$ Schwann cell-derived growth factor Vaccinia virus growth factor
Cystine Knot	NGF TGF $\beta$ PDGF v-Sis
Insulin-like	Insulin IGF-1, IGF-2 Bombyxin
Mosaic	Neu differentiation factor Glial growth facto/ARIA
Other	TNF CD40 ligand

Table 1.2.1. Possible classification of representative growth factors into structural superfamilies (McDonald and Hendrickson, 1993).

being complete to briefly discuss the growth factor cystine knot.

#### 1.2.2.1 Growth Factor Cystine Knot

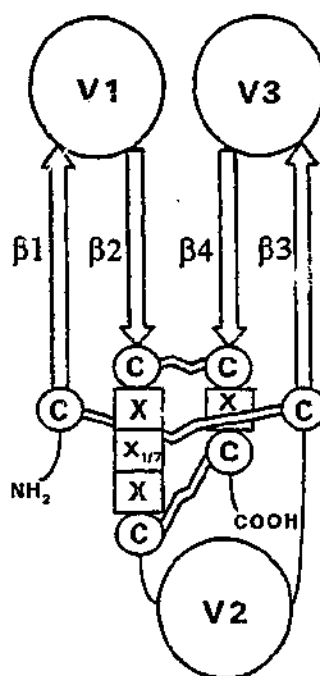
This family consists of NGF (nerve growth factor), TGF $\beta$ 2 (transforming growth factor), PDGF-BB (platelet-derived growth factor), NDP (Norrie disease protein), hCG (human chorionic gonadotrophin) (McDonald and Hendrickson, 1993; Murray-Rust *et al.*; 1993; Meltinger *et al.* 1993; Isaacs, 1995). The basic knot motif consists of a ring formed by the connecting residues of the disulphides bonds Cys(2)-Cys(5) and Cys(3)-Cys(6), through which the third disulphide Cys(1)-Cys(4) passes at approximately right angles to form a disulphide cluster. The ring size varies between the members, with NFG having a 14 membered ring, and TGF $\beta$ 2 and PDGF-BB both having an 8 membered ring.

The second common feature is four extended segments of twisted antiparallel  $\beta$ -strands, which results in two distorted  $\beta$ -hairpin loops on one side of the knot and a single loop on the other side. All of the strands ( $\beta$ 1- $\beta$ 4) vary in length and degree of twist between the family members. The fourth common strand ( $\beta$ 4) for all the members has a larger twist and is stabilised by a  $\beta$ -bulge. A schematic diagram of the growth factor cystine knot is given in Figure 1.2.4. The six half-cystines along with the  $\beta$ -strands are the only conserved features.

Other common features are the sequence Cys(3)-X-Gly-X-Cys(4), where the glycine residue has a positive  $\Phi$  main chain conformation, and also the Cys(5)-X-Cys(6) sequence. Furthermore, the hydrogen bonding is very similar particularly around the cystine cluster. However, there are some differences between the members of the family.

The number of residues between Cys(3) and Cys(4) varies from one member to the other. Also, there are three variable regions (V1-V3), occurring at the extremities of the strands. These regions together with the *N*-terminus contain the main differences between the family members and it

appears that the three variable regions contain the receptor-binding determinants. Each growth factor of the cystine knot group belongs to a distinct family of molecules and has no functional similarity, except that they each induce a response by binding to a specific cell surface receptor kinase.



**Figure 1.2.4.** Schematic diagram of the growth factor cystine knot (McDonald and Hendrickson, 1993). Half-cystine residues are indicated by open circles labelled with a C. Squares containing an X represent any amino acid and show the spacings between half-cystines. 1/7 refers to a spacing of 1 or 7 residues. Half-cystine connectivities are indicated by crooked bars. Variable regions are illustrated by a large open circle labelled with a V. The  $\beta$ -strands are designated by arrows indicating the direction of the strand.

All the cystine knot growth factors exist and are active in the form of dimers, although the mode of dimerisation and protomer interfaces differ between the family members. An indication of the differences in dimer structures is shown in Figure 1.2.5. The cystine knot motif thus acts as stable template, which can accommodate additional loops with low sequence similarity. It is these additional loops of the template that contain the specificity for receptor binding and hence the functional differentiation.

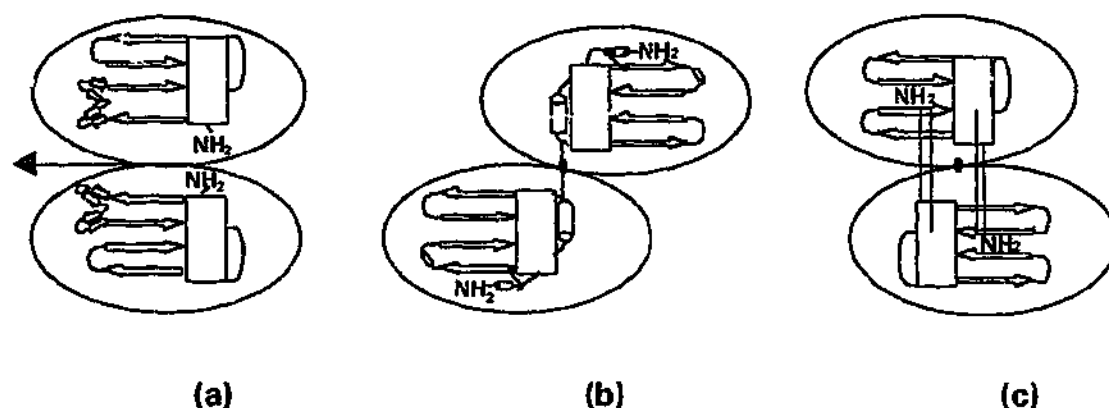


Figure 1.2.5. Schematic diagram of dimer structures of (a) NFG (b) TGF- $\beta$ 2 (c) PDGF-BB. (McDonald and Hendrickson, 1993). Arrows are  $\beta$ -strands; cylinders,  $\alpha$ -helices; rectangles, cystine knots; closed circles, intersubunit half-cystines; open circles, additional intrasubunit half-cystines. The position of the dyad axis relating two subunits in each dimer is shown. For TGF- $\beta$ 2 and PDGF-BB, the dyad axis points into the page, whereas for NGF, it is in the plane of the page.

### 1.2.2.2 Inhibitor Cystine Knot

Founding members of this family include  $\omega$ -conotoxin GVIA, an N-type  $\text{Ca}^{2+}$  channel blocking neurotoxin from the cone snail *Conus geographus*;  $\omega$ -agatoxins IVA and IVB, P-type  $\text{Ca}^{2+}$  channel blocking toxins from the American funnel web spider *Agelenopsis aperta*; kalata-B1, a cardiotoxic, uterotonic polypeptide from the tropical African plant, *Oldenlandia affinis* DC; CMTI-I, a trypsin inhibitor from pumpkin seeds of *Curcubita maxima* which belongs to the squash plant family; EETI-II, a trypsin inhibitor from the plant *Ecballium elaterium*, which belongs to the squash plant family; and CPI or PCI, carboxypeptidase potato inhibitor from a plant that belongs to the squash plant family (Pallaghy *et al*, 1994; Narasimhan *et al*, 1994; Isaacs, 1995). The sequences of these polypeptides are given below in Figure 1.2.6.

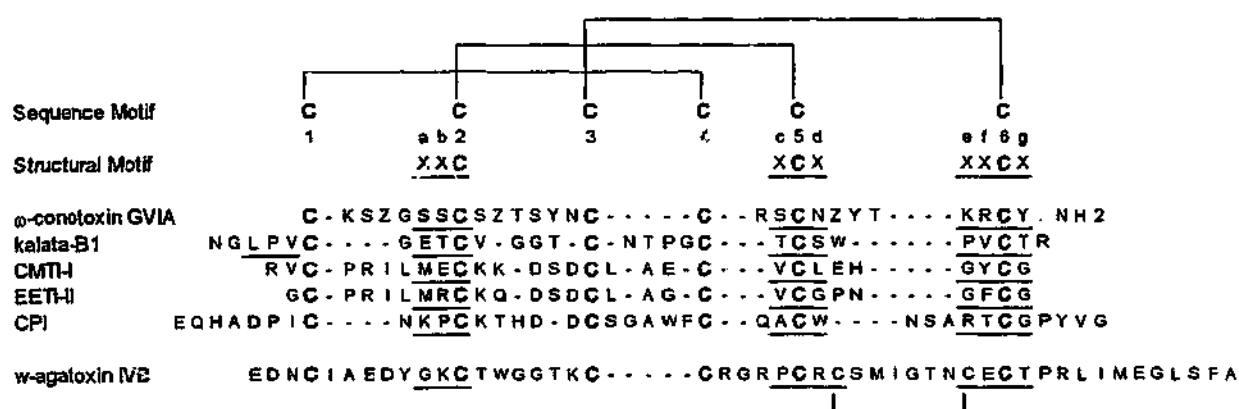


Figure 1.2.6. The sequences of the members of the inhibitor cystine knot family aligned according to the half-cysteine (in bold) positions and  $\beta$ -strand (underlined) hydrogen bonding patterns. The minimal  $\beta$ -strand motif common to all molecules XXC (strand I), XCX (strand II), and XXCX (strand III). X refers to any residue other than cysteine. Z refers to hydroxyproline.  $\omega$ -Agatoxin IVB has a fourth disulphide bridge as shown (Pallaghy *et al.*, 1994).

The inhibitor cystine knot family has since grown in size (Craik *et al.*, 2001), including polypeptides, originating from fungi, plant and animal sources, which have quite different biological activities. Furthermore, there is little sequence homology between the polypeptides from the different sources. Whilst the inhibitor and growth factor knots consist of the same main features of a cystine knot and a  $\beta$ -sheet arrangement, they are formed differently.

In the case of the inhibitor cystine knot, the ring is formed by connecting residues of the disulphides bonds Cys(1)-Cys(4) and Cys(2)-Cys(5), through which the third disulphide, Cys(3)-Cys(6), passes to essentially support a distorted triple-stranded antiparallel  $\beta$ -sheet, that contains a minimum of ten residues,  $X_aX_b\text{Cys}(2)-X_c\text{Cys}(5)X_d-X_eX_f\text{Cys}(6)X_g$ . The ring size, which determines the knot size, varies between the members. For instance, the ring of  $\omega$ -conotoxin GVIA is composed of 12 residues,  $\omega$ -agatoxins 15 residues, kalata-B1 8 residues and the squash inhibitors 11 residues and the potato inhibitor 9 residues. The smallest ring sizes in sequences originate from plants and fungi, while the larger rings are found in sequences from animals (Craik *et al.*, 2001). With one exception, the ring is asymmetric and is given by  $C(1)X_3-C(2)$  and  $C(4)X_1-C(5)$ .

The global fold of the inhibitor cystine knot can be described as being composed of four segments: a peripheral  $\beta$ -strand ( $\beta 1$ ), a connecting section containing turns or  $3_{10}$ -helix, a second peripheral  $\beta$ -strand ( $\beta 2$ ) and a central  $\beta$ -strand ( $\beta 3$ ). A turn joins the second and third strands,  $\beta 2$  and  $\beta 3$  to form a  $\beta$ -hairpin. The segments are separated by chain reversals. The Cys(2)-Cys(5) disulphide joins the first and second strands,  $\beta 1$  and  $\beta 2$ . The Cys(1)-Cys(4) disulphide links the *N*-terminus end of the main chain to the chain reversal area preceding the second  $\beta$ -strand,  $\beta 2$ . Finally the third disulphide, Cys(3)-Cys(6) which completes knot formation, joins the third  $\beta$ -strand,  $\beta 3$ , with the connecting section, linking the first and second  $\beta$ -strands. A schematic diagram of the inhibitor knot is given Figure 1.2.7.

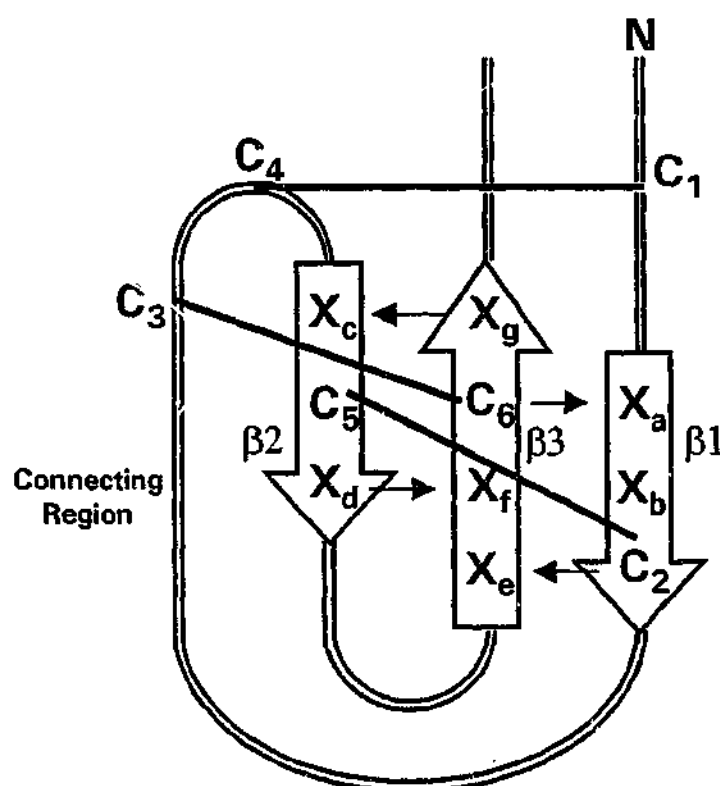


Figure 1.2.7. Schematic diagram of the inhibitor knot (Pallaghy *et al*, 1994). C represents a half-cystine residue and X represents any other residue. The common hydrogen bonds are indicated by arrows directed from the donor to the acceptor.

There are other consistent features between the family members. The gap size between consecutive half-cystines is reasonably maintained within the following pattern Cys(1) $X_{3-7}$ Cys(2) $X_{3-8}$ Cys(3) $X_{0-7}$ Cys(4) $_{1-4}$ Cys(5) $X_4$ .

<sup>13</sup>Cys(6). There is also a common hydrogen bonding pattern, with hydrogen bonds between NHCys(6)-COX<sub>2</sub>, NHCys(2)-COX<sub>3</sub>, NHX<sub>9</sub>-COX<sub>6</sub>, NHX<sub>4</sub>-COX<sub>1</sub>. In addition, there is a high degree of local structure similarity, with the chain reversals having high positional density. Chain reversals are found following the first  $\beta$ -strand, preceding the second  $\beta$ -strand and linking the second and third  $\beta$ -strands (refer to figure Figure 1.2.7). The latter chain enables the  $\beta$ -hairpin structure to be formed. The section prior to the first  $\beta$ -strand contains the most difference between the family members.

All the family members can be designated as pseudolinks, which are structures that can be unknotted by continuous deformation (Pallaghy *et al*, 1993 and 1994), with the exception of kalata-B1. Kalata-B1 unlike the other members is a cyclic polypeptide and consequently, has a more complicated topology, which is discussed in greater detail in a subsequent chapter. Kalata-B1 was the first of a multitude of homologous cyclic plant polypeptides, extracted from Rubiaceae and Violaceae plants (Craik *et al.*, 2001). Recently, these cyclic cystine knot peptides were separately sub classified as cyclotides, which are smaller, having the general structure, Cys(1)X<sub>3</sub>Cys(2)X<sub>4</sub>Cys(3)X<sub>4-7</sub>Cys(4)<sub>1</sub>Cys(5)X<sub>4-5</sub>Cys(6)X<sub>6-7</sub>. The second notable different members of the inhibitor knot family are exemplified by the American funnel web spider toxins,  $\omega$ -agatoxins IVA and IVB. These toxins have a fourth disulphide bridge, which can be considered independent of the other three that form the inhibitor cystine knot structure (Narasimham *et al.*, 1994).

The inhibitor cystine knot is a simpler structure than the growth factor cystine knot, which accommodates a system of loops and turns, and which also forms dimers that extend the diversity of functional structures. The inhibitor knot is a more compact framework that may function to present active residues for specific binding interactions (Pallaghy *et al*, 1993). Of the available data on the toxins and the plant inhibitors, the inhibitor cystine knot appears to act as an antagonist at specific targets. This is in contrast to the growth factor cystine knot polypeptides, which in the dimerised form may be

involved with signal transduction by inducing cooperative conformational changes in their cognate receptors (McDonald and Hendrickson, 1993).

Certainly, there is much work to be completed before the functions of these polypeptides are fully understood, so that this knowledge may be applied to clinical problems. To date, the above toxins have been used as probes to elucidate the distribution of and to characterise receptors (Hucho, 1995). The study of toxins may shed light on for instance the understanding of neuronal death following ischemia (Narasimham *et al.*, 1994), while the study of the growth factors may lead to enlightenment of the process of pathogenesis of diseases, such as cancer and autoimmune diseases. Determination of the 3D-structure of bioactive proteins and polypeptides is essential to gaining this knowledge.

The study of kalata-B1 and other similar cyclic plant peptides is part of the process and given their diverse bioactivities, their study could enhance our understanding of natural products, any of which may potentially be developed into lead pharmaceutical compounds in the future.



### **1.3 Peptide Folding**

During the course of the work for this thesis, synthesised fragment sequences that contain cysteine residues were oxidised and purified by high performance liquid chromatography (HPLC) before structural analysis. It is therefore important to understand the main driving forces of folding and oxidation of peptides with disulphide bonds.

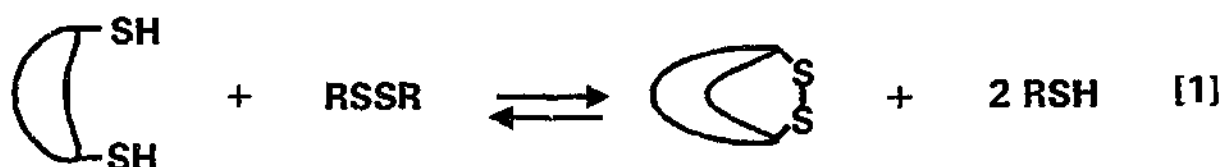
#### **1.3.1 Folding of Peptides and Proteins with Disulphide Bonds**

The presence of cysteine residues complicates folding in that the process involves, in addition, the formation of formal chemical disulphide bonds (Gilbert, 1994). Of course, with multiple cysteine residues, the number of possibilities for disulphide bond formation increases with the number of cysteines present which inherently increases the complexity of oxidative folding. The classic view of folding (Anfinsen, 1973) is that the primary sequence contains the information required to direct folding, so that the correct 3D structure is achieved and that thermodynamic stability of native structures is the driving force. Certainly, there appears to be a thermodynamic interrelation between disulphide stability and structural stability, with formation of disulphide bonds that then permit the creation of 3D structures that would not otherwise occur. Whilst the knowledge of the folding process has vastly increased over the years, there is still much that is unknown particularly within living organisms. Nonetheless, folding studies over the years have revealed that folding is a dynamic and highly co-operative process, where both thermodynamics and kinetics are major forces, which are affected and directed by environmental conditions and other interesting biochemical factors.

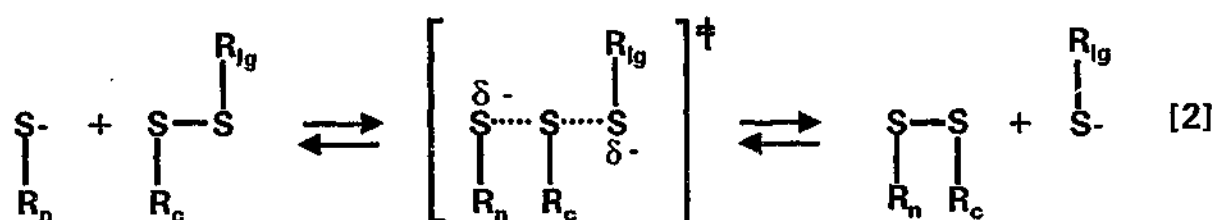
Prior to a discussion on the complexities of oxidative folding in which multiple cysteine residues are involved, it is appropriate to commence with a presentation of the important aspects of disulphide bond formation.

### 1.3.2 Chemistry of Disulphide Bond Formation

The formation of a disulphide bond requires an oxidant such as molecular oxygen in the presence of trace amounts of heavy metals or a low molecular weight disulphide. The reaction is a two electron oxidation reaction as shown in equation 1 (Gilbert, 1994).



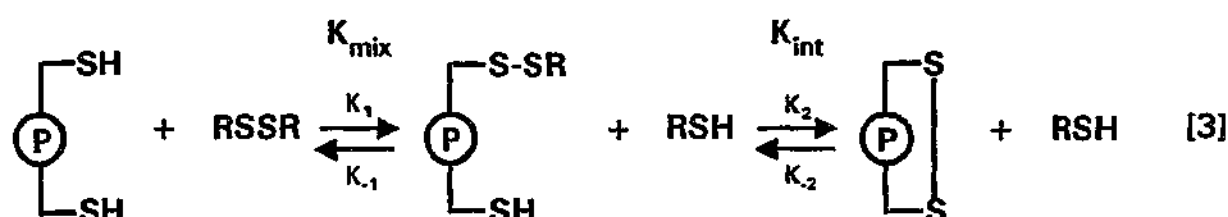
Protein and peptide thiol oxidation is thus a thiol/disulphide exchange as the oxidation equivalents are transferred from the low molecular weight disulphide to the protein or peptide. Exchange occurs through the direct attack of a nucleophilic thiolate anion on one of the sulphurs of the disulphide bond as shown in equation 2 (Gilbert, 1994). A symmetrical transition state is created in which bond formation to the nucleophilic sulphur ( $\text{R}_n\text{S}^-$ ) and the leaving sulphur ( $\text{R}_l\text{S}^-$ ) is equal and there is a small amount of negative charge on the central sulphur ( $\text{R}_c\text{S}^-$ ).



The rate of reaction increases with increasing pH until the attacking thiol is mostly in the thiolate form.

Folding of cysteine containing peptides or proteins is best carried out in a redox buffer, which typically consists of a mixture of a low molecular weight disulphide and its thiol. The reaction, which is reversible, produces a

mixed disulphide intermediate in the first instance, and then the fully oxidised product by intramolecular rearrangement. The disulphide supplies the oxidising equivalents for the protein or peptide disulphide formation, while the thiol acts to reduce trapped disulphide intermediates and to catalyse thiol/disulphide rearrangements. The reaction is shown in equation 3 (Gilbert, 1994).



The equilibrium constant for thiol exchange ( $K_{\text{mix}}$ ), in which the mixed disulphide intermediate is produced, is approximately 1 in a glutathione buffer, whereas the equilibrium constant for protein or peptide disulphide formation ( $K_{\text{int}}$ ), which occurs through intramolecular displacement, depends on the energetics of structural changes. Once mixed intermediates are formed, the reaction can proceed in either direction depending on the composition of the redox buffer. Provided there is no accumulation of mixed disulphide species, the overall oxidation equilibrium constant ( $K_{\text{ox}}$ ) in a redox buffer consisting of the monothiol (RSH) and its disulphide (RSSR) is given as follows (Gilbert, 1994):

$$K_{\text{ox}} = \frac{[\text{P(SS)}]_{\text{eq}} [\text{RSH}]_{\text{eq}}^2}{[\text{P(SH)}_2]_{\text{eq}} [\text{RSSR}]_{\text{eq}}} = K_{\text{mix}} K_{\text{int}} = \frac{K_1 K_2}{K_{-1} K_{-2}}$$

These are the basic chemical principals which govern disulphide formation in peptides and proteins. The process of folding itself is, however, a far more complex sequence of reactions, as will be illustrated in the next section.

### 1.3.3 The Folding Pathway

Several models exist to describe the folding process. An attractive model that is consistent with much of the experimental results to date is based on a diffusion-collision model. In this model (Wright *et al*, 1988), the folding process appears to commence as a result of short range interactions between the amino acids of the polypeptide chain that consequently lead to the formation of local secondary structure or hydrophobic clusters. Their formation is rapid and they exist in dynamic exchange with unfolded states. The types of local structures formed are very much determined by the local amino acid sequences. Although these initial structures are barely stable, they act as initiation sites for further co-operative growth and direct the folding pathway by reducing conformational space and permitting specific subsequent interactions to occur between more distant amino acids. Some of these interactions are more stabilising than others.

As folding continues, local elements diffuse and coalesce, forming highly localised structures that are still fairly unstable and in dynamic exchange with unfolded states. Eventually, the first observable kinetic intermediates are created as the polypeptide chain rapidly collapses to a globular state that consists of considerable secondary and supersecondary structure. This is followed by rearrangement into a compact, folded non-native structure, which is finally converted in the rate determining step into the fully folded native structure.

Formation, reduction and rearrangement of disulphide bonds of proteins and peptides containing multiple cysteines are integral to the folding process and are often used to monitor folding. Two proteins that have been

extensively studied are bovine pancreatic trypsin inhibitor (BPTI), which has three disulphide bonds, and bovine pancreatic ribonuclease A (RNase A), which has four disulphide bonds. The main reason for their study is due to their spontaneous production of the native forms in high yields, which is not typical for most proteins.

The complexities of the folding pathways involving disulphide formation are indicated in Figures 1.3.1 and 1.3.2 for BPTI and RNase A, respectively. Although there are several differences in their folding pathways, there are a few similarities (Gilbert, 1994). Firstly, both pathways are complex and produce a large number of intermediate species, yet considerably fewer than expected if the folding process is a random conformational search. At the 1- and 2-disulphide stages, there is a predominance of species with native disulphide bonds. Intermediates with 3- and 4-disulphide bonds form faster than the native structure, indicating that rearrangement of the 3- and 4-disulphide species to form the native structure is rate limiting.

Other proteins containing multiple cysteines have varying pathways and complications (Gilbert, 1994). For example, the folding of lysozyme produces a significantly lower yield of correctly folded product. The decrease in yield has been attributed to the capacity of lysozyme to form aggregates due to non-specific hydrophobic interactions. Aggregation inhibits the folding process and can be reduced by decreasing the concentration of original unfolded lysozyme or the use of agents such as non-ionic surfactants or urea that prevent hydrophobic interactions.

Cysteine-containing proteins and peptides tend to fold at a slower rate and require higher amounts of reductant to assist folding than those that are free of cysteine residues (Gilbert, 1994). This has been attributed to kinetic traps involving formation of native-like structures without the disulphide bonds, stable structures with non-native disulphides or the formation of unproductive intermediates that are too oxidised to fold correctly. Escapes

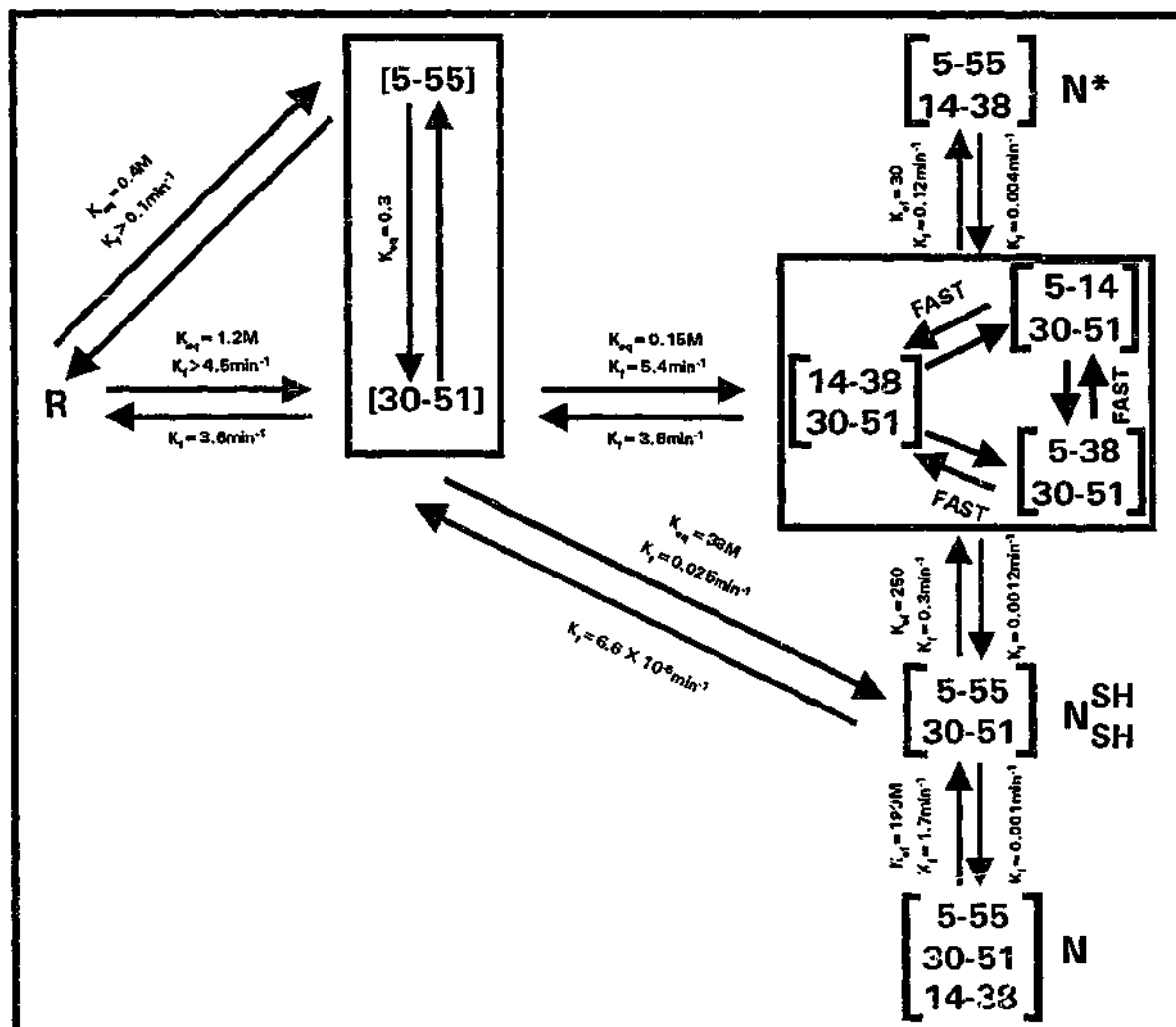


Figure 1.3.1. Oxidative folding mechanism for BPTI (Gilbert, 1994). R represents the fully reduced protein and the N variants, the intermediates. Intermediates species are indicated by square brackets within which identities of the disulphide bonds formed at that stage are given. The boxed intermediates are in equilibrium under most folding conditions.

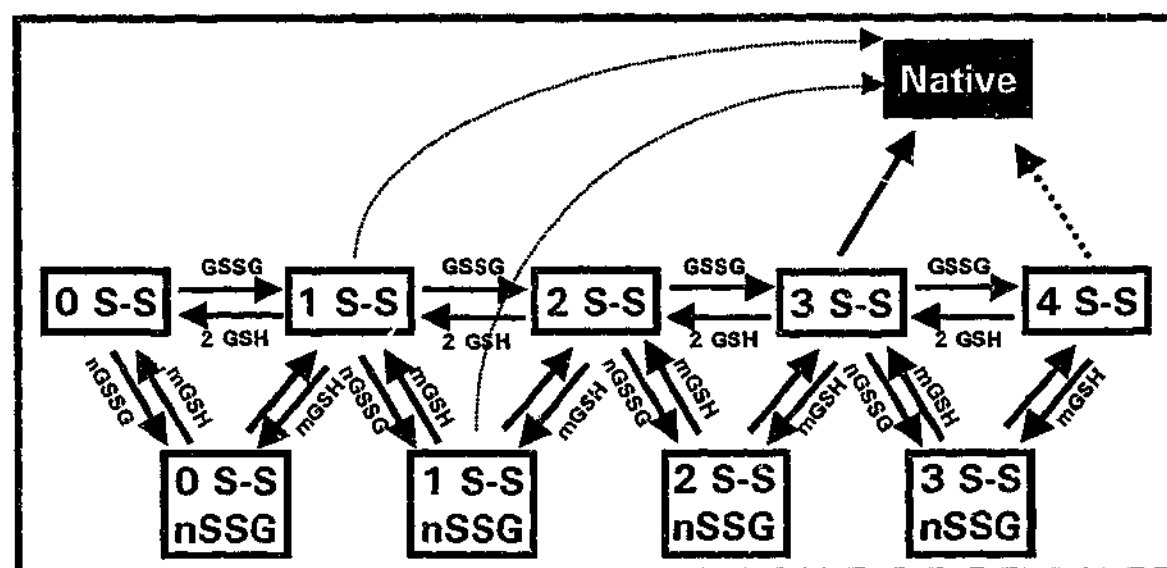


Figure 1.3.2. Oxidative folding mechanism for RNase A (Gilbert, 1994). Intermediates with 1-, 2-, 3- and 4- disulphide bonds are shown connected. Intermediates that contain n intermolecular mixed disulphides formed with the glutathione redox buffer are represented along the bottom line. Each set of intermediates represents a large number c? molecular species. The m and n letters indicate that multiple molecules of GSH and GSSH may be required to produce specific intermediates represented as collections.

from these states do exist, but generally require a backward step before continuation along a more correct pathway.

Another factor that alters the folding pathway is oxidation conditions (Gilbert, 1994). The most commonly used oxidant for folding of cysteine rich proteins and peptides is provided by a glutathione redox buffer. The use of this buffer system is generally successful in producing maximum quantities of native material, although the optimum concentrations of the glutathione and its disulphide vary between proteins. Air is another oxidant that may be utilised. However, it is difficult to control and is known to induce oxidation of certain side chains.

To date, much of the detail of folding that has been studied has been carried out *in vitro*. Such studies have been conducted on one member of the inhibitor cystine knot family, carboxypeptidase potato inhibitor (CPI) and it is reasonable to expect that there may be folding similarities between members of the same structural family. For this reason, the main findings of these studies are worthy of some discussion.

#### **1.3.4 Folding of Carboxypeptidase Potato Inhibitor**

The folding of recombinant carboxypeptidase potato inhibitor (CPI) was monitored by trapping intermediates of stop/go folding experiments with both acid and iodoacetate (Chang *et al*, 1994). Intermediates were then isolated and analysed by HPLC. The folding of CPI is found to proceed spontaneously *in vitro* and appears to occur in two stages. The first stage involves the formation of non-specific disulphide bonds through a sequential, heterogeneous dynamic flow, commencing with reduced CPI to 1-disulphide species, to 2-disulphide species and, finally, to scrambled 3-disulphide species. The flow is irreversible and the composition of these intermediates is not affected by acceleration of the flow, or by denaturant, which would indicate that non-covalent interactions do not guide the first phase of the

folding pathway. In the second stage, which is the rate-limiting step, the disulphides reshuffle to produce the native structure.

The efficiency of the separate stages may be controlled by altering the redox buffer conditions. Formation of disulphides is enhanced by cystine or oxidised glutathione. For instance, the presence of 2 mM of cystine alone incurs formation of greater than 98% scrambled intermediates after 1 minute. Reshuffling, however, requires free thiols such as cysteines, reduced glutathione or  $\beta$ -mercaptoethanol. The addition of denaturant such as 5M GdmCl or 8 M urea disrupts the reshuffling process and has no effect on the formation of early intermediates.

Knowledge of the folding of CPI could be useful to obtaining native kalata fragments. While folding is not a major part of the work for this thesis, fragments of kalata will be examined and their structure may lead to information on the folding and activity of kalata particularly, if the topology of the fragment structure deviates from the native fold. Similarly, isomers of kalata fragments could also provide information on the structure-activity relationship. Work has been attempted on another member of the inhibitor knot group,  $\omega$ -conotoxin GVIA. A brief mention of isomers in general and the results of this work indicate the difficulties involved in obtaining structural information on non-native structures.

### **1.3.5 Isomers of $\omega$ -Conotoxin GVIA**

Isomers of peptides with non-native disulphide bonds are less stable and considerably less active than the native peptide (Gilbert, 1994). They typically display conformational isomerism and it is therefore difficult to determine the 3D fold. Consequently, little is known of the structural differences between peptides with native and non-native disulphides and how this relates to activity. The results of the studies on  $\omega$ -conotoxin GVIA (Pallagay *et al*, 1993) are in line with these general observations.



Several isomers with non-native disulphides of  $\omega$ -conotoxin GVIA were identified and their structures appeared less well ordered. As expected, the main isomer retained only 1.3 % of the activity of the native and clearly lacked structural features of the native, showing significant conformational heterogeneity which may be due to the *cis-trans* isomerism that was observed. Consequently, the 3D folds could not be ascertained.

Characterisation of tertiary structure of kalata fragments with non-native disulphides would be of interest due to the difficulties involved in the analysis of isomers, as well as in the provision of structural data on non-native species. This serves as the impetus for studies on non-native fragments of kalata in this thesis.

*In vitro* studies on folding have shown that the process is elaborate, time-consuming and intricate. In consideration of living organisms that would have to contend with such problems as production of inactive isomers and aggregation, it might be expected, and indeed has been found, that conditions and processes are quite different *in vivo*. To complete this section on folding, a few words on *in vivo* folding are appropriate.

#### 1.3.6 Folding *In Vivo*

Folding *in vivo* is very much faster than *in vitro* (Gilbert, 1994). Disulphide containing proteins are found in most secretory proteins, and their folding, which involves the formation of disulphide bonds, occurs in the endoplasmic reticulum where an oxidising environment is provided mainly by glutathione and its disulphide. The first major difference between folding *in vivo* and *in vitro* is that the folding process *in vivo* begins before synthesis of the protein is complete. Secondly, folding *in vivo* appears to be assisted by other biochemical substances such as chaperones and foldases. Chaperones act to prevent non-specific aggregation by binding specifically to the unfolded protein, which is later released. Foldases behave like catalysts and hasten thiol/disulphide oxidation, reduction and rearrangement. Another contributing

feature of many secreted proteins is the presence of pre or pre-pro sequences that are part of newly synthesised proteins. These additional sequences are involved in folding and could alter the final structure produced. At the completion of folding, these sequences are removed.

Science is far from understanding the complete folding pathway within living organisms and certainly a good understanding of the folding process *in vitro* is extremely useful particularly from the point of view of larger scale production of therapeutically important proteins and peptides. In the event that the activity of either of the macrocyclic peptides shows significant potential for clinical application, knowledge of its folding pathway would be essential. Thus, information on fragments and isomers could enlighten the understanding of the delicate relationship between structure and activity.

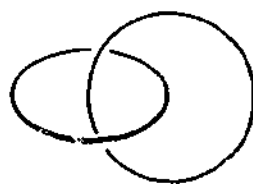
## 1.4 Molecular Topology in Proteins and Polypeptides

Topology is a branch of mathematics that studies structural properties that remain invariant under continuous deformation (Benham and Jafri, 1993). Topological concepts are consequently applicable to chemical structures where the order of atoms, covalent bonds and configuration at rigid centres define 3D spatial arrangement (Schill, 1971; Mao, 1993). In molecular topology, bonds are thought of as infinitely flexible but mathematically incapable of breaking (Mao, 1993).

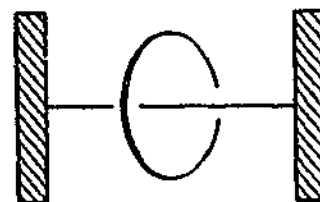
Of special interest to chemists in recent times is the discovery and synthesis of more complex topological structures such as knots, mechanically linked chains known as catenanes and structures that resemble a wheel on an axle known as rotaxanes (Schill, 1971). Simple examples to illustrate these structures are given in Figure 1.4.1.



Trefoil  
Knot



Two Ring  
Catenane



Rotaxane

Figure 1.4.1. Simple examples of complex topological structures (Schill, 1971).

These structures can be considered topological isomers as interconversion between them can only occur through breaking and reformation of chemical bonds.

Structural motifs consisting of knots and links are abundant in polynucleotides. However, in protein and polypeptide chemistry, topological stereoisomerism is uncommon, despite the myriad of interesting entanglements that exist. An explanation for the absence of topology in proteins and polypeptides is given by Mao (Mao, 1989). According to Mao, the main reason preventing the formation of the large loops that are involved

in knots and links is the occurrence of the extensive side chain interactions, which create secondary structures that are connected by short loops and held in place by hydrogen bonds. Nonetheless, many proteins and polypeptides have been named knots when in fact most folds are topologically planar and achiral (Liang and Mislow, 1996).

In this thesis, molecular topology is a significant consideration for cyclic peptides with three disulphide connections such as kalata, which displays topological stereoisomerism (Liang and Mislow, 1995). To demonstrate the application of, and interesting features in molecular topology, a discussion of stereoisomerism in proteins and polypeptides will be presented. It is important to realise the role played by disulphide bonds in topology in proteins and polypeptides. Not only are they key elements to formation of topological features but the pattern of bonding is a determining factor in the presence or absence of topology. The following presentation utilises two-dimensional graphs which are often used in molecular topology to highlight main features in relation to the constitutional formulae.

#### 1.4.1 Achiral, Planar Proteins

The first example, Ribonuclease, is characteristic of the large majority of proteins and polypeptides in that it is planar and achiral in a topological sense. Ribonuclease has four disulphides bonds. A diagrammatic representation, known as a condensed or covalent structure graph, which shows the constitutional formula of ribonuclease, is shown below in Figure 1.4.2. (Liang and Mislow, 1994a).

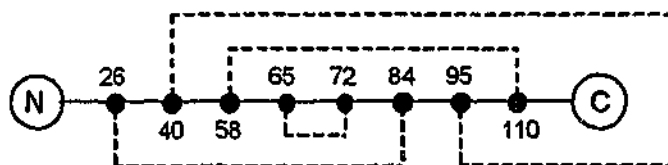


Figure 1.4.2. Covalent structure graph of Ribonuclease. The vertical line represents the polypeptide chain from the N to the C terminal. Vertices (solid circles) correspond to the  $\alpha$ -carbon of the disulphide cysteine residues and edges (dashed lines) correspond to covalent linkages between two cysteine  $\alpha$ -carbons (Liang and Mislow, 1994a).

This graph consists of vertices (solid circles), which correspond to the  $\alpha$ -carbons of the disulphide cysteine residues, and of edges (dashed lines), which correspond to covalent linkages between two cysteine  $\alpha$ -carbons. The graph can be embedded in the plane without crossing of any edges, indicating that ribonuclease is topologically planar. For planar graphs, there is only one topology (Mao, 1989). A consequence of this means that ribonuclease can be considered topologically achiral whereby its structure can be converted onto other 3D arrangements by continuous deformation, unlike in the case of topologically chiral proteins.

#### 1.4.2 Simple, Chiral, Nonplanar Proteins

The first native proteins shown to be nonplanar were two mammalian active neurotoxins from scorpions (Mao, 1989). These were variant 3 toxin from a North American scorpion species *Centruroides sculpturatus* Ewing and toxin II from a North African scorpion species *Androctopus australis* Hector. Both native toxins have four disulphide bonds. The covalent structure graph for variant 3 toxin is shown in Figure 1.4.3 (a) (Mao, 1989).

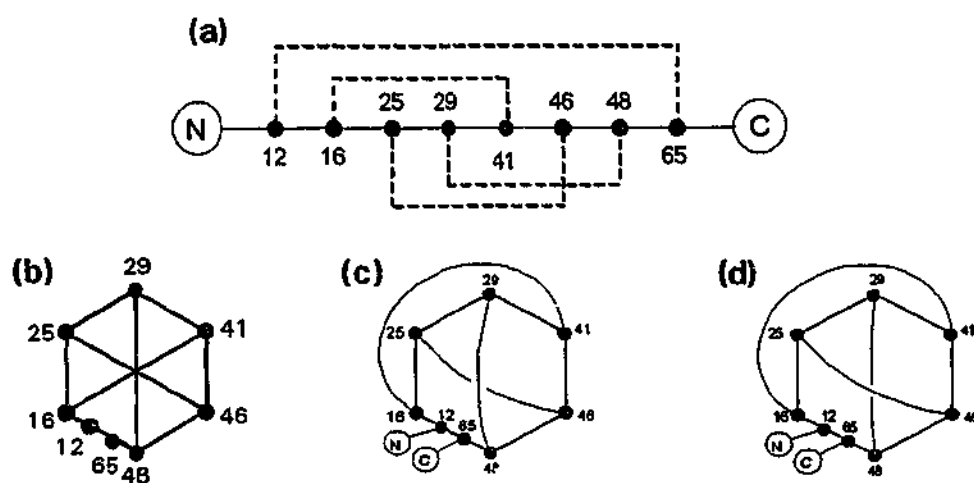


Figure 1.4.3. Graphs for Variant 3 scorpion toxin from *Centruroides sculpturatus* Ewing (Mao, 1989 and 1993). (a) Covalent structure graph illustrating nonplanarity. (b) Complete bipartite  $K_{3,3}$  graph illustrating chirality. (c) Reduced graph illustrating native D topology. (d) Reduced graph illustrating L topology.

Since the covalent structure graph cannot be drawn without crossing of edges, the graph and therefore the structure is designated topologically nonplanar. Another feature of these proteins are that the graphs contain  $K_{3,3}$  subgraphs. A  $K_{3,3}$  subgraph is defined to consist of two disjoint sets of three vertices each, with each vertex of one set adjacent to all three of the other and the presence of a  $K_{3,3}$  subgraph confers nonplanarity (Liang and Mislow, 1994a; Walba, 1985). The complete bipartite  $K_{3,3}$  graph for the variant 3 toxin chain is given in Figure 1.4.3 (b) (Mao, 1989). Since the vertices of scorpion toxin graphs are all labelled differently as they represent chemically different entities, the  $K_{3,3}$  subgraph cannot be converted into its topological enantiomorph by continuous deformation. These scorpion toxins are consequently topologically chiral. Furthermore, for a graph such as the scorpion toxins, topological stereoisomers fall into two classes, **D** and **L** (Mao, 1993). The **D** and **L** topological stereoisomers are displayed in a reduced graph form in Figures 1.4.3 (c) and (d), respectively. It is the crossing of the edges 29,48 and 25,46 which defines the configuration. In the case of the native scorpion toxins, the topology is **D**.

Another nonplanar, chiral protein for the same reasons and arguments as for the scorpion toxins is the light chain of the quinoprotein, methyl-amine dehydrogenase (MADH) (Mao, 1993). This protein has six disulphides. Schematic drawings of the molecular topology and the covalent structure graph are shown in Figures 1.4.4 (a) and (b), respectively.

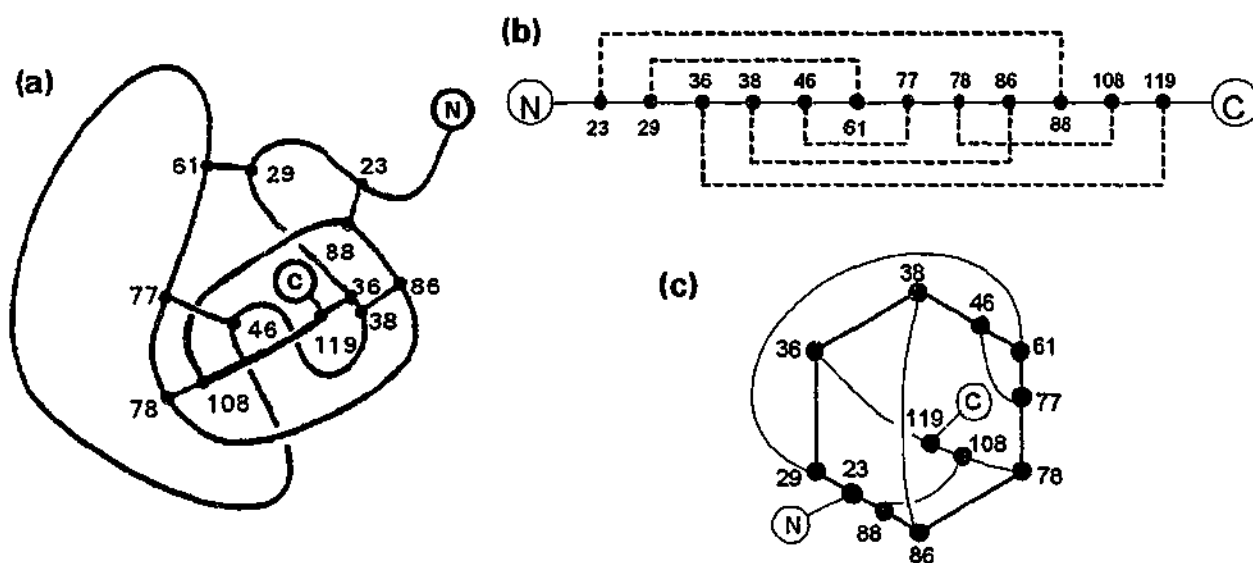
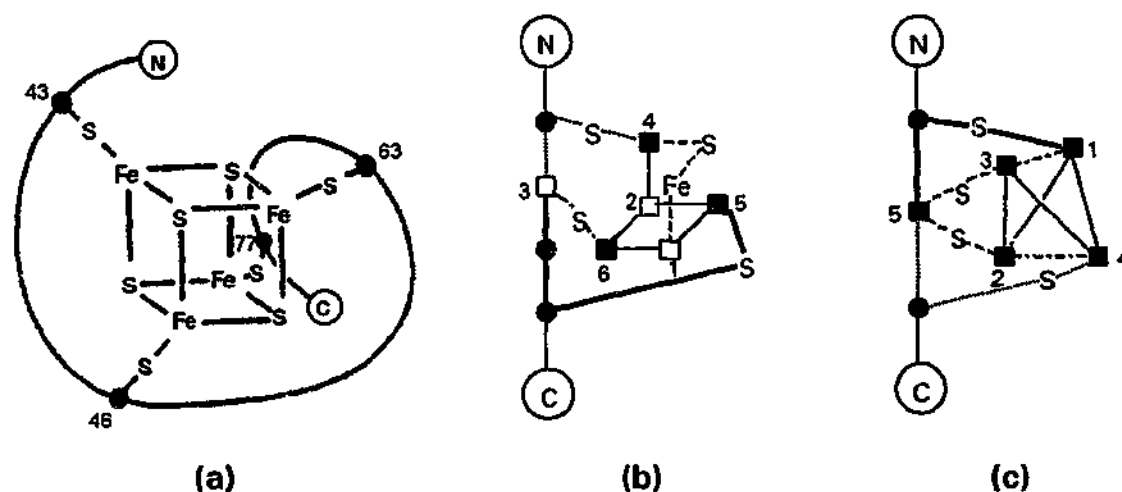


Figure 1.4.4. Representations for methyl-amine dehydrogenase (Mao, 1993). (a) Schematic drawing of the molecular topology. (b) Covalent structure graph. (c) Reduced graph illustrating native **D** topology.

Similar to the scorpion toxins, there are two possible topologies for this protein, D and L and like the scorpion toxins, the topology is D as shown in the reduced graph of Figure 1.4.4 (c). MADH is one of a family of such proteins, all of which display nonplanar, chiral topology (Liang and Mislow, 1994a).

Nonplanar and chiral topology are also found for several proteins without disulphide bonds that have cofactors (Liang and Mislow, 1994a). Given that there are a many proteins, which contain covalently bound cofactors that are integral to the function of the protein, the possibility of nonplanar, chiral topology is greater than originally thought. Examples of such proteins include iron-sulphur proteins such as *Chromatium* high potential iron protein (HiPIP), HiPIP from *Rhodocyclus tenuis* and flavoprotein trimethylamine dehydrogenase. Iron-sulphur clusters of these proteins covalently bind to the polypeptide chain resulting in a topological structure with both  $K_{3,3}$  and  $K_5$  subgraphs. This is the first known case of a  $K_5$  subgraph of a protein. A  $K_5$  subgraph is defined as consisting of five vertices that are adjacent to one another (Walba, 1985). A schematic drawing of the molecular topology and contractions to  $K_{3,3}$  and  $K_5$  subgraphs for *Chromatium* HiPIP, are shown in Figure 1.4.5 (Liang and Mislow, 1994a).



**Figure 1.4.5.** Representations for Fe<sub>4</sub>S<sub>4</sub> cluster containing-iron sulphur protein, *Chromatium* high potential iron protein (Liang and Mislow, 1994a). (a) Schematic drawing of molecular topology. (b)  $K_{3,3}$  subgraph. The two disjoint sets of vertices of the  $K_{3,3}$  subgraph are shown in open and solid squares. The heavy and dotted type lines represent atom chains which form the remaining four edges. (c)  $K_5$  subgraph. Vertices are shown in solid squares. The heavy and dotted type lines represent atom chains which form the remaining four edges.

The final point concerning the scorpion toxins, the quinoproteins and the cofactor proteins is that although they are all topologically nonplanar and chiral, they are defined as topologically simple structures in that they do not consist of knots or links or other complexities.

### 1.4.3 Knots and Links in Proteins

Knots and links are more complex structures that are inherently topologically nonplanar, chiral objects (Liang and Mislow, 1994a). Trefoil knots and catenated links have been identified for several proteins (Liang and Mislow, 1994b; Liang and Mislow, 1995) where the combination of disulphide crosslinks and covalently bound cofactors appears to be essential in the formation of more complex topology.

One of the first described topological complexities in proteins and polypeptides is given for MADH quinoproteins (Liang and Mislow, 1994a), which in addition to six disulphide bonds, have covalently bound cofactors that crosslink amino acid residues in the chain. The covalent structure graph and a schematic drawing of the molecular topology of MADH from *Thiobacillus versutus* (TV-MADH) (Liang and Mislow, 1994a) are given in Figures 1.4.6 (a) and (b), respectively. The consideration of the cofactor link together with the disulphide bonds leads to the catenated structure shown in Figure 1.4.6 (c). Similar considerations apply to form other complex topological structures.

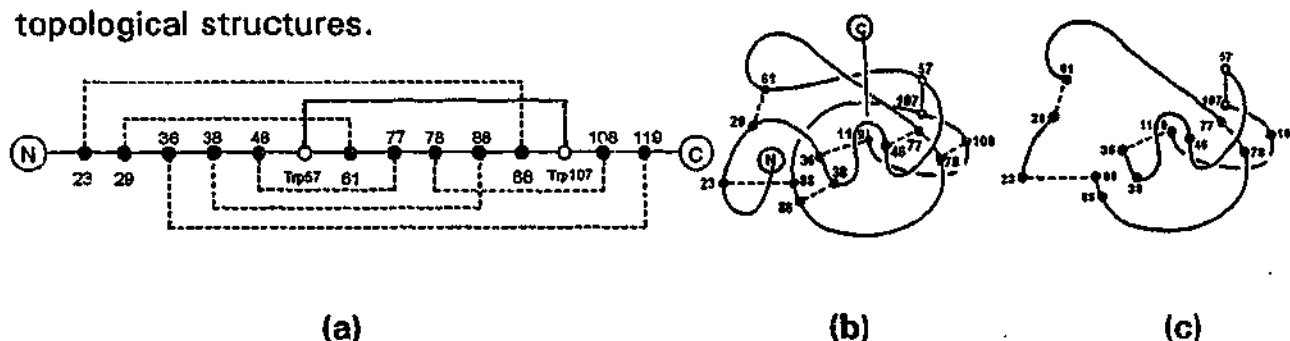


Figure 1.4.6. Representations for MADH from *Thiobacillus versutus*. (a) Covalent structure graph. The covalently bound cofactors are indicated by a heavy line and the  $\alpha$ -carbons of the residues connected by the cofactors are indicated by open circles (Liang and Mislow, 1994a; Liang and Mislow, 1995). (b) Schematic drawing of molecular topology (Liang and Mislow, 1994a). (c) The topological catenated structure derived from the schematic drawing (Liang and Mislow, 1994a).



Several metalloproteins have also been determined to contain trefoil knots (Liang and Mislow, 1994b; Liang and Mislow 1995). The multicopper enzyme ascorbate oxidase (AOase) has two copper atoms bonded to nitrogen atoms and two disulphide bonds that form the trefoil knot with L chirality as shown in Figure 1.4.7.

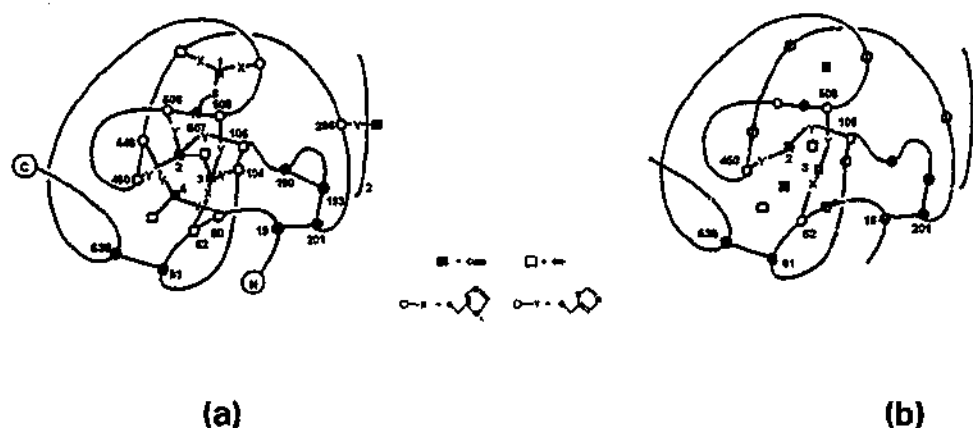


Figure 1.4.7. (a) Schematic drawing of molecular topology of dimer subunit of ascorbate oxidase. The  $\alpha$ -carbons of cysteine residues are represented by filled circles and those of histidine residues by unfilled circles. (b) The topological trefoil structure derived from the schematic drawing (Liang and Mislow, 1994b).

Similarly, the C-lobe of human lactoferrin (hLF) has an iron atom bonded to oxygen atoms and one disulphide bond that form the trefoil knot with D chirality.

These three proteins serve as examples of some of the known knots and links in protein and polypeptide chemistry and are used in this thesis for comparisons to the topological features in members of the cystine knot family, ironically, none of which form knots (Liang and Mislow, 1996). However, some members of the cystine knot family are topologically complex.

#### 1.4.4 Molecular Topology in the Cystine Knot Family

The use of the word knot has been technically incorrectly applied to many proteins and polypeptides (Liang and Mislow, 1996). With regard to the cystine knot family which includes the growth factors, potato inhibitor of

carboxypeptidase A, trypsin inhibitors CMTI-I and EETI-II, neurotoxin  $\omega$ -conotoxin GVIA and inhibitor kalata-B1, only two of the members have interesting topological features (Liang and Mislow, 1994b; Liang and Mislow, 1995).

The first interesting member is the growth factor, human chorionic gonadotrophin (hCG). This protein is a heterodimer of which the  $\alpha$ -subunit is planar. However, the  $\beta$ -subunit has two catenated substructures as depicted in Figure 1.4.8.

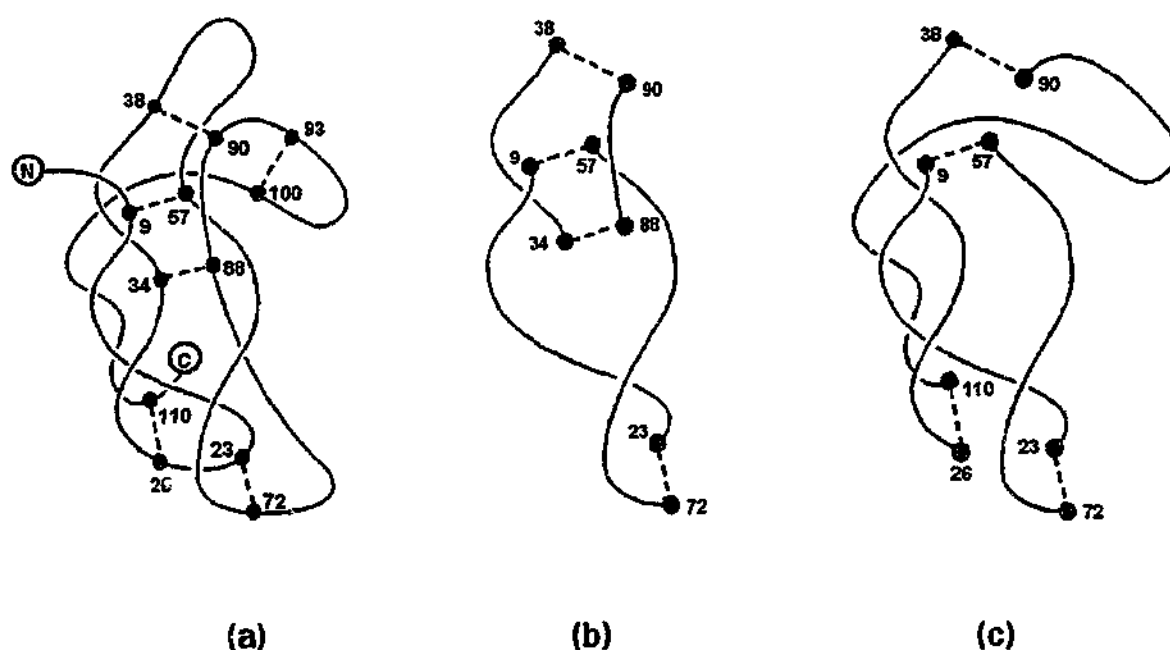


Figure 1.4.8. (a) Schematic drawing of molecular topology of  $\beta$ -subunit of human chorionic gonadotrophin. (b) A topological catenated structure derived from the schematic drawing. (c) A second topological catenated structure derived from the schematic drawing (Liang and Mislow, 1995).

The hCG protein is the first known topological link which consists of entirely amino acid residues.

The second interesting member is kalata-B1. Due to the cyclic nature of kalata-B1 in combination with the pattern of three disulphide bonds, the kalata-B1 structure contains a  $K_{3,3}$  subgraph and is consequently topologically nonplanar. Kalata-B1 is the simplest polypeptide known with topological chirality. It could be proposed that other cyclic peptides with identical

disulphide bonding to kalata-B1 are likely to be similarly topologically nonplanar, chiral and simple.

The remaining members of the cystine knot family are topologically planar and achiral. Computer graphics are useful for the display of topological features. The planar, achiral topology of the potato inhibitor of carboxypeptidase A has been demonstrated using animated graphical images known as kinemages (Benham and Jafri, 1993).

#### 1.4.5 Kinemages

Kinemage is an abbreviation for kinetic image, which is used to display and highlight 3D information of proteins and polypeptides often through basic animations. The software (Richardson and Richardson, 1992) used to produce and illustrate kinemages, also allows operations on images such as rotation, distance and dihedral angle measurement, and zooming to name a few. A kinemage, created for potato inhibitor of carboxypeptidase A (Benham and Jafri, 1993), shows the disulphide bond pattern and apparent loop threading formed from the disulphides. An associated animation shows deformation of the structure to produce a planar structure. The potato inhibitor of carboxypeptidase A structure is essentially unknotted indicating the absence of chiral and complex topology.

Animated kinemages have also been created for one of the scorpion neurotoxins and MADH discussed in section 1.4.2 (Mao, 1993). The animation shows the transformation of the 3D structure to the 2D reduced graph representation that demonstrates the nonplanarity and chirality of the structure. The backbone and disulphides are moved, bent and stretched though not passed through each other in the process.

These examples illustrate the value of kinemages in communicating 3D information. As such, kinemages will be used in the thesis to illustrate structural and topological features of significance to the study of the cystine knot family.

The consideration of topology in protein and polypeptide chemistry is relatively recent. To date, it is limited by the amount of 3D structural information available which albeit is expanding at an exponential rate. Consequently, it can be expected that the identification of more structures with topological stereoisomerism will appear, particularly with larger proteins. Topological features are of interest in the study of proteins and polypeptides due to the constraining effect they have on structure. With increased topological complexity, molecular rigidity is increased concomitant with a decrease in internal mobility (Liang and Mislow, 1995). The ramifications of this effect on the folding mechanism as well as activity are unknown and accordingly act as the impetus for the understanding of the relation between molecular topology, structural activity and the folding pathway.

**CHAPTER TWO**

**2D NUCLEAR MAGNETIC**

**RESONANCE TECHNIQUES**

**AND PEPTIDE STRUCTURE**

**DETERMINATION**

## 2.1 Introduction

With the development of 2D NMR spectroscopy, the possibility and ease of detailed structural analysis of molecules by NMR spectroscopy was greatly advanced. Not only was the full identification of structural moieties assisted, but also information with regard to proximities between non neighboring atoms could be extracted from one experiment. Coupling this distance data with molecular dynamical calculations allows 3D structures to be generated. This method of determining structures is a powerful tool in biotechnology, where knowledge of 3D topology is essential in the process of rational drug design. The work for this thesis is centered on the combination of 2D NMR spectroscopy and molecular dynamics to elucidate 3D structures of macrocyclic peptides. An overview of the specific use and application of the techniques, along with some basic concepts, is presented in this chapter.

## 2.2 2D NMR Spectroscopy

### 2.2.1 The 2D Experiment

The quintessential 2D NMR pulse sequence can be characterised by four stages: preparation, evolution, mixing and detection (Jeener and Alewaeters, 1971). During the preparation stage, the spin systems are permitted to achieve equilibrium. The evolution stage is the crux of the 2D experiment, during which magnetization induced during preparation evolves over a time period,  $t_1$ . This may be followed by a mixing stage, where nuclear magnetization is redistributed, so that correlations with other related spins can occur. Finally, the resulting free induction decay (FID), which contains important structural information, is recorded over the time period,  $t_2$ . Repetition of this sequence, in which  $t_1$  is successively incremented by increasing time periods, constitutes the 2D experiment. Fourier transformation with respect to  $t_1$  and  $t_2$  produces a two dimensional spectrum, with

frequency domains,  $F_1$  and  $F_2$  and peaks represented by contours. The diagonal contours of this spectrum correspond to the 1D spectrum and the off diagonal contours or cross peaks correspond to the interacting spins.

The basic 2D NMR experiment can be varied through the application of for instance, different pulses, decoupling or pulse field gradients during mixing, so that specific scalar and NOE interactions may be observed from the different pulse sequences. Over the years, a suite of standard 2D NMR experiments has been developed. Six of these 2D pulse sequences involving both homonuclear and heteronuclear experiments have been utilised in this thesis and these are schematically shown in Figure 2.1.

### 2.2.2 Homonuclear Pulse Sequences

The first and simplest of these pulse sequences to be used are the correlation spectroscopy (COSY) type sequences. COSY spectra give information on scalar couplings that are the result of neighbouring through bond connections of up to three covalent bonds. The identification of proton pairs in direct proximity is thus enabled. COSY spectra are recorded in phase sensitive mode, so the cross peaks exhibit positive and negative lobes. Variants of the COSY pulse sequence reduce the complexity of coupling and allow coupling constants to be measured. In general, COSY spectra are used to identify two to three bond connectivities and to determine chemical shifts and three bond coupling constants. Two types of COSY spectra were applied for these purposes.

The double quantum filtered (DQF) COSY pulse sequence (Rance *et al.*, 1983; Shaka and Freeman, 1983) consists of three  $90^\circ$  pulses. The first  $90^\circ$  pulse is followed by the evolution time. The second  $90^\circ$  pulse produces multiple quantum transitions. Double quantum coherence can exist in any coupled spin pair and is thus the coherence of interest. The third  $90^\circ$  pulse converts double quantum coherence in single quantum coherence and phase cycling is used to select the converted double quantum coherence without





unwanted signals. Consequently, the strong diagonal peaks are suppressed to some extent and therefore, cross peaks near to the diagonal are no longer obscured as they are in straight COSY spectra. Resolution is improved and line shapes are narrower. DQF-COSY spectra can be used to accurately determine  $^3J_{\text{NH-}\alpha\text{H}}$  coupling constants.

Exclusive spectroscopy (E-COSY) (Griesinger *et al.*, 1987) is another COSY variant, which filters coupling information, so that only the active couplings of the complex multiplet coupling are selected and recorded. This is particularly useful for determining  $^3J_{\alpha\text{H-}\beta\text{H}}$  coupling constants. However, there are a few problems with the COSY experiment.

One of these problems is that assignment of certain spin systems with long side chains such as arginine, lysine, leucine, isoleucine often cannot be completed, in particular in the 0-2 ppm spectral region, where there is more spectral overlap. Another problem frequently encountered is that some cross peaks are not observed. This applies to  $\text{H}_\gamma$  and  $\text{H}_\delta$  of arginine, leucine, lysine and proline, as these particular protons couple with several neighbouring protons and therefore, produce broad multiplets with low intensity. Moreover, the cross peaks of the COSY experiment are antiphase and this can result in poor sensitivity. Many of these problems are overcome in alternative 2D experiments, such as in total correlation spectroscopy (TOCSY).

In the TOCSY experiment (Braunschweiler and Ernst, 1983; Bax and Davies, 1985), the second  $90^\circ$  pulse of the COSY sequence is replaced by an MLEV-17 spin locking pulse. The MLEV-17 spin locking sequence is composed of a series of composite low pulses, which "spin lock" or align the magnetization along the y-axis. Under this condition, spins do not precess; they share the same energy levels and can therefore, exchange energy. Magnetization is propagated along the bonds within the same spin system of each individual amino acid system comprising the polypeptide. The extent of magnetization transfer along the bonds is determined by the duration of the mixing time, in other words, the application of the spin locking field.

At short mixing times of less than 20 ms, only directly coupled connectivities are observed in the spectrum. In this sense, the data obtained from the TOCSY spectra are akin to that of the COSY spectra, although coupling constants cannot be measured. At longer mixing times of between 65 to 120 ms, relayed, as well as, direct connectivities are observed, so that ultimately, the entire spin system of each individual amino acid is recorded and may be identified. In addition to the increased data that is obtained from TOCSY experiments, the cross peaks are resolved more clearly than the antiphase COSY cross peaks. Furthermore, TOCSY data acquisition is much less time consuming, as it requires minimal phase cycling. TOCSY data are very useful in the assignment of individual spin systems. To completely determine the residues in sequence, another 2D experiment, nuclear Overhauser and exchange spectroscopy (NOESY) was employed.

The NOESY experiment (Kumar *et al.*, 1980) takes advantage of an alternative relaxation pathway, in which excited nuclei separated in distance by less than a maximum of 5 Å may exchange their magnetization, due to interaction of their magnetic dipoles. This is known as the nuclear Overhauser effect or enhancement. The rate of magnetization transfer is inversely proportional to the sixth power of the distance between the nuclei and is given by the following equation:

$$\text{NOE} \propto 1/(r^6) \cdot f(\tau_c)$$

where,  $r$  is the distance between the nuclei and  $\tau_c$  is the correlation time, which is inversely related to the rate of molecular tumbling in solution. The actual pulse sequence is very similar to that of the DQF-COSY. However, during the mixing stage, when cross relaxation takes place, it is zero order coherence that is selected for phase cycling, resulting in the observation of an NOE cross peak in the 2D spectrum for every spatially interacting pair of nuclei. Hence, the NOESY data provide information on protons spatially connected, so that as well as intraresidue connectivities, sequential residues

may be identified. Furthermore, mid and long range interresidual connectivities in close proximity, due to the molecular fold, are observed. This information is pertinent to supplying distance restraints for structure calculations and it is the most important application of NOESY experiments.

However, there is a potential problem in NOESY experiments that requires attention, as it can lead to incorrect assignments from false NOE cross peaks. Another relaxation mechanism, known as spin diffusion (Neuhaus and Williamson, 1989), can arise from the diffusive transfer of magnetization through other neighbouring protons, leading to undesirable cross peaks, which can be confused with NOE cross peaks. The spin diffusion process is secondary to the NOE build up and occurs at longer mixing times, affecting larger molecules more profoundly. Spin diffusion can be avoided or reduced by using shorter mixing times, and NOESY experiments where mixing times of less than 250 ms are applied, in general, will not display significant levels of spin diffusion (Bradley *et al.*, 1990). A symptom of the occurrence of spin diffusion is the appearance of cross peaks between distant protons in the molecule. The mixing time at which NOE cross relaxation is most efficient and spin diffusion is ineffective, may be ascertained by profiling the NOE build up curve for a particular molecule (Kessler and Seip, 1994). Consideration for spin diffusion was given during structural analysis throughout this work. In addition to the use of these 2D homonuclear NMR techniques, heteronuclear pulse sequences were also utilised.

### 2.2.3 Heteronuclear Pulse Sequences

The heteronuclear multiple quantum correlation (HMQC) pulse sequence (Bax and Subramanian, 1986) and the HMQC-TOCSY (Bax and Lerner, 1986) pulse sequences assist in the assignment of protons and allow carbon-13 chemical shifts to be determined through the correlation of  $^{13}\text{C}$  nuclei with their directly bonded  $^1\text{H}$  nuclei. Due to the large chemical shift dispersion of  $^{13}\text{C}$  nuclei, data from heteronuclear pulse sequences will resolve most, if not

all, ambiguous proton assignments from homonuclear experiments. For instance, protons bound to the same carbon may be differentiated from those of vicinal neighbours as in the case of the  $\beta$  and  $\gamma$  protons of leucine which are often degenerate.

The HMQC pulse sequence accumulates data relatively quickly, as minimal phase cycling is required. Moreover, the HMQC pulse sequence is highly sensitive. The actual pulse sequence is simple, consisting of only four pulses and a GARP decoupling sequence during detection. Initially, a  $90^\circ$  pulse on the protons induces transverse magnetization, i.e. first order coherence, which is permitted to evolve for a defined period until the  $^{13}\text{C}$  coupled  $^1\text{H}$  vectors are  $180^\circ$  out of phase. A  $90^\circ$  pulse is applied on the coupled  $^{13}\text{C}$  nuclei, converting the transverse magnetization into zero and double quantum coherence, which evolves in  $t_1$ . During this time, through bond magnetization transfer occurs. Midway through  $t_1$ , a  $180^\circ$  pulse is applied to the protons. This interconverts zero and double quantum coherence. Finally, another  $90^\circ$  pulse is applied on the coupled  $^{13}\text{C}$  nuclei, producing single quantum proton coherences. After a delay, this magnetization is simultaneously decoupled and detected as a proton FID. The resulting spectrum consists of  $^1\text{H}$  and  $^{13}\text{C}$  axes and cross peaks due to directly bound  $^1\text{H} - ^{13}\text{C}$  pairs.

The HMQC-TOCSY experiment is similar to the HMQC experiment. However, a spin locking pulse sequence is included prior to detection. This allows for correlation of all the protons within the amino acid spin systems to take place. This experiment is considerably less sensitive than HMQC and TOCSY experiments. However,  $^{13}\text{C}$  chemical shifts may be unequivocally assigned. Critical to the success of these heteronuclear experiments is broadband decoupling, which improves the sensitivity and spectral appearance.

With nuclei such as  $^{13}\text{C}$ , which has a low natural abundance, it is desirable to remove coupling interactions with  $^1\text{H}$  nuclei to avoid splitting of signals and therefore, further reduced intensities that would otherwise occur.

This is achieved by broadband decoupling of the  $^1\text{H}$  spectral width. In this work, 2D heteronuclear pulse sequences incorporated a GARP-1 pulse sequence (Shaka *et al.*, 1985) during detection for this purpose. This pulse sequence consists of a series of composite pulses that eliminate  $z$  magnetization of the  $^1\text{H}$  nucleus. In addition to broadband decoupling, solvent suppression is an important consideration in optimising sensitivity and spectral appearance for all the pulse sequences.

#### 2.2.4 Solvent Suppression

Often in NMR spectroscopy, the solvent signal is considerably more intense than the signals of the sample. In this work, peptides were analysed in aqueous solutions, where the water signal was a dominating spectral feature. To improve the dynamic range and reduce problems due to baseline roll, baseline modulation and the broad  $F_1$  noise stripe at the position of the water signal, suppression of the water signal is imperative to obtaining satisfactory NMR data. A number of methods designed to reduce the water signal were employed.

A well known method, presaturation, was used in all 2D experiments at some stage and was the method of choice in COSY type pulse sequences, where sensitivity is compromised using other techniques. Solvent presaturation involves the application of a continuous, weak, radiofrequency pulse at the frequency of the water resonance. After presaturation for a length of time, the  $z$  component of magnetization is removed, so that the water signal is greatly reduced. However, presaturation is limited, as peptide signals such as the alpha protons close to the water signal, are frequently saturated and obliterated from view. Furthermore, NH resonances exchanging with the solvent are also not observed. These problems are particularly acute at higher presaturation power. Another technique, in which only peptide resonances are excited, was utilised.

There are a variety of versions of this type of off-resonance sequence. In this work, a 1-1 binomial pulse sequence (Plateau and Geuron, 1982; Hore, 1993) was used in TOCSY and NOESY experiments. The sequence consists of two pulses separated by a delay, such that a null occurs at the water resonance and excitation of water molecules does not take place. The 1-1 binomial pulse sequence reduced the effect of the water signal dramatically, without significant loss of peptide resonances. However, there is a drawback in that a residual water resonance is obtained and this is phase shifted by  $180^\circ$ , resulting in loss of peptide signals near to the water signal. This method was used in combination with either presaturation, or a more recent technique, which uses a pair of pulsed field gradients of opposite sign.

The pulse field gradients method operates by dephasing magnetization caused by an initial gradient pulse and then, refocusing magnetization using a second gradient pulse. Relative molecular diffusion rates govern the success of solvent suppression by this method, as peptide molecules diffuse much more slowly than the small water molecules. After the first gradient pulse, all coherences are dephased. The larger peptide molecules remain in the same magnetic field throughout the 2D NMR pulse sequence. However, the mobile water molecules diffuse, sampling different spatial volumes. The second field gradient pulse will only refocus magnetization resident in the same sample volume. Consequently, coherences due to solute species are rephased, while the water molecules, which do not have any net magnetization, are dephased even further. The particular pulse sequence, used in NOESY and TOCSY spectra was a modified version of water suppression by gradient tailored excitation (WATERGATE) (Piotto *et al.*, 1992). In essence, the sequence is a spin echo technique and was the most successful method for water suppression.

The combination of data from all the aforementioned pulse sequences was used to assign chemical shifts according to a general strategy described in the following section.

### 2.2.5 Assignment Strategy

The NMR spectra of polypeptides are highly characteristic due to the repetitive amino acid structure. For instance, proton resonances of the different functional groups are observed typically within specific spectral ranges. The  $^{13}\text{C}$  resonances of the various functional groups are even more static. These ranges are displayed in Figure 2.2 and serve as an initial guide in assignment. Nonetheless, the 1D spectra of even small polypeptides are complex and it is considerably difficult to assign the spectra without knowledge of the polypeptide sequence and also the relationships between the signals. Once the sequence is determined, using generally enzymatic degradation and mass spectrometry techniques, multidimensional NMR spectroscopy can be effectively applied to assign the spectra. In this work, the polypeptides analysed are of a sufficiently small size that their structures can be fully examined by 2D NMR spectroscopy, without information from higher dimensional NMR spectroscopy of labeled peptide material.

Well established resonance assignment strategies were utilised (Wüthrich, 1986; King and Mackay, 1996). The general strategy is in part based on the relatively small deviation of the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of the individual amino acid residues from predefined values. Proton chemical shifts of the twenty common amino acids have been established from NMR measurements in aqueous solution of linear, random coil tetrapeptides, G-Gly-Gly- $X_{xx}$ -Ala-OH, where  $X_{xx}$  is each of the 20 different amino acids (Wüthrich, 1986). Hence, spin systems or sets of protons characterise each amino acid residue. Unique proton spin systems exist for glycine, alanine, valine, leucine, isoleucine, threonine, lysine, arginine and proline. Similarly, typical  $^{13}\text{C}$  chemical shift values for the amino acids residues have been established from the tetrapeptides and the general study of protein  $^{13}\text{C}$  chemical shifts (Howarth and Lilley, 1978). These defined  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts were used initially, as a starting point, to compare and assign the 2D NMR data.

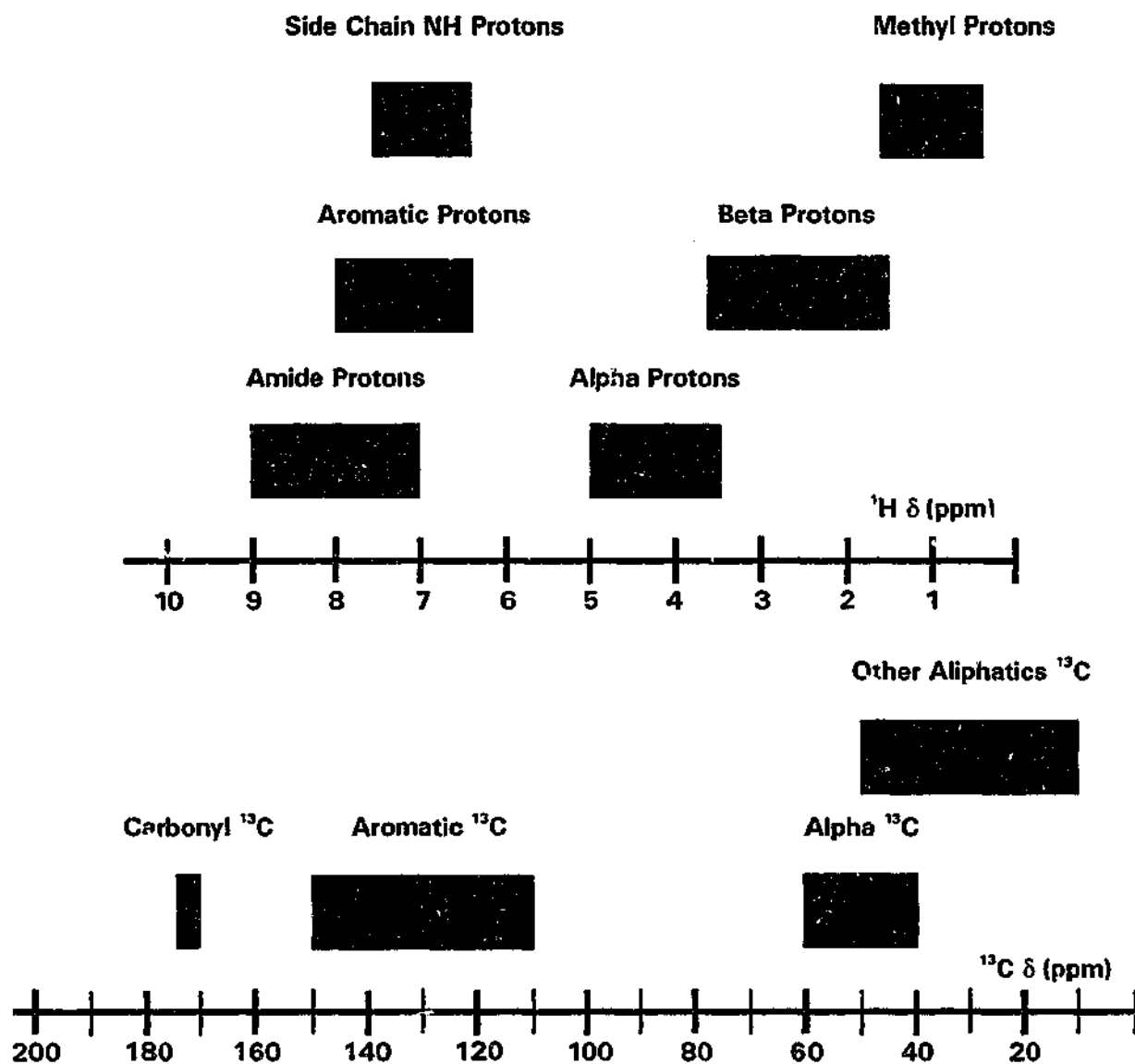


Figure 2.2.  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift ranges of amino acid residue functional groups.

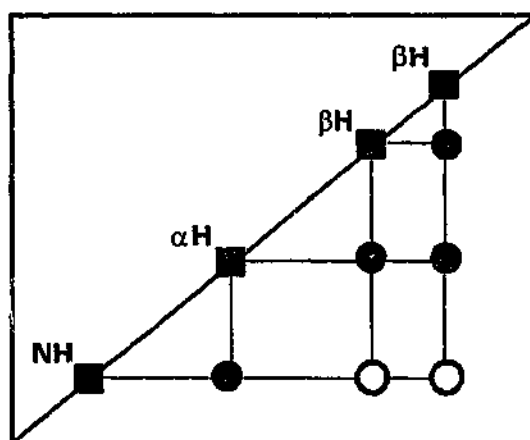


Figure 2.3. Diagram of COSY connectivities (filled circles) and longer mixing time, TOCSY connectivities (filled and unfilled circles) for a typical amino acid residue,  $\text{NH}-\alpha\text{CH}-\beta\text{CH}_2$ . The squares represent the diagonal cross peaks.



However, confirmed assignments are ultimately derived from unambiguous connectivity information from the 2D NMR spectra.

In the first step of assignment, the individual spin systems of the residues of the polypeptide in water are identified from scalar coupled experiments, such as COSY type and TOCSY spectra. All the cross peaks of these spectra relate to intraresidue connectivities and permit the entire network of connected protons to be established. TOCSY data allow identification of all the protons within a spin system, while COSY data distinguish those protons that are directly connected. The information that may be derived from COSY and TOCSY spectra is demonstrated in Figure 2.3. It is worthwhile recording TOCSY data with short, medium and long mixing times, so that the intensity variation of individual cross peaks can give additional and decisive information on direct and indirect connectivities. Furthermore, both TOCSY and COSY type experiments should also be performed on deuterated solutions, as alpha protons obscured by incomplete water suppression in water may be observed, and slowly exchanging amide protons in structured regions of the peptide may be determined.

In the second step of assignment, NOESY spectra of the aqueous peptide solution are used to correlate the amino acid spin systems to their position in the peptide sequence. The cross peaks of NOESY data relate spatially connected protons within 5 Å separation, thereby giving interresidual information. Medium to strong NOE cross peaks due to adjacent residues, involving the amide proton and the amide/alpha/beta protons of the previous residue, are used to assign spin systems specifically. It is important to be wary of the potential for confusion between sequential NOE cross peaks and other NOE cross peaks due to topological proximity. The sequential assignments can be validated when a continuous relationship between each amide proton and the previous alpha proton can be united in the fingerprint region. This is known as the NOESY walk, or circuit. This circuit is broken at proline residues, but proline delta protons may be used to complete the cycle. The NOESY data are particularly useful for identifying the non unique amino

acids. The analysis of these 2D homonuclear NMR spectra in conjunction allows for the majority of proton peptide resonances to be assigned. However, often there are ambiguous assignments remaining, especially for the side chains of longer chain amino acid residues.

In a third step of assignment, heteronuclear experiments may be used. These experiments are optional, as it is generally the less significant protons that at this point remain unassigned. However, these experiments also provide  $^{13}\text{C}$  assignments, which contain structural information. It is the larger chemical shift dispersion and the more stable nature of  $^{13}\text{C}$  chemical shifts of the component amino acids that enable both  $^1\text{H}$  and  $^{13}\text{C}$  data to be fully assigned. Having completely assigned the NMR data, information on the peptide secondary structure can be extracted.

## **2.3 Secondary Structure Determination by NMR Spectroscopy**

### **2.3.1 The Chemical Shift**

There are several parameters that may be measured from NMR data that provide insight into the secondary structure of peptides. Firstly, the chemical shift is considerably sensitive to the environment and it is well known that the alpha proton chemical shifts of polypeptides are strongly dependent on the local secondary structure. Deviations from their random coil values are directly related to this structure. For example, alpha protons that are part of an  $\alpha$ -helix conformation move upfield in the NMR spectrum, whilst alpha protons existing in extended or  $\beta$ -strand conformation move downfield.

A statistical technique, based on this movement in a collection of proteins, where the secondary structure is known, has been developed to rapidly and simply predict the location of secondary structure (Wishart *et al.*, 1992). The method generates CSI indices of +1, 0 or -1, for every alpha proton in a peptide, depending on the deviation from a defined range for alpha protons in unstructured regions. For a proton chemical shift lower than the

predefined range for the particular amino acid, a CSI value of "-1" is assigned, within the range, zero is given, and below the range "+1" is designated. Secondary structure is determined from the number and type of contiguous CSI indices along a stretch of sequence.

An  $\alpha$ -helix is indicated by a sequential grouping of four or more "+1" CSI values and  $\beta$ -strand is represented by a sequential grouping of three or more "-1" CSI indices. The termination points of secondary structure are shown by two consecutive zero CSI indices, or by the first appearance of a CSI value of the opposite sign. The occurrence of "-1, 1" or "1, -1" may be due to a reverse or  $\beta$ -turn. All other regions are considered as coil. The graphical representation used to illustrate CSI values of the component amino acids in a sequence is given in Figure 2.4. The method claims a 90 to 95% accuracy to predict common secondary structure from alpha protons. The predictive capability can be increased to greater than 92% accuracy when CSI values based on  $^{13}\text{C}$  chemical shifts are combined.

Similar methods have been developed using  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$  and  $^{13}\text{C}'$  chemical shifts (Wishart and Sykes, 1994). The  $^{13}\text{C}^\beta$  chemical shifts also move upfield for  $\alpha$ -helices and downfield for  $\beta$ -strands. However, the  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}'$  chemical shifts move in the opposite directions for  $\alpha$ -helices and for  $\beta$ -strands. In the case of  $^{13}\text{C}^\beta$  CSI values, four or more consecutive "+1" indices are required to indicate  $\beta$ -strand. Only  $\beta$ -strand can be predicted from  $^{13}\text{C}^\beta$  CSI values due to overlap problems, which result in unreliable predictive power for helices. CSI values based on  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}'$  chemical shifts can be used to preempt the presence of  $\alpha$ -helices with four or more consecutive "+1" CSI indices, and  $\beta$ -strands with three or more consecutive "-1" CSI values. Termination points are indicated as for alpha protons and the remainder of the sequence that precludes the presence of secondary structure is also considered as coil.

In addition to secondary structure determination,  $^{13}\text{C}$  data may also be used to ascertain the *cis/trans* conformation of proline residues (Kessler and Seip, 1994). The difference between the  $^{13}\text{C}^\beta$  and  $^{13}\text{C}^\gamma$  chemical shifts is

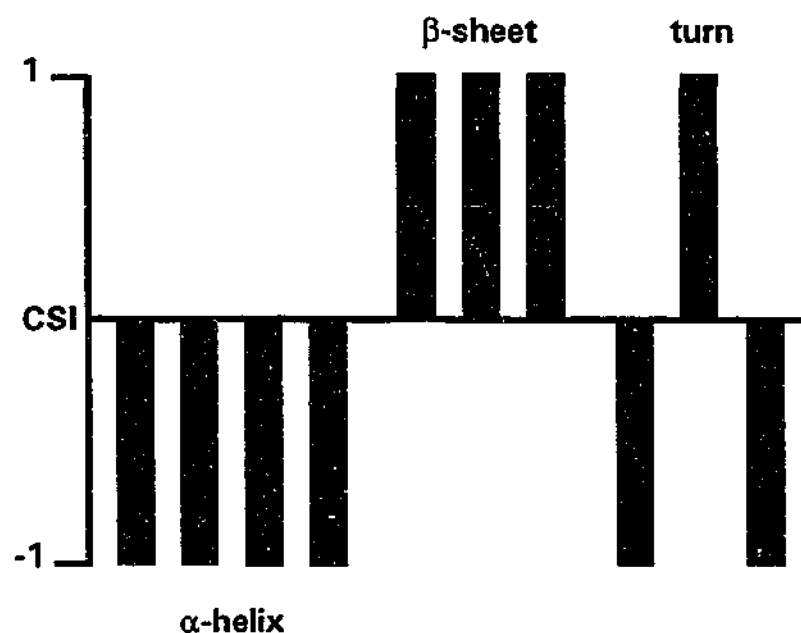


Figure 2.4. Graphical representation of the  $^1\text{H}$  CSI value patterns that indicate types of secondary structural elements. The vertical axis corresponds to the polypeptide sequence.

$\chi_1$	$60^\circ$	$180^\circ$	$-60^\circ$
$^3J_{\text{H}\alpha\text{-H}\beta 2}$ (Hz)	2.6 -5.1	2.6 -5.1	11.8 -14.0
$^3J_{\text{H}\alpha\text{-H}\beta 3}$ (Hz)	2.6 -5.1	11.8 -14.0	2.6 -5.1
NOE( $\text{H}^\alpha\text{-H}^{\beta 2}$ )	s	s	m - w
NOE( $\text{H}^\alpha\text{-H}^{\beta 3}$ )	s	w - m	s
NOE( $\text{H}^{\text{N}}\text{-H}^{\beta 2}$ )	w - m	s - m	s
NOE( $\text{H}^{\text{N}}\text{-H}^{\beta 3}$ )	s - m	s	w - m

Figure 2.5. The NMR criteria for stereo assignment of the three conformations involving the  $\text{C}\alpha\text{-C}\beta$  bond of amino acid residues in polypeptides. The letter "s" refers to strong intensity, "m" to medium intensity and "w" to weak intensity.

sensitive to the configuration of proline. For a proline involved in a *cis* arrangement, there is an increase in this chemical shift difference, with  $\Delta\delta_{\text{C}\beta\text{-C}\gamma}$  being around 10 ppm, whereas for a proline involved in a *trans* arrangement, this difference is lower, being around 5 ppm. Chemical shift data is thus extremely valuable for the prediction of secondary structure. Additional information can be obtained from coupling constants, as described below.

### 2.3.2 Coupling Constants

Spin-spin couplings between nuclei result in the appearance of dispersive cross peaks in COSY type spectra. The frequency separation between cross peaks is known as the coupling constant,  $J$ . The three bond coupling constant is related to dihedral angles in polypeptides according to the Karplus equation (Karplus, 1959):

$$^3J = A\cos^2\theta + B\cos\theta + C$$

where,  $\theta$  is the H-X-Y-H dihedral angle and A, B and C are variable values dependent on the type of nuclei and the environment. For NH- $\alpha$ H coupling constants, values of  $A = 6.4$ ,  $B = -1.4$  and  $C = 1.9$  are applicable (Pardi *et al.*, 1984), while for H $\alpha$ - $\beta$ H coupling constants, values of  $A = 9.5$ ,  $B = -1.6$  and  $C = 1.8$  are used (DeMarco *et al.*, 1978; Bystrov, 1976).

$^3J_{\text{NH-}\alpha\text{H}}$  coupling constants can be measured from DQF-COSY data (Marion and Wüthrich, 1983) to indicate  $\phi$  ( $\theta + 60^\circ$ ) angles. This angle is dependent on local secondary structure, so that typical  $^3J_{\text{NH-}\alpha\text{H}}$  values may be related to this structure. In general,  $^3J_{\text{NH-}\alpha\text{H}}$  coupling constants of less than 5 Hz are characteristic of  $\alpha$ -helix, and values above 8 Hz are found for random coil and  $\beta$ -strand conformations. Values of between 6 - 7 Hz are obtained for regions of helical conformations where averaging is taking place. Not only are  $^3J_{\text{NH-}\alpha\text{H}}$  coupling constants useful as a diagnostic tool in secondary structure

determination, but they provide a source of additional torsion angle restraints for 3D structure calculations, along with  $^3J_{\alpha\beta}$  coupling constants.

$^3J_{\alpha\beta}$  coupling constants are measured from E-COSY data (Griesinger *et al.*, 1987), and are related to the  $\chi_1$  angle in polypeptides. For the gauche-gauche conformation where  $\chi_1 = 60^\circ$ ,  $^3J_{\alpha\beta 2} = ^3J_{\alpha\beta 3} = 3.4$  Hz are found. For gauche-trans conformations,  $^3J_{\alpha\beta 2} = 3.4$  Hz and  $^3J_{\alpha\beta 3} = 12$  Hz or *vice versa* are observed for  $\chi_1 = 180^\circ$ , or  $-60^\circ$  (where H $\beta$ 2 and H $\beta$ 3 refers to the IUPAC-IUB conventions, 1970). The coupling constants collectively, with specific NOE intensity data, may be used to stereo specifically assign the beta protons for 3D structure calculations (Wagner *et al.*, 1987; Nagayama and Wüthrich, 1986). The criteria for stereo assignments are shown in Figure 2.5.

### 2.3.3 Exchanging Amide Protons

Amide protons, slowly exchanging with deuterium from  $^2\text{H}_2\text{O}$  solvent, are characteristic of the presence of hydrogen bonds that occur in secondary structural elements. Protons involved in secondary structure are in a protected environment, where the protons are significantly less exposed to solvent. Slowly exchanging amide protons are identified by recording a series of 1D and TOCSY spectra, at increasing intervals of time, upon dissolving peptide material in highly pure  $^2\text{H}_2\text{O}$  solvent. Those amide protons that do not partake in secondary structure formation, exchange rapidly with  $^2\text{H}_2\text{O}$  molecules, and their amide protons are not observed in the spectra. However, those amide protons that are more shielded, exchange slowly at a rate dependent on the degree of solvent exposure, in other words, the extent of involvement in secondary structure formation. These protected amide protons are observed in the spectra for hours and even days. TOCSY spectra are used to identify the amino acids with slowly exchanging amide protons from the observation of its entire connected spin system.

In this context, the presence of secondary structure and the amide components of hydrogen bonds are identified. Such information is also useful for providing additional distance restraints for 3D structure calculations. After preliminary trial calculations, the family of structures generated may be used to indicate the carbonyl acceptor of the hydrogen bond. It may be then possible to obtain further information on secondary structure. For instance, in  $\alpha$ -helices and  $3_{10}$ -helices, hydrogen bonds between the carbonyl of one residue, and the amide group of another residue four or three amino acids respectively along the sequence, are typical. For  $\beta$ -sheet structure, which exhibits a dense network of hydrogen bonds, hydrogen bonds are observed between residues forming the sheet.

#### 2.3.4 Temperature Coefficients

Determination of temperature coefficients of amide bonds may give further evidence for the presence of intramolecular hydrogen bonding (Hruby, 1994). The amide proton chemical shift is known to move upfield at increasing temperatures, due to decreasing hydrogen bonding with solvent molecules. Temperature coefficients can be measured by recording a series of 1D spectra at varying temperature. The amide chemical shift for each amino acid can be plotted as a function of temperature. The gradient of each line corresponds to the NH temperature coefficients, which are defined as the rates of change of the amide chemical shifts with temperature in ppb/ $^{\circ}\text{C}$ . A temperature coefficient of magnitude greater than 6 is obtained for an amide group which is completely solvent exposed, whereas a temperature coefficient of less than 3 is observed for an amide group that is solvent protected and involved in intramolecular hydrogen bonding. Intermediate values could be due to the presence of conformational equilibria between hydrogen bonded and solvent exposed protons.

### 2.3.5 Interresidual NOEs

A wealth of structural information is contained within the NOE cross peaks of NOESY data. Cross peaks are observed between protons that are within 5 Å distance. As mentioned previously, the intensities of these peaks are related to the inverse sixth power of the internuclear distance. Consequently, the closer two protons are to each other, the stronger the resulting cross peak in the NOESY spectrum. The NOE cross peaks are therefore a reflection of the secondary and tertiary structure of the polypeptide, and thus characteristic NOE cross peak patterns and intensities can be defined for secondary structural elements.

To describe the NOE relationship between protons, standard nomenclature is used (Wüthrich, 1986). The NOE  $d_{AB}(i,j)$  refers to the NOE between the A type proton of residue  $i$  and the B type proton of residue  $j$ . For example, the NOE  $d_{\alpha N}(i,i+4)$  denotes the NOE between any alpha proton and the amide proton, four residues along the sequence. Often NOEs are also considered in terms of their range being, either short, medium or long range. Short range NOEs signify that the NOE is between protons of adjacent residues. Medium range NOEs occur between non-sequential residues within a segment of five consecutive residues and long range NOEs arise between protons separated by more than five residues. This terminology is used in the classification of regular secondary structure.

The  $\alpha$ -helical section of peptides, which consist of 3.6 residues per turn are primarily characterised by a series of medium intensity  $d_{\alpha N}(i,i+3)$  and  $d_{\alpha\beta}(i,i+3)$  NOEs. In addition, they exhibit weaker  $d_{\alpha N}(i,i+1)$  and  $d_{\beta N}(i,i+1)$  NOEs, relatively strong  $d_{NN}(i,i+1)$  NOEs and weak  $d_{\alpha N}(i,i+2)$  and  $d_{\alpha N}(i,i+4)$  NOEs. The  $3_{10}$ -helix, which consists of 3 residues per turn, also displays these NOEs, but has weaker  $d_{\alpha\beta}(i,i+3)$  NOEs, and  $d_{\alpha N}(i,i+4)$  are not seen. The NOE pattern differs for  $\beta$ -strand character.



The  $\beta$ -strand presents with strong  $d_{\alpha\alpha}(i,i+1)$  NOEs, but  $d_{\alpha\beta}(i,i+1)$  NOEs are either very weak or absent. The anti-parallel  $\beta$ -sheet is more difficult to identify. However,  $d_{\alpha\alpha}(i,j)$ ,  $d_{\alpha\beta}(i,j)$  and  $d_{\beta\beta}(i,j)$  NOEs across residues of contacting strands are expected to be observed. Random coil, or unstructured peptides, do not give rise to any medium or long range NOE patterns. Similar to the  $\beta$ -strand, random coil and unstructured peptides show only strong  $d_{\alpha\alpha}(i,i+1)$  NOEs and perhaps some weak  $d_{\alpha\beta}(i,i+1)$  NOEs.

NOEs can also be used to determine *cis/trans* conformation of prolines (Wüthrich, 1986). The distances between the alpha/delta protons of the proline and the amide/alpha protons of the preceding residue are influenced by *cis/trans* geometry. As the *cis* configuration permits closer contact between the alpha proton of the proline and the alpha/amide protons of the preceding residue, NOEs due to through space couplings between these protons could be observed in the NOESY spectrum. The *trans* configuration results in shorter distances between the delta protons of the proline and the amide/alpha protons so that, NOEs due to these through space couplings could be manifested in the NOESY data.

Finally, NOE data provides the fundamental input distance data for 3D structure calculations. The NOEs are hence a most important parameter that demand confidence in interpretation and assignment for the generation of reliable 3D structures.

## 2.4 3D Structure Calculations

### 2.4.1 Experimental Restraints

The overall process of 3D structure calculations involves firstly assignment and classification of NOES from NOESY data. Distances are then determined on the basis of the NOE intensities, which are related to the separations or distances between interacting nuclei. It is important to be aware of other mechanisms that can affect intensity and give rise to false

NOEs. Such mechanisms include spin diffusion, internuclear exchange and varying correlation times of individual nuclei. Having taken these effects and limitations into account, a conservative approach to assigning distances has been adopted in this work. The peak volumes of the NOE cross peaks were visually assessed and categorized according to strong, medium, weak and very weak. Corresponding upper distance limits of 2.7, 3.5, 5.0 and 6.0 Å, respectively, were ascribed (Clore *et al.*, 1986b; Williamson *et al.*, 1985). Pseudoatom corrections were added to upper limits to allow for maximum possible error in circumstances where hydrogens could not be uniquely specified. This included non-stereospecific methylene protons (+ 1.0 Å), aromatic protons (+ 2.0 Å), methyl group protons (+ 1.5 Å) and indistinguishable methyl groups (+ 3.0 Å) of valine and leucine residues (Wüthrich, 1986).

It has been ascertained that an average of fifteen NOE distance restraints for every amino acid is required to generate a potentially realistic, accurate set of final structures in calculations based on distance restraints (Clore *et al.*, 1993). The addition of restraints from hydrogen bonds and dihedral angles improves the final structures further. However, initially 3D calculations were carried out using only NOE NMR derived distances. It is essential that the assignments are correct, as errors in data will bias the calculations to produce structures that do not emulate the real situation. In an iterative process, the set of NOE restraints may be updated and expanded until an optimum family of structures is obtained. This permits NOE distance restraints, which may be in error to be corrected and examination of the resulting structures can be used to resolve ambiguous NOEs, which may be included in subsequent calculations. The calculations are repeated, using also additional information from coupling constants and hydrogen bonds, where the carbonyl acceptors have been implicated from the initial calculated structures. If the family of structures meet certain requirements, the structures are ultimately refined to produce a final set of structures, which represent the 3D topology of the polypeptide.

### 2.4.2 Dynamical Simulated Annealing

There are several computational procedures that can be used in 3D structure determination of polypeptides from NMR experimental data. All of these procedures basically aim to explore conformational space within the bounds of the set of restraints. In this work, the method of dynamical simulated annealing (Nilges *et al.*, 1988) was applied. This method is well known and widely used, performing calculations within a reasonable amount of computational time and resources. The final structures are generally in accord with the NMR experimental restraints and are consistent with ideal covalent geometry.

Dynamical simulated annealing entails the application of molecular dynamics under a simplified force field, at high temperature, followed by slow cooling. Initially, a set of structures with random  $\phi$  and  $\psi$  angles and extended side chains was generated. These structures were minimised to establish correct covalent geometry and to remove improper contacts before use as the starting structures for dynamical simulated annealing. The dynamics involves the numerical integration of Newton's equation of motion, using very small time steps to give atom velocities following the time step and displacement during the time step. This equation is given by:

$$d^2r_i(t)/dt^2 = m_i^{-1} F_i \quad (i = 1, \dots, n)$$

where,  $F_i$  represents the forces acting on  $n$  particles with masses  $m_i$  and positions  $r_i$ , and the quantity  $d^2r_i(t)/dt^2$  is the acceleration. The forces  $F_i$  are related to the negative gradient of the potential energy function  $E$  with respect to Cartesian coordinates.

The potential energy function, or force field, includes terms to maintain the covalent geometry and ensure proper chiralities and planarities, and terms for experimental NMR restraints. In addition, there is a simple van der Waal's

repulsive potential, which represents the non bonded interactions, by treating the atoms as soft spheres without attractive or long range non-bonded interactions. This simpler repulsion term replaces the more complicated Lennard-Jones and Coulombic potentials used in other molecular dynamical algorithms, and consequently the calculations are hastened. The entire potential energy function is given by:

$$E = \sum_{\text{bonds}} k_b (r - r_o)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_o)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n_\phi + \delta)] \\ + \sum_{\text{impropers}} k_\phi (\phi - \delta)^2 + \sum_{\text{nonbonded}} k_{\text{repel}} \{\max[0, (sR_{\text{min}})^2 - R^2]\}^2 \\ + \sum_{\text{distance restraints}} k_d \Delta_d^2 + \sum_{\text{angle restraints}} k_a \Delta_a^2$$

where,  $k_b$ ,  $k_\theta$ ,  $k_\phi$ ,  $k_{\text{repel}}$ ,  $k_d$ , and  $k_a$  are the various force constants,  $r$  and  $r_o$  the actual and correct bond length, respectively,  $\theta$  and  $\theta_o$  the actual and correct bond angle,  $\phi$  the actual dihedral angle,  $n$  the number of minima of the dihedral angle potential,  $\delta$  an offset of the dihedral angle and improper potentials,  $R_{\text{min}}$  the distance where the van der Waal's potential has its minimum,  $R$  the actual distance between a non bonded atom pair,  $s$  a scaling factor, and  $\Delta_d$  and  $\Delta_a$  the size of the distance and dihedral angle restraint violation.

The molecular dynamics simulation permits the peptide molecule to explore the potential energy surface to locate the global minimum. This search is expanded with the incorporation of simulated annealing into the process, whereby molecular dynamics is carried out in stages commencing at high temperature (eg 1000K) for several picoseconds. During high temperature dynamics, low weightings are placed on the repel force constant and the NOE restraints. The increased kinetic energy and the weak constraint on non bonded interactions allows the atoms to pass through, or near each other, so that the molecule can overcome being trapped in local minima and cross potential energy barriers. In this way, the molecule can sample a larger region

of conformational space, which increases the probability of locating the global minimum. This session of molecular dynamics is followed by another several picoseconds of dynamics, also at the same high temperature, but with the force constant on the NOE restraints increased. Subsequently, the system is cooled with dynamics carried out at each temperature step decrease and with increases on the repel force constant, which ensures that there are no non-bonded contacts.

To improve the structural energies and geometries, the structures generated were used as input for further simulated annealing using final values on NOE and repel force constants. These structures were then refined in a force field, where NMR experimental restraints are maintained by a different potential function. Furthermore, a modified molecular topology definition, which prevents unrealistic interactions between side chains and backbone atoms from occurring, was utilised during refinement. The 3D structure calculation protocol is schematically outlined in Figure 2.6.

The structures were examined for violations of the NMR restraints after the first round of dynamical simulated annealing. Ambiguous NOE restraints could be resolved and hydrogen bonds could be determined by analysis of the structures in relation to the experimental data. The process of initial dynamical simulated annealing and examination of results was repeated until the best and most accurate NOE restraints, giving rise to structures with the lowest energies were obtained. To assess structural quality, several measures were taken.

#### **2.4.3 Analysis of Structural Quality**

The first of the indicators of structural quality are the residual restraints' violations. A list of violations above an acceptable limit may be generated at the end of structure calculations. This list is examined for consistently violating experimental restraints over the lower energy structures. These violations were compared to the structures and referred to the NMR data in

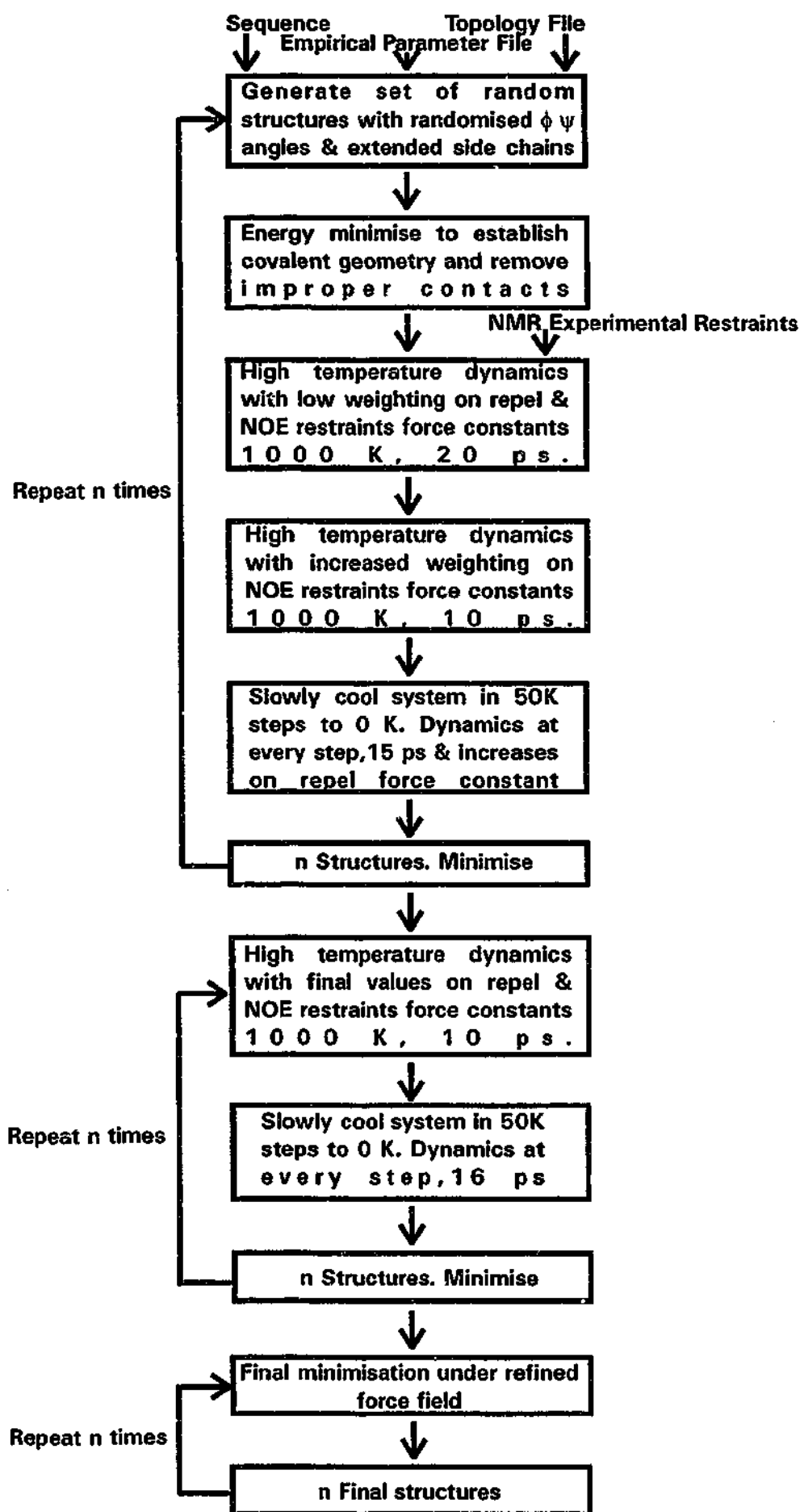


Figure 2.6. Dynamical simulated annealing protocol used to generate 3D structures from NMR experimental restraints.

order to determine possible errors of assignment, due to for example, spectral overlap. The use of residual violation lists enables the correction of NOEs, so that a set of accurate NOE restraints could be established to calculate optimum, realistic structures.

As well as lists of residual restraints' violations, lists of energy values for structures may be obtained. The values of the total, bond, angle, dihedral, improper, van der Waal's and NOE energies and their corresponding deviations may be used to evaluate structural quality. The aim is to produce structures where these energies and rms deviations are as low as possible. There should be minimum variation of the engeries overall in the best set of final structures. It is important to realise that these values are not necessarily a measure of accuracy, but rather precision. Acceptable energetic and geometric statistics may also be obtained for a set of structures trapped in a local, low minimum or an inaccurate fold, where the NOEs have been misassigned, but the structures may satisfy the restraints. The energy values should be used in conjunction with all structural information.

Another useful measure is the pairwise root mean square deviation (RMSD) for a given set of atoms, such as the backbone atoms, or all the heavy atoms, of an ensemble of structures. This parameter is used to estimate the degree of convergence of a set of structures and may also be used to indicate regions that are well defined. Low RMSD values indicate better convergence of structures and/or good structural definition. Similar to energetic and geometric statistics, the pairwise RMSD is a measure of precision and should be viewed with caution.

Additionally, there are two very useful programs that may be applied to analyse structures. The program, PROCHECK (Laskowski *et al.*, 1993), was used to examine the stereochemical quality. Among many functions of this program, the  $\phi$  and  $\theta$  angles can be plotted to determine compliance with Ramachandran allowed regions (Ramachandran and Safisekharan, 1968). Full compliance is an indication of a structure with ideal geometry. The second program, PROMOTIF (Hutchinson and Thornton, 1996), was applied to

identify classical structural features. This program automatically analyses the structure files and determines the presence and types of turns, helices,  $\beta$ -structures and disulphide conformations. The program creates postscript files, which describe in detail the identified structural elements.

Finally, the structures must be visually regarded, especially in relation to the experimental data. Interactive graphical analysis of structures permits the viewer to develop a 3D image of the general fold, as well as, detect unrealistic or unusual features. Crucial to the success of 3D structure calculations based on NMR spectroscopic data is purity of the peptide sample under investigation, as the presence of impurities can lead to NOE misassignment or unwanted chemical interactions. The following section describes the methodology used in this work to obtain pure material for study.

## **2.5 Peptide Purification**

There are many different techniques that are currently available for separation and isolation of polypeptides from mixtures. One of the most well established and most popular techniques in protein chemistry is reverse phase high performance liquid chromatography (RP-HPLC). RP-HPLC gives high mass recoveries and is particularly useful for separation of closely related species, such as those of the crude mixture from polypeptide synthesis. RP-HPLC was the technique of preference used in this work and was applied to isolate and concentrate the prime polypeptide component after synthesis. For good separation, it is important to understand the underlying mechanisms and conditions required to optimise separation.

### **2.5.1 Separation Mechanism**

RP-HPLC is characterised by the nature of the column surface, which governs the separation process. The RP column, referred to as the stationary phase, is composed of silica supports that are surface modified by chemically



bonded hydrophobic groups generally, alkyl chains of varying length. It is at this hydrophobic surface where interactions and separation takes place. The separation mechanism relies on relative hydrophobicities of the polypeptide components of a mixture, where the overall hydrophobicity of a polypeptide is determined by the fold and residual composition and distribution. The mobile phase in RP-HPLC consists of water and a miscible organic solvent, which functions as an organic modifier of the mobile phase environment. A low level of an additive is also present to act as an ion pairing agent or buffer and to aid polypeptide solubility. UV detection of the peptide bond at  $\sim 206$  nm and or aromatic functional groups at  $\sim 280$  nm is used for recording polypeptide retention during a changing mobile phase environment.

Gradient elution is preferred to isocratic elution, as often the retention characteristics and identities of the component polypeptides are unknown. Upon injection of the polypeptide mixture into the HPLC system, the chromatographic process commences with the passing of 100% water at a low flow rate through the column system. The polypeptide components adsorb onto the surface of the column to an extent determined by the hydrophobic attraction of the particular polypeptide to the column surface and the hydrophilic repulsion of the particular polypeptide from the solvent environment. The composition of the mobile phase is incrementally changed over time, with increasing proportion of the organic modifier simultaneous with equivalent decreasing proportion of water. Thus, an increasing hydrophobic mobile phase environment is created and the dynamics of attraction between polypeptides and the column surface are decreased as the attraction between the polypeptides and the solvent environment is increased.

At critical concentrations of organic modifier, the attraction between individual polypeptides and the solvent environment is sufficiently strong that the polypeptide molecules rapidly desorb from the column surface and complete the transit through the column to be detected and collected. For every different polypeptide in the mixture, the concentration of organic modifier required for desorption is different, depending on the hydrophobic

character which defines the dynamic equilibrium between the polypeptide, column and mobile phase. This method of gradient separation is so successful because elution is extremely sensitive to minute changes in organic solvent strength. For this separation system, more hydrophilic polypeptides are eluted sooner, while more hydrophobic polypeptides have longer retention times. There are a number of parameters that can be controlled by the operator to optimise separation.

### 2.5.2 Chromatographic Conditions

The conditions, which can be altered to improve the separation, include mobile phase composition and column properties. Standard mobile phase and stationary phases were employed in this work. The mobile phase is the easiest to manipulate and was chosen not only for the separation power, but also for ease of use.

The mobile phase used for most separations and for all fraction collecting consisted of aqueous acetonitrile with 0.1 % trifluoroacetic acid (TFA). The use of acetonitrile as an organic modifier avoids separating problems from high column back pressure that may be experienced using the more viscous isopropanols, that are also a popular choice, due to their excellent eluting capabilities. Acetonitrile lies between simple alcohols and isopropanols in organic strength. Acetonitrile is highly volatile and is easily removed from collected fractions during lyophilization. Furthermore, acetonitrile is relatively UV transparent, so eluted peaks may be detected.

Similarly, TFA is also very volatile. The addition of TFA serves a number of purposes. TFA is an excellent for solubilising many polypeptides. Moreover, TFA acts as a hydrophobic, anionic ion pairing reagent, that complexes with and therefore masks the positively charged, basic residues of the polypeptide. TFA maintains a low pH of 2-3, where -COOH functional groups of the polypeptide residues are forced into the protonated form. The overall effect is to reduce the polar or hydrophilic character of the polypeptide, so that

interaction between the hydrophobic portions of the polypeptide and the column matrix are enhanced.

Additionally, TFA ion pairs with free, accessible silanol groups of the column, concealing the polar Si-O-H groups with hydrophobic  $-CF_3$  groups, thereby marginally increasing the hydrophobicity of the column surface. In its entirety, this particular mobile phase system is driven by competing relative hydrophobicities. The use of aqueous acetonitrile with TFA was the preferred mobile phase, as all the components can be removed upon lyophilization, leaving pure polypeptide material for NMR analysis. However, the addition of alternative minor additives to TFA was useful in analytical separations to obtain a different retention profile by altering column and polypeptide surface character, in order to observe the presence of other polypeptides that may have co-eluted with other fractions when TFA was used.

On occasion, the TFA was replaced by non-volatile phosphoric acid. Similar to TFA, phosphoric acid interacts with basic polypeptide residues masking these residues, and also protonates the acid polypeptide residues. However, phosphoric acid is a weaker acid than TFA, and hence protonation is less effective. Consequently, the hydrophilicity of the polypeptide is less reduced in phosphoric acid than in TFA. In respect of the column surface, phosphoric acid ion pairs with the free accessible silanol groups, forming polar P-O-H groups on the column surface. The P-O-H group is more polar than the Si-O-H group of the column, and the effect of phosphoric acid, as opposed to TFA, is to increase column surface hydrophilicity. A third mechanism was also used, where phosphoric acid and triethylamine were added. These two additives are likely to form a complex together that has little or no effect on either the column or polypeptide surfaces, and so this system represents the unadulterated dynamics of column-polypeptide-mobile phase interactions. Having manipulated the mobile phase conditions, the stationary phase may be also altered simply to optimise separation.

Commercially available RP columns for HPLC are excellent for polypeptide chromatography. The silica based columns that are surface

modified with covalently bonded alkyl hydrophobic groups are most commonly used. These columns differ in the length of the alkyl group. The most popular columns have surface moieties due *n*-butyl (C4), *n*-octyl (C8), *n*-octadecyl (C18) and the phenyl group. Columns with shorter alkyl groups (C4) are best suited for more hydrophobic peptides, while those with longer alkyl groups (C18) are better for more hydrophilic polypeptides. The longer the alkyl chain, the stronger the hydrophobic interaction between the polypeptide and the stationary phase, and at some point this may have a negative effect. In other words, the interactions may be too strong and the polypeptide may not be removed. There are other parameters which determine the separating performance of a particular column.

Column dimensions, particle size and pore size can all affect polypeptide retention. A longer, thinner column can improve resolution required for complex mixtures in comparison to a shorter, fatter column which, permits greater loading capacity required for bulk separation. The particle sizes are generally between 5 and 10  $\mu\text{m}$  and this range does not affect performance greatly. However, theory suggests that column performance should increase with decreasing particle size, due to the greater opportunity for polypeptide interactions. In response, columns of particle sizes 2 to 3  $\mu\text{m}$  are available and improve resolution. However, such columns were not utilised in this work. Lastly pore size should be selected on the basis of the polypeptide size to avoid restricted physical access for polypeptide interactions. Pore size is standardised and sizes of 80 Å are adequate for smaller molecules, but 300 Å should be used for larger samples. Finally, consideration for contamination, chromatographic reproducibility and loss of biological integrity requires attention.

Steps in the process of chromatographic separation may be taken to avoid problems and to retain sample purity and quality. Contamination from successive separations and reproducibility of data can be assured by cleaning the column between separations and equilibrating before each separation commences. In addition, only the highest grade of mobile phase components

should be used and filtered before use to eliminate rogue solid particles. It is worthy of note that the column degrades over time with exposure to the acidic mobile phase environment. The operator should be attentive to this occurrence and the lifetime of a column may be increased by washing thoroughly with, and storing the column in, non aggressive solvents.

Finally, although some polypeptides are highly unstable, many polypeptides may be preserved by taking precautions. These include minimising the column transit time and the time collected fractions are exposed to the mobile phase environment. The latter is achieved by freezing the sample in a dry ice/alcohol mixture and lyophilisation of the sample as soon as possible. Adhering to these guidelines led to successful separation of specific polypeptide fractions. The use of RP-HPLC formed an important part of this work.

## **2.6 Applications**

The RP-HPLC methodology described above was used to isolate and concentrate a key polypeptide fragment of kalata-B1. This fragment could then be structurally analysed through the application of 2D NMR techniques outlined in this chapter. The 2D NMR techniques, involving both  $^1\text{H}$  and  $^{13}\text{C}$  nuclei, were applied to further study the secondary and tertiary structure of kalata-B1. Subsequently, the 2D  $^1\text{H}$  NMR techniques were utilised to identify the secondary structures and provide experimental data for computational structure calculations on two novel, related macrocyclic plant peptides with anti-HIV activity. The results and discussion of the findings are presented in the ensuing chapters.

**CHAPTER THREE**

**TOPOLOGICAL CONSIDERATIONS**

**FOR CYCLIC KNOTTED PEPTIDES**

### 3.1 Introduction

Knowledge of the three dimensional topology of bioactive peptides and proteins is a crucial aspect in rational drug design. It is essential to understand the effect of topological features and constraints on specific activities and also folding processes. From this understanding, pertinent information on structure/activity relationships, binding sites, recognition and initial interaction mechanisms and synthetic pathways may be derived. The aim of this chapter is to undertake a theoretical analysis of the topology of cyclic cystine knot peptides for the purpose of determining key structural features that may impact directly on folding, activity and synthesis. As mentioned in Chapter 1.4, kinemages\* (Richardson and Richardson, 1992), which serve as a convenient tool for graphical illustrations of 3D structures, will be employed and created to demonstrate several points of interest. The first section of this chapter will consider the strategic analysis of cyclic cystine knots, using kalata-B1 as the prototypic member of the family.

### 3.2 Strategic Structural Analysis of Cyclic Cystine Knot Peptides

There are multiple different techniques that could be employed to analyse structures of polypeptides through the introduction of some perturbation. An approach that could be adopted involves selected linear segments of kalata-B1. Linear segment synthesis and analysis are one of the simplest methods in which to examine sequence dependent structure formation.

---

\* All kinemages presented in this chapter are found on the floppy disk accompanying this thesis, along with the program MAGE that can be used to view the kinemages on a PC operating under windows 95/98/2000. Specific kinemages are referred to throughout the chapter. To view the kinemages, execute MAGE and open the kinemage file of interest. The kinemage files have the extension of .kin. Animations may be viewed by depressing the keyboard letter "a" for observation of each successive picture of the animation. To obtain information on kinemages, consult the web page at the following address, <http://kinemage.biochem.duke.edu/website/kinhome.htm>. To gain further knowledge on the use of MAGE, download the tutorial by Gary Manfredy at this website.

### 3.2.1 Linear Segments

Short linear segments, which are of structural importance in proteins, are often studied exclusively to establish any propensity to form structure in isolation. To investigate the contribution of sequence to structure in kalata-B1, various linear segments, known to partake in local secondary structure of the 3D fold, could be synthesised and analysed by NMR spectroscopy. Although the 3D structure of kalata-B1 has been described in general in Chapter 1, a detailed examination of the structure is required in order to identify the important segments.

Kalata-B1 consists of three short antiparallel strands held together by its disulphide bonding pattern to form a compact sulphur core. Turns and loops connect these strands. A diagram of kalata-B1 labeled with elements of local secondary structure is shown in Figure 3.1 (Saether *et al.*, 1995).

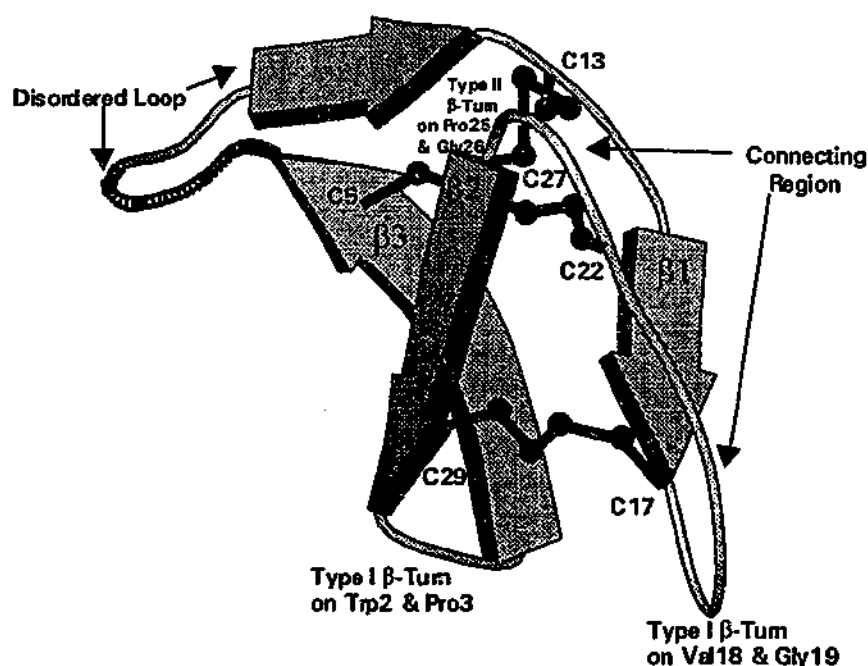


Figure 3.1. Richardson-type diagram of kalata-B1 labeled with elements of local secondary structure (Saether *et al.*, 1995).  $\beta$ -Strands are represented by block arrows and disulphides by thick black lines. The strands of the triple stranded antiparallel  $\beta$ -sheet are labeled  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ . The half-cystines of the disulphides are also labeled.



In the first instance, it could be postulated that the segments of greatest biological importance are likely to be the sections of sequence involved in the formation of turns, as turns are often found at active sites (Richardson, 1981). Through careful examination of the structure of kal<sub>2</sub>a-B1, it seems that many of the turns and significant local features begin and end near the cystine residues. Consequently, the set of linear segments that could be analysed are synonymous with combinations of neighbouring loops designated in the global alignment of all the known cyclic cystine knot peptides (Figure 1.1.11). These loop segments are described below. The naming of the linear segments refers to the particular defined loops that comprise the segment, with the "+" sign in the name indicating neighbouring loops are added together sequentially to form the segment.

(1) Loops 6 + 1

6 7 8 9 10 11 12 13 14 15 16

Thr-Arg-Asn-Gly-Leu-Pro-Val-Cys-Gly-Glu-Thr

This section consists of mainly residues of the disordered loop, Asn8 to Cys13. The section protrudes somewhat from the rest of the peptide, and so it could be a probable point of attachment to receptors. It also includes two charged residues, Arg7 and Glu15, which are located at each end of the segment and are on opposite sides of the hydrophobic surface of the original peptide. These residues have been postulated to provide recognition of and binding to receptors (Saether *et al.*, 1995).

(2) Loops 4 + 5

28 29 1 2 3 4

Thr-Cys-Ser-Trp-Pro-Val

This section involves most of the residues of the  $\beta$ -hairpin, which includes the type I  $\beta$ -turn on Trp2 and Pro3 (Saether *et al.*, 1995). The Trp2 residue is bifunctional, is often found at active sites (Walton, 1981) and together with the turn, renders this segment a likely receptor binding point.

(3) Loops 1 + 2

14 15 16 17 18 19 20 21

Gly-Glu-Thr-Cys-Val-Gly-Gly-Thr

This section is part of the connecting region, which joins one extended strand to the  $\beta$ -hairpin. It involves a type I  $\beta$ -turn on residues Val18 and Gly19 (Saether *et al.*, 1995).

(4) Loops 3 + 4

23 24 25 26 27 28

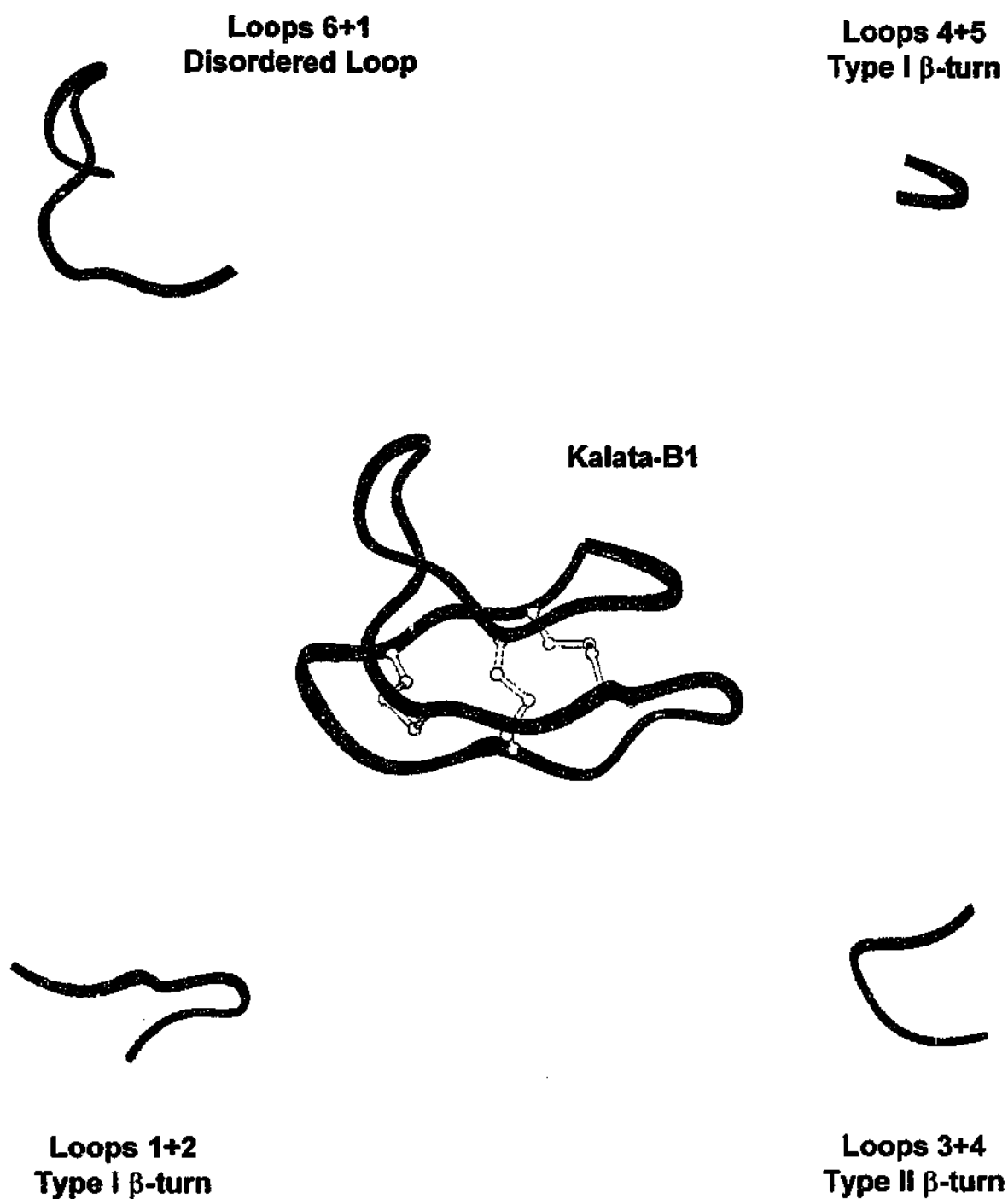
Asn-Thr-Pro-Gly-Cys-Thr

This section involves the remaining residues of the connecting region, not part of segment (3) above. A type II  $\beta$ -turn on residues Pro25 and Gly26 is included (Saether *et al.*, 1995).

These four segments, displayed in Figure 3.2, capture the obviously significant secondary structure features of kalata-B1. However, the segments are short, comprising 11, 6, 8 and 6 residues, respectively, and it is well known that peptides with less than 20 residues commonly adopt many conformations in rapid interconverting equilibrium (Kessler *et al.*, 1996; Dyson and Wright, 1991), rendering structural analysis by NMR spectroscopy often difficult. For this reason, an alternative set of peptide fragments, where the sequences are longer was proposed. The members of the set were referred to as acyclic permutants.

### 3.2.2 Acyclic Permutants

The proposition of the acyclic permutants is based on the premise that activity is dependent on conformation, as well as the particular sequence of amino acids. From a close examination of the 3D fold of kalata-B1, it can be seen that by synthesising small linear sections, it would be difficult to confer structure. All the three disulphide bonds are crucial to the overall 3D fold by



**Figure 3.2.** Linear segments of kalata-B1. These segments capture the significant secondary structural features of kalata-B1. Entire kalata-B1 is displayed in the centre.

virtue of the nonplanar, chiral nature of kalata-B1 procured by the specific disulphide bonding pattern of kalata-B1.

These observations lead to an alternative approach, which involves the synthesis of the set of the largest fragments, where all three disulphide bonds and the general structure are conserved, but with omission of the sequences between successive pairs of cystine residues. The entire set then comprises all the possibilities of individually absent loop sequences. The structural analysis of the resulting fragments enables investigation of the significance of small prominent sections and local secondary structure on the overall fold, through the omission of particular loop sections. Once the structures of the fragments have been resolved, the activities could also be measured to ascertain the effect of any structure pertaining to specific activities. These fragments form a set of non-cyclic variants or permutants of kalata-B1, and have been appropriately named acyclic permutants to reflect their origin. The acyclic permutants, numbered according to the excluded loop sections, are listed below, together with the secondary structure of the omitted fragment in kalata-B1 that the acyclic permutant is designed to investigate.

(1) Acyclic Permutant 1

17 18 19 20 21 22 23 24 25 26 27 28 29 1 2 3 4 5 6 7 8 9 10 11 12 13  
Cys-Val-Gly-Gly-Thr-Cys-Asn-Thr-Pro-Gly-Cys-Thr-Cys-Ser-Trp-Pro-Val-Cys-Thr-Arg-Asn-Gly-Leu-Pro-Val-Cys

Residues Gly14 to Thr16 are excluded. This fragment examines the effect of the absence of a centrally located, relatively linear sequence.

(2) Acyclic Permutant 2

22 23 24 25 26 27 28 29 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17  
Cys-Asn-Thr-Pro-Gly-Cys-Thr-Cys-Ser-Trp-Pro-Val-Cys-Thr-Arg-Asn-Gly-Leu-Pro-Val-Cys-Gly-Glu-Thr-Cys

Residues Val18 to Thr21 are excluded. This fragment examines the effect of the absence of some of the residues of the connecting region, which joins one

extended strand to the  $\beta$ -hairpin and the type I  $\beta$ -turn on residues Val18 and Gly19 (Saether *et al.*,1995).

### (3) Acyclic Permutant 3

27 28 29 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22  
Cys-Thr-Cys-Ser-Trp-Pro-Val-Cys-Thr-Arg-Asn-Gly-Leu-Pro-Val-Cys-Gly-Glu-Thr-Cys-Val-Gly-Gly-Thr-Cys

Residues Asn23 to Gly26 are excluded. This fragment examines the effect of the absence of the remaining residues of the connecting region not examined by acyclic permutant 2, and the type II  $\beta$ -turn on residues Pro25 and Gly26 (Saether *et al.*,1995).

### (4) Acyclic Permutant 4

29 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27  
Cys-Ser-Trp-Pro-Val-Cys-Thr-Arg-Asn-Gly-Leu-Pro-Val-Cys-Gly-Glu-Thr-Cys-Val-Gly-Gly-Thr-Cys-Asn-Thr-Pro-Gly-Cys

Only one residue, Thr27, is excluded. This fragment examines the effect of the absence a second complete  $\beta$ -strand, which forms part of the triple stranded  $\beta$ -sheet (Saether *et al.*,1995).

### (5) Acyclic Permutant 5

5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29  
Cys-Thr-Arg-Asn-Gly-Leu-Pro-Val-Cys-Gly-Glu-Thr-Cys-Val-Gly-Gly-Thr-Cys-Asn-Thr-Pro-Gly-Cys-Thr-Cys

Residues Ser1 to Val4 are excluded. This fragment examines the effect of the absence of some of the residues of the  $\beta$ -hairpin and the type I  $\beta$ -turn on Trp2 and Pro3 (Saether *et al.*,1995).

### (6) Acyclic Permutant 6

13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 1 2 3 4 5  
Cys-Gly-Glu-Thr-Cys-Val-Gly-Gly-Thr-Cys-Asn-Thr-Pro-Gly-Cys-Thr-Cys-Ser-Trp-Pro-Val-Cys

Residues Thr6 to Val12 are excluded. This fragment examines the effect of the absence of residues of the disordered loop, Asn8 to Cys13 (Saether *et al.*, 1995).

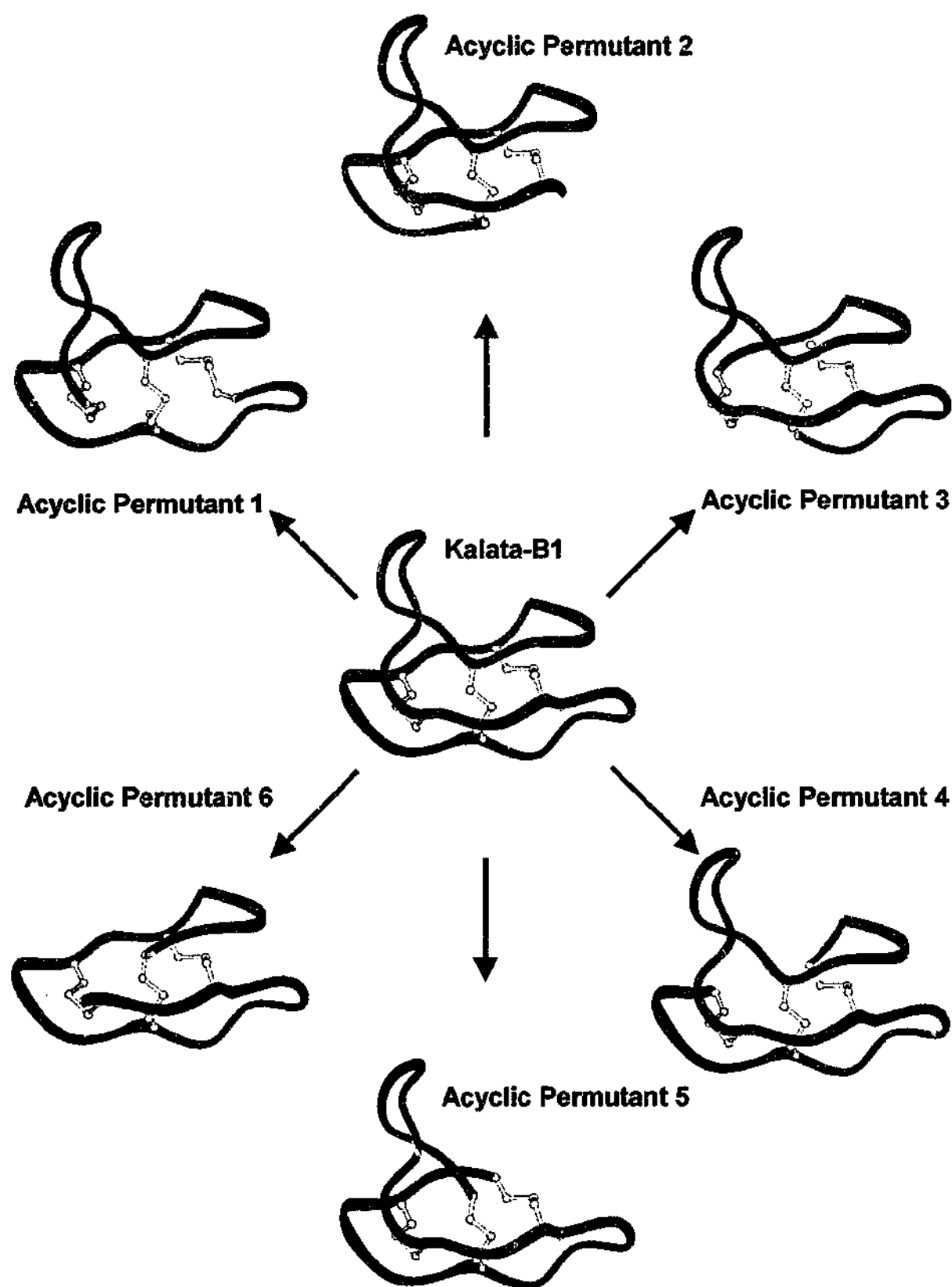
The acyclic permutants are graphically shown in Figure 3.3.

The synthesis of acyclic permutants could potentially result in the production of a multitude of non-native varieties, as well as the native forms. The analysis of non-native permutants is the prime incentive behind work presented in a subsequent chapter in this thesis (Chapter 5). The development and investigation of acyclic permutants with native structure was carried out by other members of the research group, through financial support from the Australian Research Council. The results of the study have been recently published (Daly and Craik, 2000) and are summarised below for comparison with studies of non-native forms described in Chapter 5.

### 3.3 The Study of Native Acyclic Permutants

In a previous chapter, the division of the cyclic cystine knot into six loops, defined by successive cystine residues was discussed (Chapter 1.1). Effectively, the complete set of acyclic permutants represents the six possible topologically distinct permutants that may be derived from the cyclised parent peptide, in which the backbone has been opened in every loop. The six acyclic permutants for kalata-B1 were synthesised, analysed by NMR spectroscopy and tested for hemolytic activity (Daly and Craik, 2000). The concept, synthesis and analysis of the acyclic permutants form a novel approach to the study of folding and structure/activity relationships uniquely offered by cyclic peptides. The results of the investigation provided an insight into the structural imperatives of cyclic cystine knot peptides.

Of the six acyclic permutants synthesised, four folded into the native conformation. These acyclic permutants correspond to the acyclic permutants,



**Figure 3.3.** The set of native acyclic permutants. The parent, kalata-B1, is displayed in the centre.

2, 3, 5 and 6. The acyclic permutants corresponding to the acyclic permutants 1 and 4 did not fold to produce the native structure. Furthermore, none of the native acyclic permutants registered hemolytic activity, which occurs for the original intact cyclic parent peptide. With reference to the missing residues of the acyclic permutants, these results infer the following:

- 1/ the presence of the embedded ring of the cystine knot (Chapter 1.2) is absolutely critical for folding into the native form,
- 2/ disruption of the  $\beta$ -sheet region is inconsequential for folding into the native structure, with the *proviso* that the cystine knot core is maintained,
- 3/ removal of turns causes minimal structure disruption,
- 4/ cyclisation is crucial for hemolytic activity.

Evidently, the cystine knot core and cyclisation are the structural features necessary for correct folding and at least hemolytic activity, respectively. It is also remarkable, that the residues associated with the embedded ring, are highly conserved regions, both in terms of size and content, throughout the cyclic cystine knot family. The native acyclic permutants were also used to formulate ideas on the natural synthetic pathway of macrocyclic peptides, by comparison to another non-cyclic member of the inhibitor knot family.

### 3.4 Identification of the Natural Synthetic Pathway

As kalata-B1 is cyclic, the initial residue of the natural synthetic pathway is not obvious, assuming that the molecule is derived from the cyclisation of linear precursor polypeptide. Furthermore, the gene coding of kalata-B1 is not yet known. Consequently, there are potentially 29 equally possible linear precursors, with different *N*- and *C*-termini that may be synthesised by the plant. To assist in the identification of the most likely region for the initial residue of the natural synthetic pathway of a cyclic cystine knot, kalata-B1 may be topologically compared, in two dimensions, to non-cyclic members of the cystine knot family. By this comparison, a

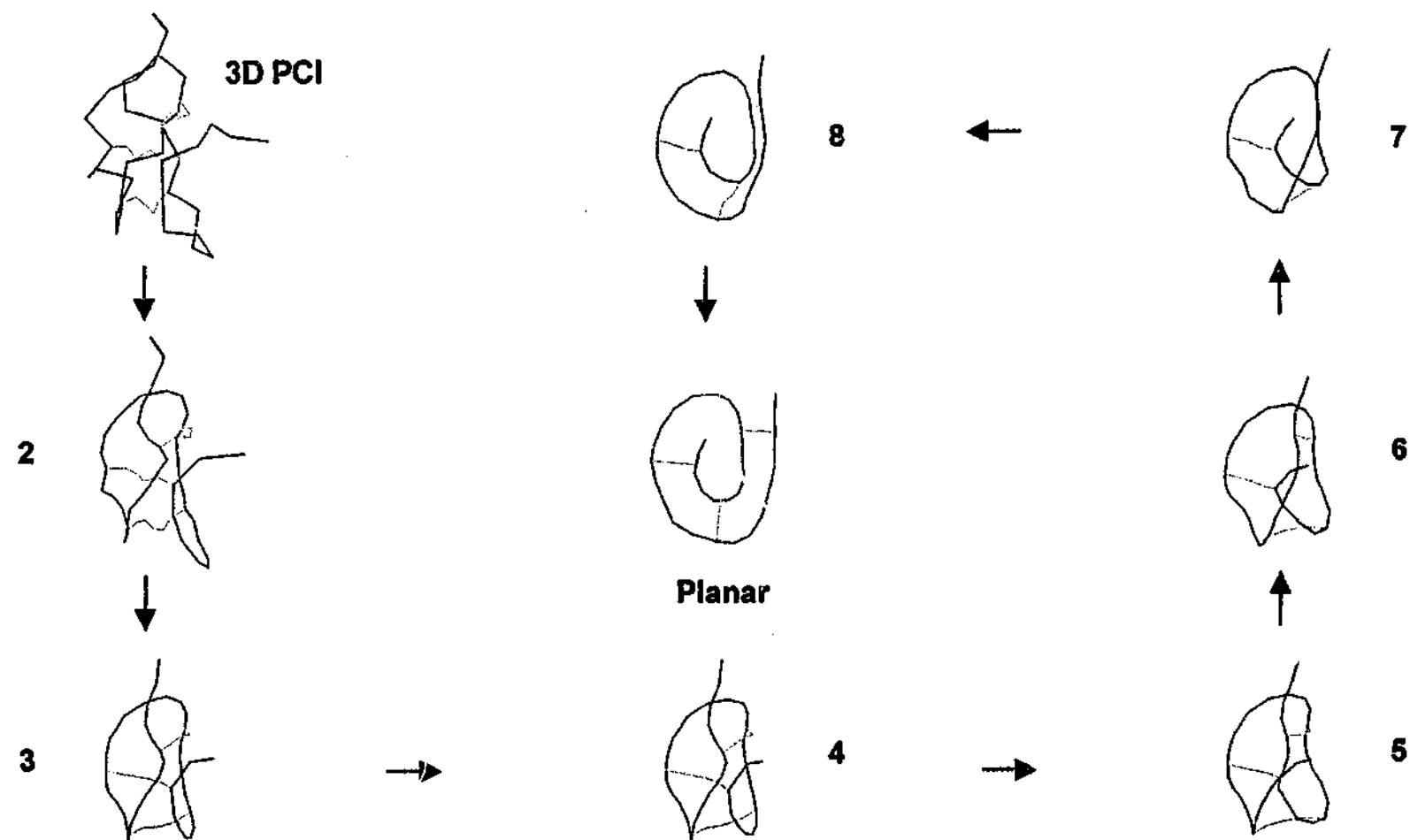


topological basis can be used to predict the possible position of the starting residue. The 2D topology of the non-cyclic relative, potato carboxypeptidase inhibitor (PCI), that has been established, may be used for the comparison.

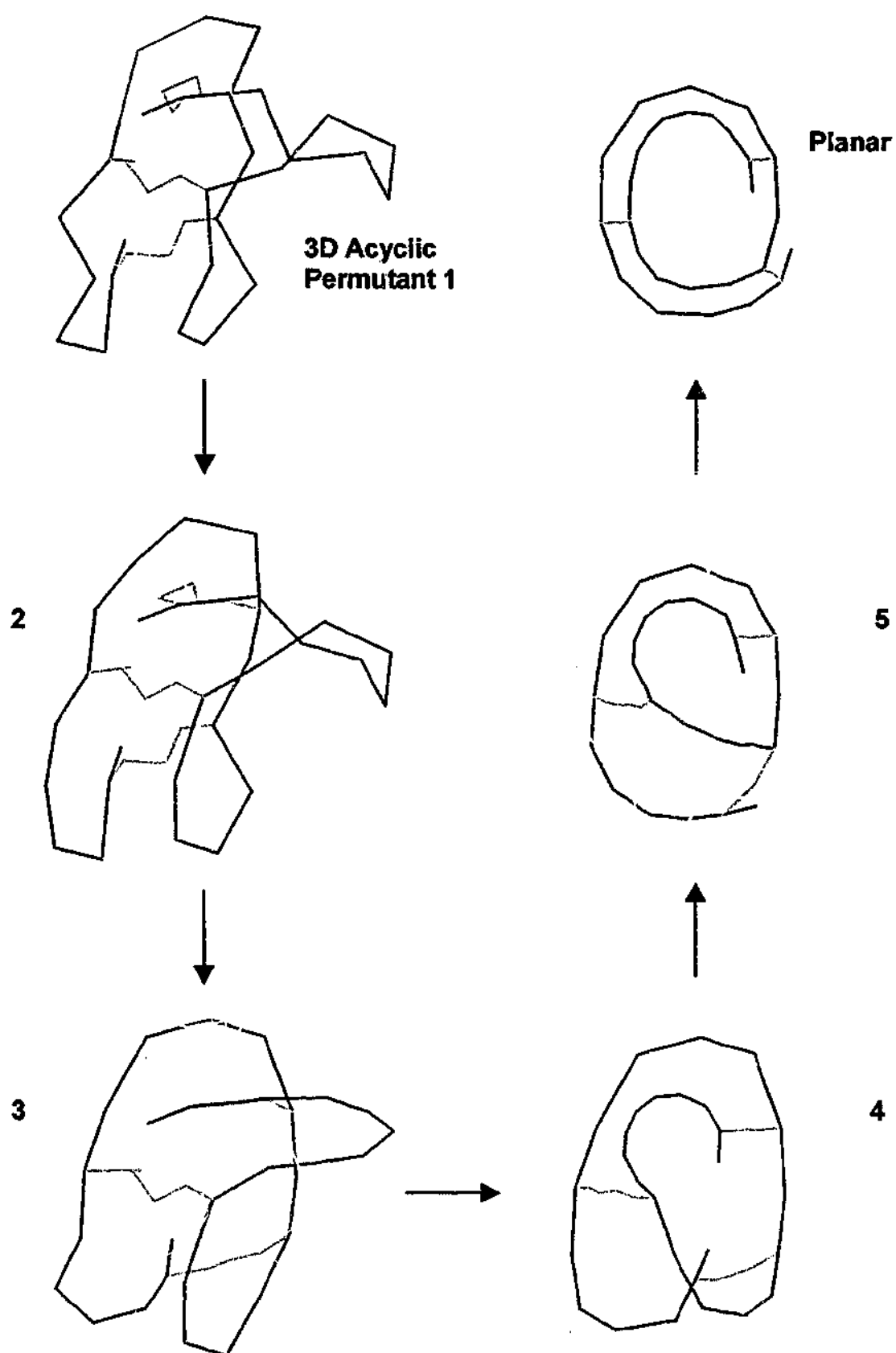
A kinemage of PCI has been created (Benham and Jafri, 1993), in which the 3D structure is unknotted to form the planar structure, by pushing the C-terminus and the connected backbone, through the ring of the cystine knot. There may be a realistic significance in this animation, as it has been suggested that a similar unraveling of a member of the growth factor cystine knot family, human nerve growth factor, may be involved in very slow unfolding (de Young *et al.*, 1996). The animated kinemage is presented in Figure 3.4 (kinemage PCI.kin). The final 2D structure illustrates the planarity and absence of complex topology of non-cyclic cystine knot peptides. Moreover, the disulphide bonding pattern, projected in 2D, is displayed and this pattern is used for comparison.

Kalata-B1, however, is chiral and nonplanar, and cannot be unknotted. Nonetheless, the acyclic permutants are no longer cyclic. They are achiral and mathematically planar, and can be unknotted to give 2D structures, which can be compared to that of PCI. It may be proposed that the 2D topology of the polypeptide backbone and disulphide bonds of the acyclic permutant, which is identical to that of PCI, may be indicative of the starting residue originating from the absent segment of the particular acyclic permutant. Animated kinemages, in which the acyclic permutants were individually unknotted, were produced. An additional length was added to each end of the acyclic permutants for clarity. These animations are shown in Figures 3.5 to 3.10 (kinemages AP\*.kin).

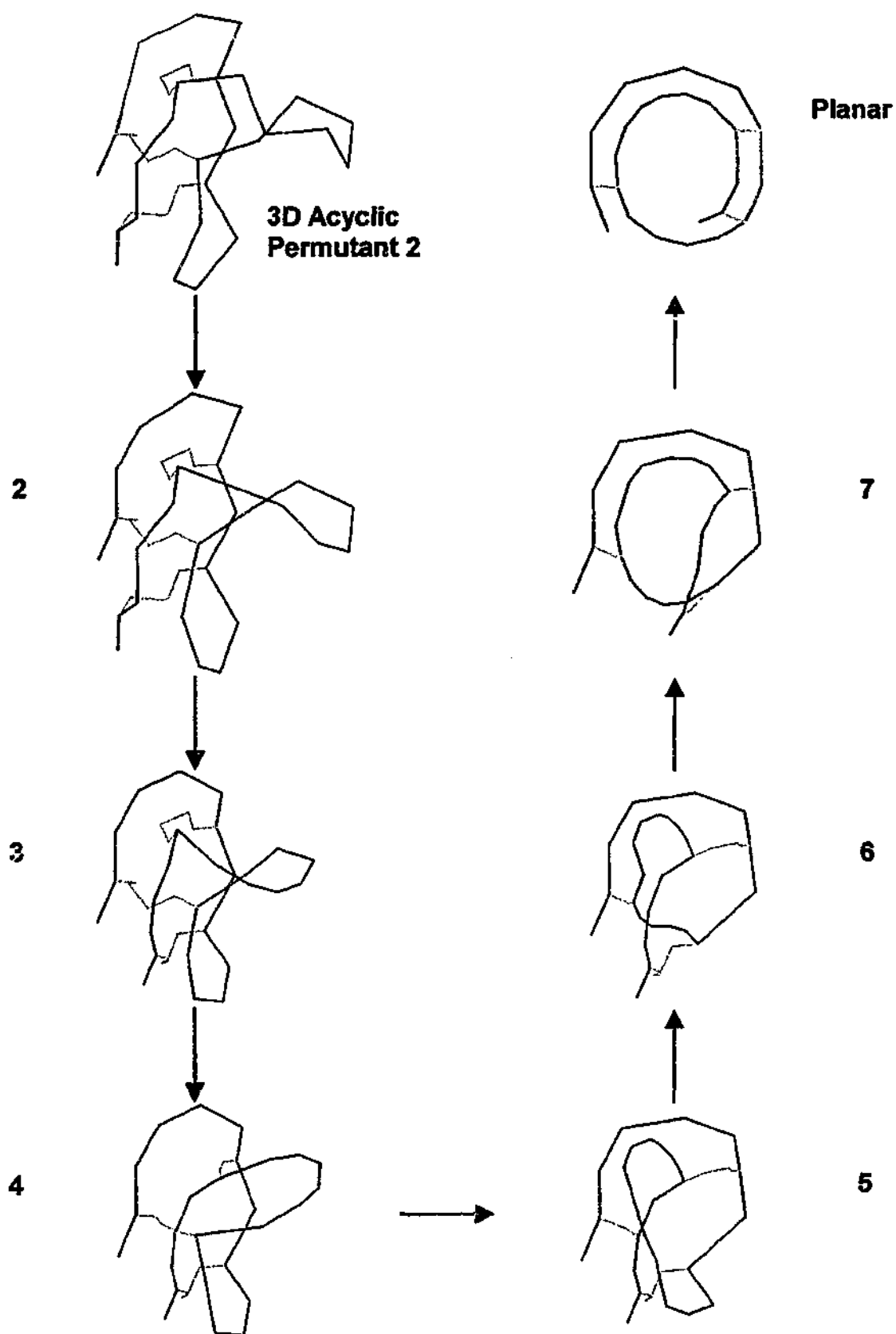
The final planar structures for each acyclic permutant are compared to that of PCI in Figure 3.11. In this diagram, it can be seen that the planar structure of the acyclic permutant 6 is identical to that of PCI. This suggests that the starting residue for the synthesis in the plant, quite possibly, originates in the region of residues Cys5 to Cys13, in other words, loop 6 which comprises the disordered region in kalata-B1. While this result may



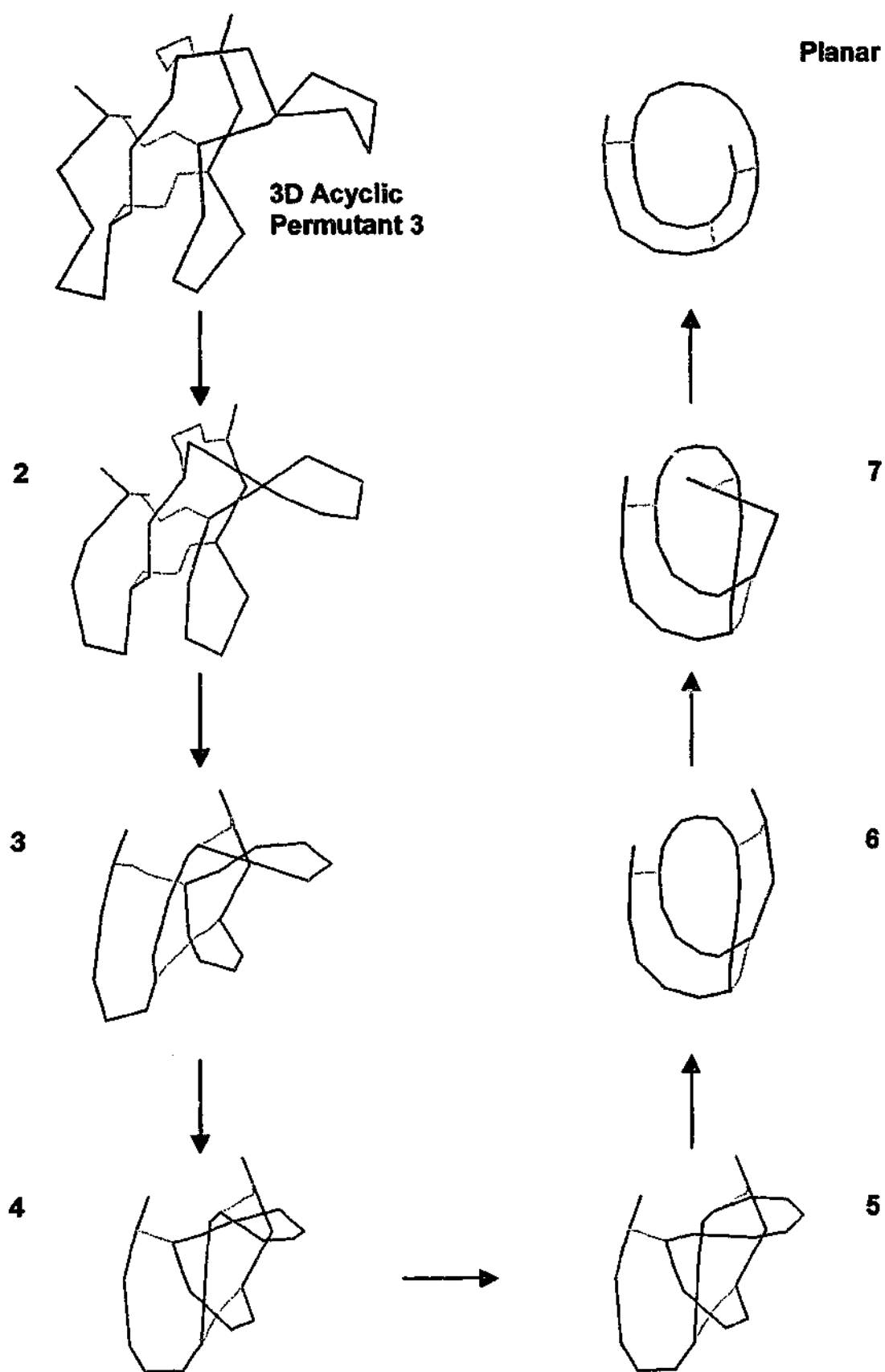
**Figure 3.4.** Animation of the unknotting of the 3D structure of potato carboxypeptidase inhibitor to the planar structure (Benham and Jafri, 1993). The backbone is shown in green and the disulphides bonds are shown in gold.



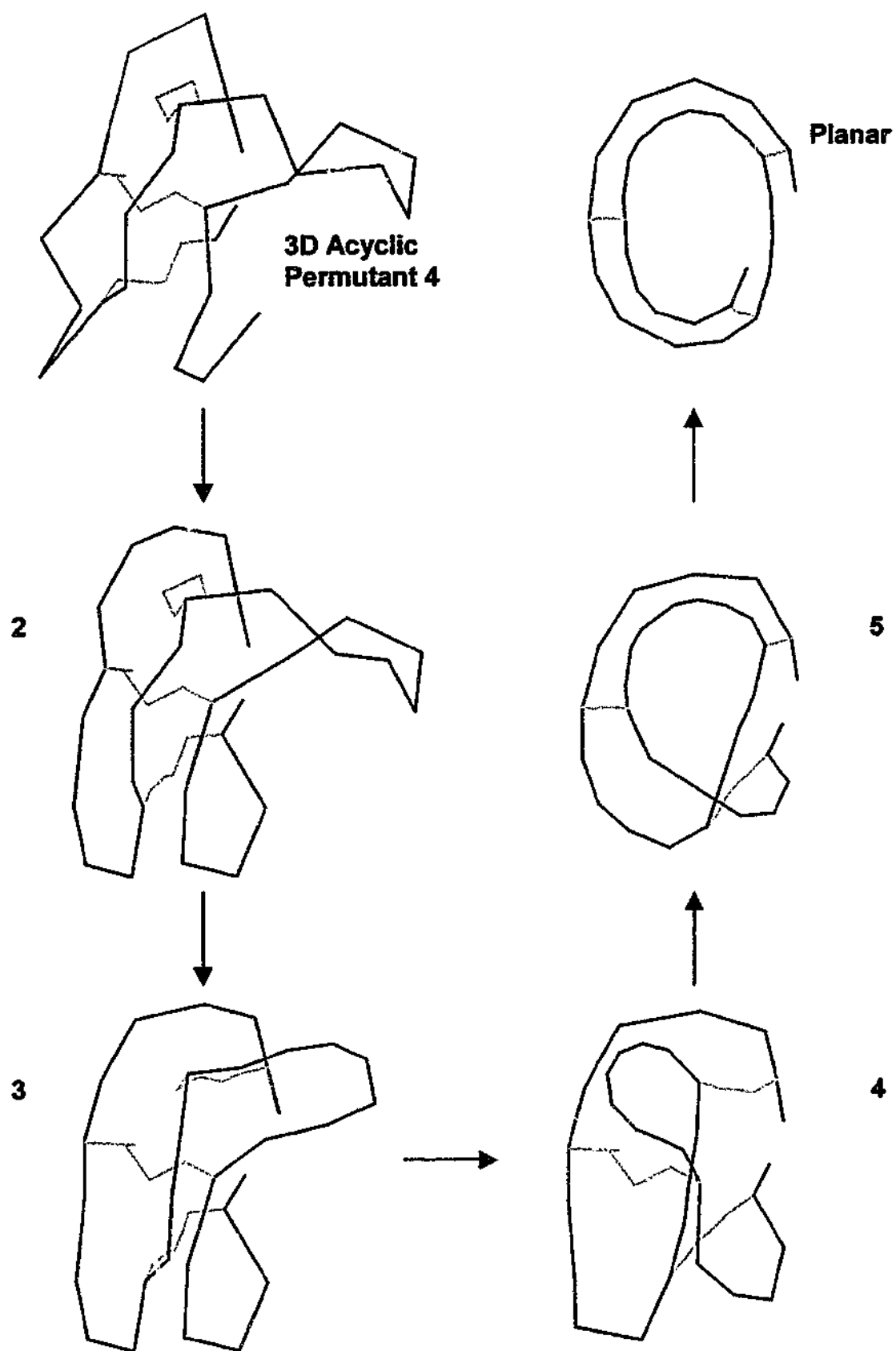
**Figure 3.5.** Animation of the unknotting of the 3D structure of acyclic permuted 1 to the planar structure. The backbone is shown in blue and the disulphide bonds are shown in gold.



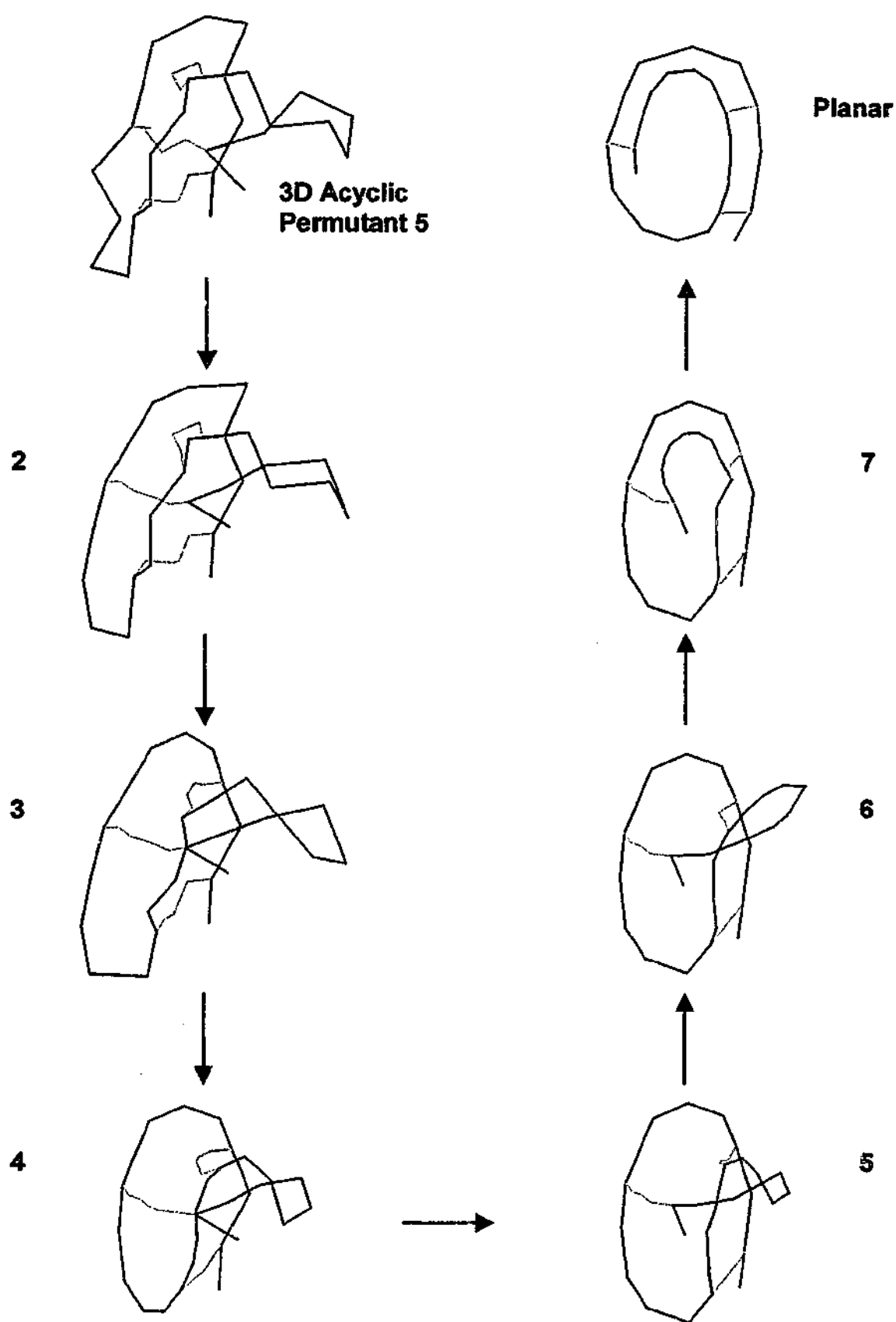
**Figure 3.6.** Animation of the unknotting of the 3D structure of acyclic permutant 2 to the planar structure. The backbone is shown in blue and the disulphide bonds are shown in gold.



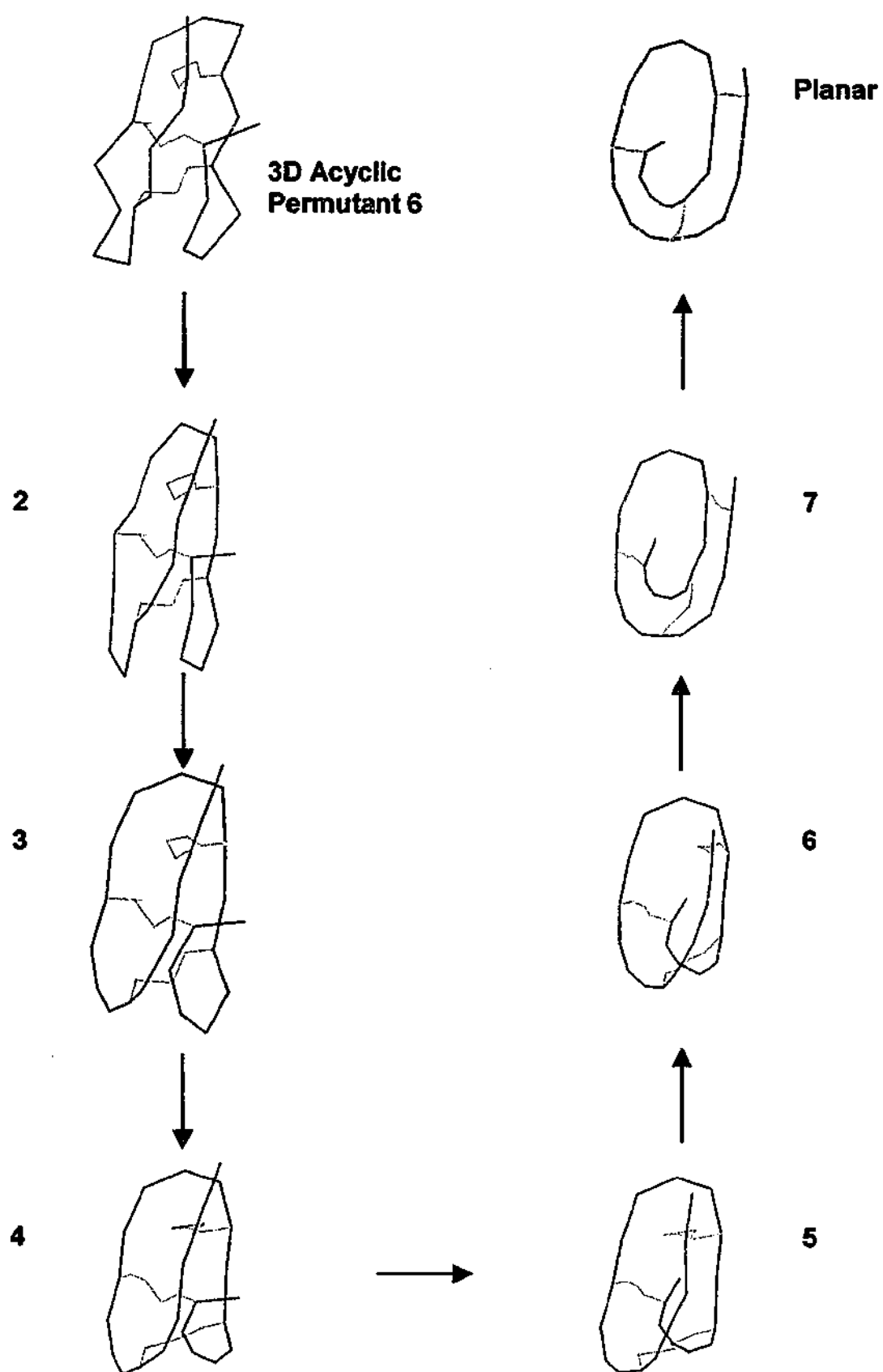
**Figure 3.7.** Animation of the unknotting of the 3D structure of acyclic permuted 3 to the planar structure. The backbone is shown in blue and the disulphide bonds are shown in gold.



**Figure 3.8.** Animation of the unknotting of the 3D structure of acyclic permutant 4 to the planar structure. The backbone is shown in blue and the disulphide bonds are shown in gold.



**Figure 3.9.** Animation of the unknotting of the 3D structure of acyclic permutant 5 to the planar structure. The backbone is shown in blue and the disulphide bonds are shown in gold.



**Figure 3.10.** Animation of the unknotting of the 3D structure of acyclic permutant 6 to the planar structure. The backbone is shown in blue and the disulphide bonds are shown in gold.



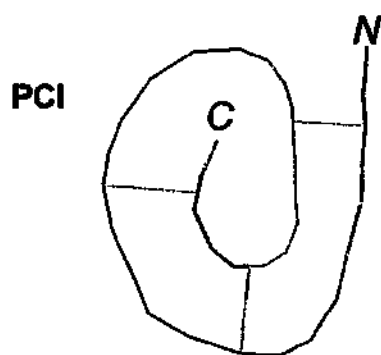
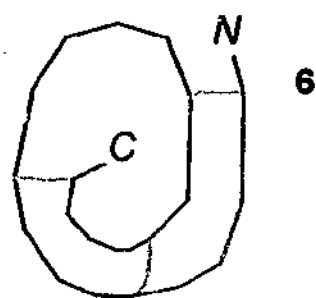
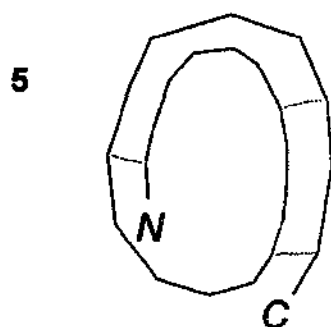
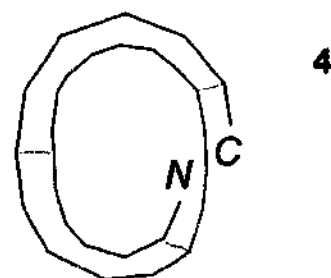
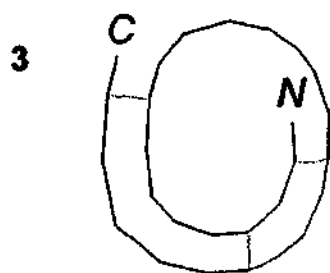
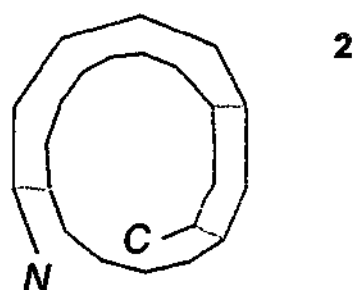
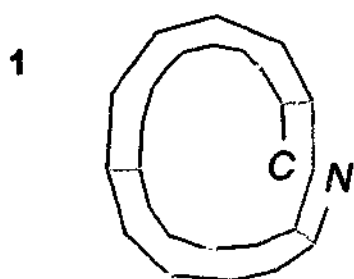


Figure 3.11. Comparison of the planar structures of the native acyclic permittants to the planar structure of PCI. The backbone of PCI is shown in green and of the acyclic permittants in blue. The disulphide bonds are shown in gold and the residues of the embedded ring in pink. The N- and C-termini are marked.



have been predicated from the sequential alignment, the animations of untying the backbones of the acyclic permutants provide a topological argument.

The animations also illustrate that unknotting the acyclic permutants, where the embedded ring of the cystine knot is present (Figures 3.6, 3.7, 3.9 and 3.10) is more complicated than where the ring is absent (Figures 3.5 and 3.8). In addition, when the ring is absent, there are more ways to form the planar structure, as the backbone virtually falls apart. When the ring is present, the backbone can only be untied by passing through the ring itself.

The relative ease of untying the two acyclic permutants where the ring is absent could explain the experimental finding (Daly and Craik, 2000) that these acyclic permutants do not form the native structure. The animations show that the three disulphide bonds and the ring act as four different topological restrictions and when one these, such as the ring, is removed, untying is much simpler, as backbone directionality is increased. Untying may be related to folding. The disulphides and the ring restrictions influence the folding pathway and removal of even one restriction, opens the possibilities for other pathways. In the final section of this chapter, the topological isomerism of cyclic cystine knots is considered.

### **3.5 Topological Isomerism of Kalata-B1**

Kalata-B1 and other cyclic plant polypeptides of the cystine knot family are topologically interesting, as they are nonplanar and chiral. The cyclic cystine knot peptides are thus the simplest known peptides with topological chirality, as well as being part of a minority of polypeptides that are topologically complex in any way. To ascertain and demonstrate the isomerism of cyclic cystine knot peptides, an animation of kinemages can be constructed to transform the 3D structure into a 2D reduced graph, which clearly illustrates nonplanarity and chiral isomerism. With the backbone and disulphide bonds forming edges, and the alpha carbons of the cystine residues acting as vertices, a kinemage animation for the scorpion neurotoxin, variant 3

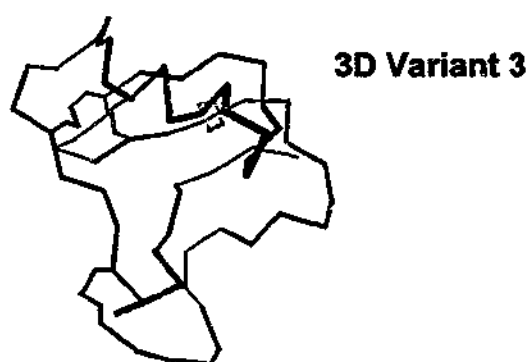
toxin from the North American scorpion *Centruroides sculpturatus* Ewing, has been created, by bending and stretching the backbone and the disulphide bonds (Mao, 1993). The animation is given in Figure 3.12 (kinemage scorpion.kin).

Viewing the animation for the scorpion neurotoxin, it can be seen that the polypeptide has D chirality, as the DG edge overlaps the CF edge. A similar animation was produced for kalata-B1 and is demonstrated in Figure 3.13 (kinemage kalata.kin). Regarding the animation for kalata-B1, it can be seen that the CF edge overlaps the BE edge, indicating that kalata-B1 too, has D chirality. Kinemages have been extremely useful in demonstrating topological isomerism, and indeed throughout this chapter in the structural consideration of cyclic cystine knotted peptides.

### 3.6 Summary

Topological considerations for the cyclic cystine knotted peptides have produced an interesting proposition for their study. A detailed examination of the secondary and tertiary structure, using kalata-B1 as a prototypic member of the cyclic cystine knot family, was conducted. This led to a proposition to analyse the structures of the complete set of six fragments of kalata-B1, in which each member consisted of the three disulphide bonds intact, but with a different segment of the backbone residues, namely sections or loops between successive cystine residues of the parent, omitted. The fragments were named "acyclic permutants" and the study of native and non-native variants of acyclic permutants represents a novel approach in the study of structure/activity relationships that is uniquely offered by cyclised polypeptides. Non-native versions of acyclic permutants are the focus of a portion of experimental work carried out in this thesis.

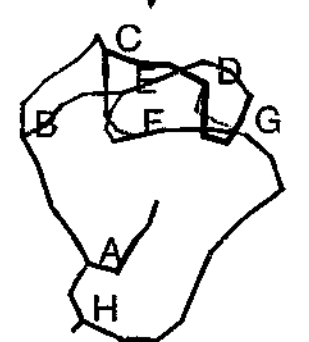
The six native versions of acyclic permutants were then used to indicate the potential location of the starting residue in plant synthesis. Creating animations of kinemages, the backbone of each native acyclic



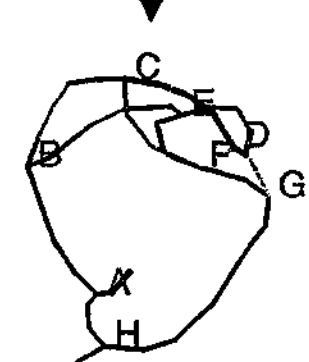
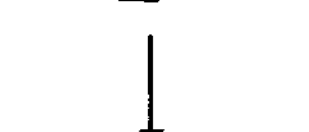
**Figure 3.12.** Animation showing the transformation of the 3D structure of the scorpion neurotoxin, variant 3, into the nonplanar reduced graph. The disulphide bonds are coloured and labeled. The overlap of the disulphide edge, DG, above the disulphide edge, CF, indicates D chirality.



2



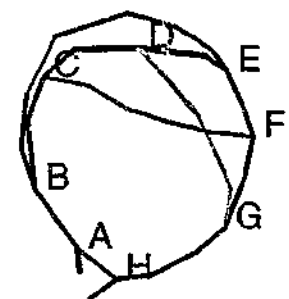
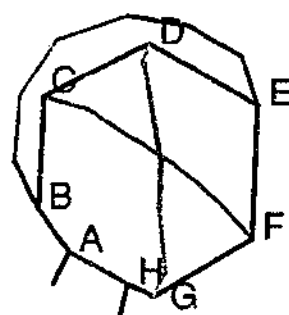
3



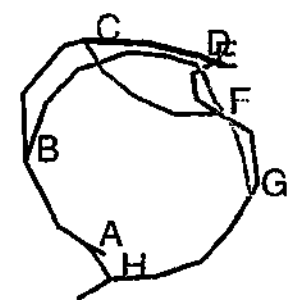
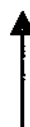
4



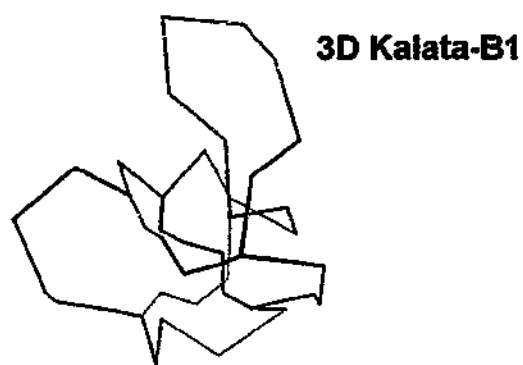
**Reduced Graph**



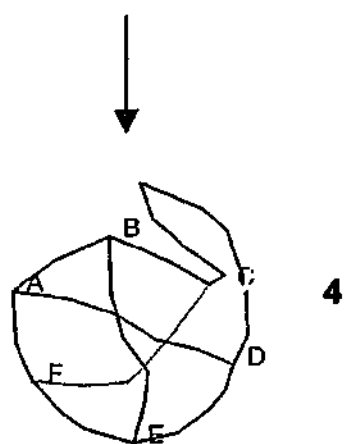
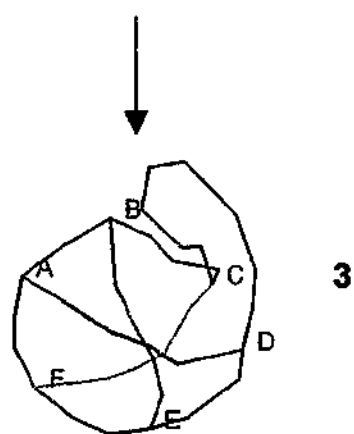
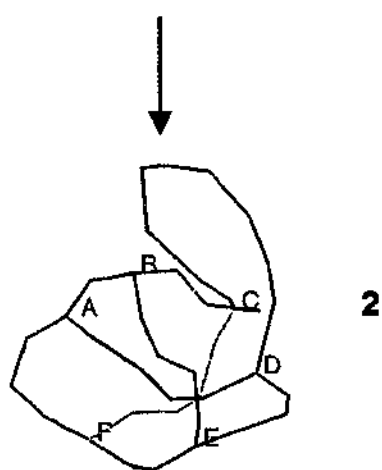
6



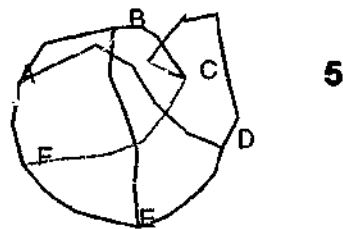
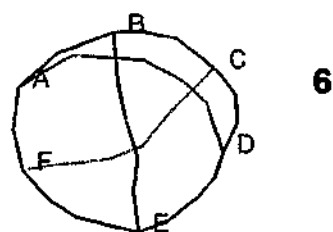
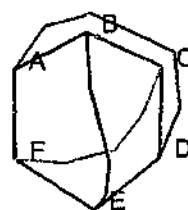
5



**Figure 3.13.** Animation showing the transformation of the 3D structure of kalata-B1 into the nonplanar reduced graph. The disulphide bonds are coloured and labeled. The overlap of the disulphide edge, CF, above the disulphide edge, BE, indicates D chirality.



**Reduced Graph**



permutant was progressively unthreaded to form the 2D planar structure, which was compared to the unthreaded planar structure of PCI, another well studied, non-cyclic member of the cystine knot family. The planar arrangement of the native acyclic permutant, in which residues of the disordered loop of kalata-B1 are absent, is identical to that of PCI. The exercise of unthreading the native acyclic permutants provides a topological rationale for the location of the starting residue being within the disordered loop.

Furthermore, the animations of the unthreading of native acyclic permutants indicate an increased order of complexity in unthreading, where the embedded ring of the cystine knot is present. If threading can be legitimately likened to unfolding, the ease of unthreading of the two native acyclic permutants, where the cystine knot is absent, may explain the experimental findings on the real equivalent fragments, which do not fold into the native form. Conversely, for the four native acyclic permutants, where the ring is present, the unthreading process is more difficult and highly restricted. This may explain the experimental findings on the real equivalent fragments, which do fold into the native configuration. Once folded, the tendency of these fragments to fall apart, is prevented. The kinemages were useful in demonstrating the unthreading process.

Kinemages were also effective in illustrating the topological isomerism of kalata-B1. An animation of kinemages was created to obtain the 2D reduced graph. This graph clearly shows that kalata-B1 has D chirality, similar to other known polypeptides with four disulphide bonds. Topological considerations for the study of kalata-B1 are an introduction to the study of the cyclic polypeptides in this thesis.

## **CHAPTER FOUR**

### **STUDIES OF THE SOLUTION**

### **STRUCTURE OF KALATA-B1**

## 4.1 Introduction

Experimental work for this thesis was commenced with continuing studies on the structure of kalata-B1. Because of the intriguing three dimensional (3D) fold of this cyclic biomolecule, it was of interest to further study the 3D structure. This chapter focuses on the application of homonuclear and heteronuclear NMR spectroscopy to study the structure of kalata-B1 under varying temperature and solvent environments. The results will be used to determine if information from  $^{13}\text{C}$  NMR spectroscopy could be used to refine and to predict structure of cyclic peptides. Furthermore, the effect of the hydrophobic solvent trifluoroethanol (TFE) on the structure of kalata-B1 is investigated through analysis of chemical shift data and simulated annealing structure calculations, which are based on distance data obtained from 2D-homonuclear experiments. The presence of TFE in the solvent environment is well known to induce significant structural changes and it was of importance to test the stability of the molecular framework of the cystine knot as represented by kalata-B1.

The results and their implication in relation to the structure of kalata-B1 are presented in the following pages, which begins with a description of materials and methods utilised. The numbering system used in the original elucidation of the structure of kalata-B1 (Saether *et al.*, 1995) is retained in this work.

## 4.2 Materials and Methods

An overview of methods used in this thesis is given in Chapter 2. In the following section, specific details relevant to the study of kalata-B1 are recorded.



#### 4.2.1 Materials

The kalata-B1 used in experiments was the same solution of kalata-B1 in an aqueous environment as that used for the determination of the 3D structure (Saether *et al.*, 1995). Half of this solution was sequentially diluted with MilliQ water and TFE to give two solutions of varying trifluoroethanol (TFE) content which were individually analysed. In the first dilution, half of the solution of kalata-B1 was diluted to give a 1.3 mmol kalata-B1 solution in 6% TFE. The solution was further diluted to give the second kalata-B1 solution of 0.68 mmol in 20% TFE. The remaining half of the original sample was diluted with deuterated water. These three samples were used for all the NMR experiments detailed below.

#### 4.2.2 NMR Experiments

NMR spectra were recorded on a Bruker ARX 500, except for select NOESY spectra, which were recorded on a Bruker DMX 750 Spectrometer. All two dimensional NMR spectra were acquired in the phase sensitive mode using time proportional phase incrementation for quadrature detection in the  $t_1$  dimension (Marion and Wüthrich, 1983).

TOCSY spectra (Brauschweiler and Ernst, 1983) were recorded using an MLEV-17 spin lock sequence (Bax and Davis, 1985) and a mixing time of 80 ms. NOESY (Jeener *et al.*, 1979) spectra were recorded with a mixing time of 200 ms. Solvent suppression for NOESY and TOCSY spectra was achieved using a modified WATERGATE sequence (Piotto *et al.*, 1992) in which two gradient pulses of 2 ms duration and 6 G cm<sup>-1</sup> strength were applied either side of a binomial 3-9-19 pulse. The spectra were acquired with 4096 complex data points in  $F_2$  and 400 or 600 increments in the  $F_1$  dimension with 48 or 64 scans per increment for the TOCSY spectra and 512 increments in the  $F_1$  dimension with 32 scans per increment for the NOESY spectra. Both

TOCSY and NOESY spectra were collected over a spectral width corresponding to 11 ppm for both dimensions.

HMQC (Bax and Subramanian, 1986) and HMQC-TOCSY (Bax and Lerner, 1986) spectra were recorded with detection of  $^1\text{H}$  in  $F_2$  with a GARP composite pulse and decoupling of the  $^{13}\text{C}$  nuclei (Shaka *et al.*, 1985). Suppression of the water signal was achieved by selective irradiation during the relaxation delay. The HMQC-TOCSY spectrum was recorded using an MLEV-17 spin lock sequence (Bax and Davis, 1985) and an 80 ms mixing pulse. HMQC and HMQC-TOCSY spectra were acquired with 2048 complex data points in  $F_2$  and 158 increments in the  $F_1$  dimension with 256 scans per increment for the HMQC spectrum and 128 increments in the  $F_1$  dimension with 512 scans per increment for the HMQC-TOCSY spectrum. HMQC and HMQC-TOCSY spectra were collected over a spectral width corresponding to 11 ppm in the  $F_2$  dimension and 70 ppm in the  $F_1$  dimension.

TOCSY and NOESY spectra were obtained at a temperature of 303 K on the 6% TFE sample and at 333 K on the 20% TFE sample. In addition, a TOCSY spectrum at 280 K and a NOESY spectrum at 300 K were recorded on the 20% TFE sample. HMQC and HMQC-TOCSY spectra were obtained at 333 K on the 20% TFE sample and an HMQC spectrum on the 6% TFE sample was acquired at 303 K. Constant temperature was maintained using a temperature controlled stream of air which was supplied by a Bruker BCU05 refrigeration unit and controlled by a B-VT 2000 control unit. The  $^1\text{H}$  scale was referenced to residual water solvent peaks, calibrated externally using 4,4-dimethyl-4-silapentane-1-sulfonate (DDS) at 0.00 ppm. The  $^{13}\text{C}$  scale was calibrated internally relative to TFE.

The data were processed on a Silicon Graphics (4D/30) computer using the UXNMR 940501.2 software package (Bruker). Both  $F_1$  and  $F_2$  dimensions for NOESY and TOCSY spectra were zero filled to 2048 real data points with both dimensions multiplied by a square sine-bell or sine-bell function phase-shifted by 60 - 90° prior to Fourier transformation.  $F_1$  and  $F_2$  dimensions for HMQC and HMQC-TOCSY were zero filled to 1024 real data points with the

F<sub>2</sub> dimension multiplied by a sine-bell function phase-shifted by 90° and the F<sub>1</sub> dimension multiplied by an exponential function with line broadening of 5.0. Polynomial baseline correction was applied to the spectral data to improve the appearance of the spectra. With the data processed, distance restraints for structure calculations could be determined.

#### 4.2.3 Structural Restraints

Distance restraints were determined from the NOESY spectrum on the 20% TFE solution of kalata-B1 acquired at 300 K on a 750 MHz spectrometer. Peak volumes were visually classified as strong, medium, weak or very weak and upper bound interproton distance restraints of 2.7, 3.5, 5.0 and 6.0 Å, respectively, were assigned (Williamson *et al.*, 1985; Clore *et al.*, 1986a). Pseudoatom corrections were applied to non-stereospecifically assigned methylene and methyl protons (Wüthrich *et al.*, 1983) so that for distances involving methylene protons, 1.0 Å was added to the upper limit and for distances involving methyl protons, 1.5 Å was added to the upper limit. The classified distance restraints as a set were used as input for dynamical simulated annealing structure calculations, which were conducted as follows.

#### 4.2.4 Structure Calculations

Three dimensional structures were generated using simulated annealing and energy minimisation protocols in the program XPLOR 3.1 (Brünger, 1992). The calculations were carried out on a Silicon Graphics (4D/30) computer. Initially, an *ab initio* simulated annealing protocol (Nilges *et al.*, 1988) was used to generate a set of 50 structures, starting from template structures with randomised  $\phi$  and  $\psi$  angles and extended sidechains. The backbone was cyclised throughout the calculations. At this stage, the disulphide bonds were included as pseudo-NOE restraints and the simulated annealing protocol

consisted of 20 ps of high temperature molecular dynamics at 1000 K with a low weighting on the repel force constant and the NOE distance restraints.

This was followed by a further 10 ps of high temperature dynamics with an increased force constant on the NOE distance restraints. The disulphide bonds were formally bonded and the dihedral force constant increased prior to cooling the system to 0 K and increasing the repel force constant over 15 ps of dynamics. The NOE restraints were checked for violations and ambiguous cross peaks were resolved on the basis of interproton distances in the family of structures. A further 50 structures were calculated and these structures were refined using the conjugate gradient Powell algorithm with 1000 cycles of energy minimisation (Clare *et al.*, 1986b) and a refined force field based on the program CHARMM (Brooks *et al.*, 1983).

No hydrogen-bonding restraints or dihedral angle restraints were incorporated in the calculations. All peptide bonds were defined as *trans* except where experimental data indicated the *cis* conformation as in the case of certain proline residues. Structures were analysed and displayed using INSIGHT (Biosym Technologies, San Diego, CA). These structures together with NMR data produced some interesting results for discussion.

## 4.3 Results and Discussion

### 4.3.1 Proton Resonances

In general, the chemical shifts of kalata-B1 in the TFE environment at the various temperatures were similar to those of kalata-B1 in the aqueous environment. This greatly assisted assignment. As with the spectral data of kalata-B1 in water (Saether *et al.*, 1995), the resonances of the 2D data are well dispersed, so that the amino acid spin systems and the through-space cross peaks are readily recognised and assigned. The sets of <sup>1</sup>H assignments

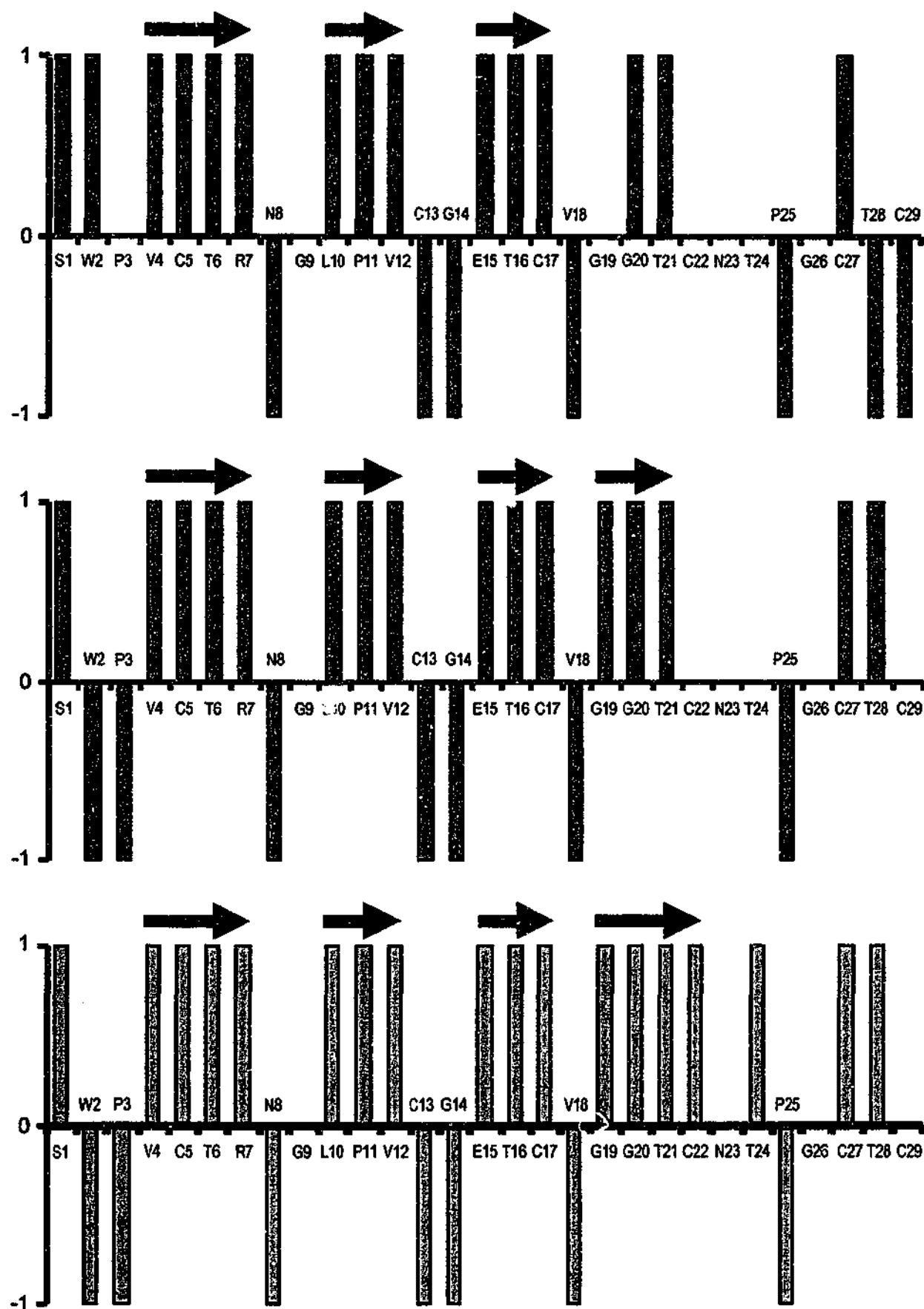
of kalata-B1 under the different temperature and solvent conditions are given in Tables AI.1 to AI.4 in Appendix I.

Similarly, complete cycles of  $\alpha$ H-NH sequential connectivity were also obtained in the fingerprint regions of NOESY spectra under the varying TFE concentration and temperature conditions. Only at the proline residues were these cycles broken. However, NOE cross peaks between  $\alpha$ H of preceding residues and the appropriate proline protons allowed for completion of the cycles. Diagrams of the assignments of the amino acid spin systems of the fingerprint regions of TOCSY data, and also the complete cycles of  $\alpha$ H-NH sequential connectivity of the fingerprint regions of the NOESY data for both kalata-B1 in 6% and 20% TFE are illustrated in Figures AII.1 to AII.4 of Appendix II. With the proton data assigned, information on local secondary structure changes in the TFE environment could be assessed. In the first instance, CSI values of alpha protons were calculated and examined.

#### 4.3.1.1 CSI Values from Alpha Proton Chemical Shifts

CSI values based on  $\alpha$ H chemical shifts were calculated for kalata-B1 in water (Saether *et al.*, 1995), 6% TFE and 20% TFE. These CSI values are illustrated in Figure 4.1. Solid arrows indicate the predicted location of  $\beta$ -strand. The  $\beta$ -strand structure is correctly predicted for three of the four known  $\beta$ -strands of kalata-B1 in water by the CSI values. These include the  $\beta$ -strands over residues V4 to R7, L10 to V12 and E15 to C17. Similarly, three of the four  $\beta$ -turns centred on residues N8/G9, V18/G19 and P25/G26 are predicted by a differentiated "-1" CSI value occurring for the first residue in the sequence of the turn. The  $\beta$ -bulge involving residues C13/G14 is indicated by two "-1" CSI values between two  $\beta$ -strands. Neither the  $\beta$ -strand involving residues C27 to C29 nor the  $\beta$ -turn centred on W2/P3 are predicted by CSI values for kalata-B1 in water for which the 3D fold has been determined.

The presence of TFE, however, does seem to induce changes in CSI values in certain regions of kalata-B1. These CSI values are identical for the



**Figure 4.1.** CSI values of alpha protons at around 300 K for kalata-B1 in water (Saether *et al.*, 1995) (blue bars), in 6% TFE (pink bars) and in 20% TFE (purple bars). The prediction of  $\beta$ -strand is indicated by solid arrows.

environment of water and TFE over residues V4 to V18 and differ for many residues over S1 to P3 and G19 to C29. The most striking change is the prediction of the formation and development of  $\beta$ -strand structure in the fairly featureless connecting region of kalata-B1 in water. In 6% TFE, a  $\beta$ -strand is predicted over residues G19, G20, T21. As G19 is part of a  $\beta$ -turn in kalata-B1 in the aqueous environment, this result implies that this  $\beta$ -turn, formally centred on V18/G19, is perhaps much sharper, as G19 of kalata-B1 in 6% TFE forms part of a  $\beta$ -strand. In 20% TFE, the  $\beta$ -strand is predicted to extend by encompassing also C22, and a "+1" CSI value appears for T24. The CSI values for both are zero for kalata-B1 in water. Furthermore, there are significant changes in the region T28 to P3, where the CSI indices seem to almost correctly predict the  $\beta$ -strand over C27 to C29 and the  $\beta$ -turn centred on W2/P3, although no clear local secondary structure is evident from CSI values in this region.

While the accuracy of CSI values in predicting local secondary structure is a question (typically 90 - 95 %, Wishart *et al.*, 1992), the results of CSI analysis for kalata-B1 in TFE suggest the formation of  $\beta$ -strand in the connecting region of kalata-B1, and changes in the structure over residues T28 to P3. To determine any trends with regard to TFE concentration, amide proton chemical shifts were considered.

#### **4.3.1.2 Effect of TFE Concentration on Amide Proton Chemical Shifts**

Amide proton chemical shifts are sensitive to hydrogen bonding changes. Even slight local structural changes can affect both the local and global hydrogen bonding network. Addition of TFE to the aqueous environment of kalata-B1 seems to result in shifting of many amide proton chemical shifts. Some of these shifts are minimal, being less than 0.05 ppm, while some of these are quite pronounced, being greater than 0.1 ppm.

In comparison with the aqueous environment at approximately 300 K, the following shift trends were observed for kalata-B1 in 6% TFE. Large shift values are given in brackets (ppm).

**Large downfield shift** - V4 (0.13), T6 (0.13), C17 (0.24), N23 (0.65),  
T24 (0.36);

**Medium downfield shift** - S1, L10, G20, T28;

**Large upfield shift** - W2 (0.21), C13 (0.15), G19 (0.14);

**Medium upfield shift** - C29;

**Minimal shift** - C5, R7, N8, V12, G14, E15, V18, T21, C22, C27;

**No shift** - G9, T16, G26.

As the concentration of TFE was increased from 6% to 20%, the following shift trends were observed. Large shift values are given in brackets (ppm).

**Large downfield shift** - C17 (0.14), N23 (0.3);

**Medium downfield shift** - S1, V4, T6, R7, L10, V12, E15, G20, C27, T28;

**Large upfield shift** - W2 (0.15);

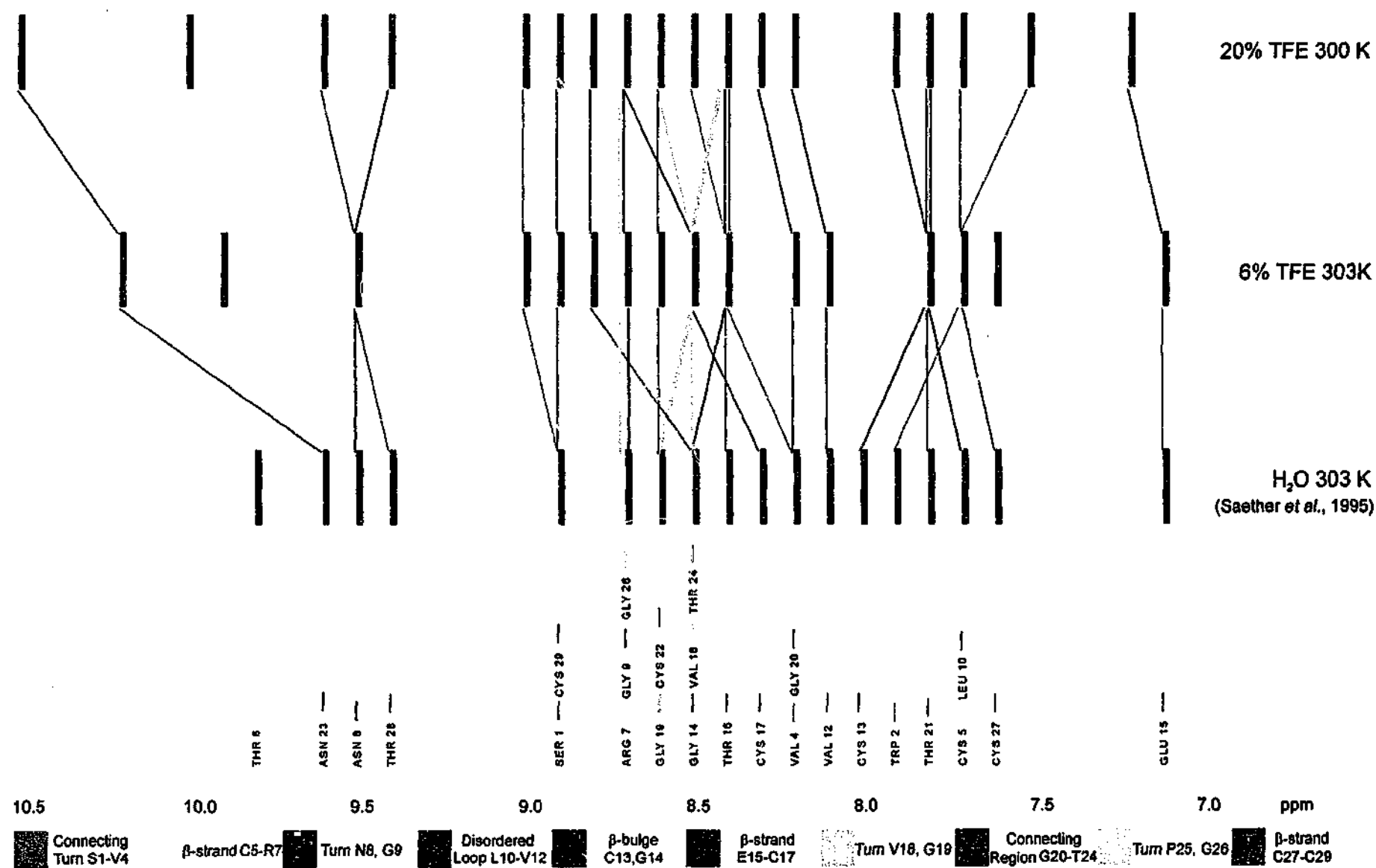
**Medium upfield shift** - C13, G19;

**Minimal shift** - C5, N8, G14, V18, T21, C22, C29;

**No shift** - G9, T16, T24, G26.

The amide proton chemical shift values and their movement with TFE concentration at around 300 K are shown in Figure 4.2. The shift changes are indicated by coloured lines, which are coded according to the region of the molecule. Obvious systematic changes are not readily apparent from a glance at this diagram. However, in the context of results from CSI values obtained in previous section and the 3D fold of kalata-B1, several observations can be made.





**Figure 4.2.** Amide proton chemical shifts in ppm (thin black bars) of kalata-B1 at varying TFE concentration. Colored lines link chemical shifts of the same residue. The colors of these lines are coded according to the region of the molecule.

Firstly, the residues with large, consistent downfield amide proton shift with increasing TFE concentration are C17, N23 and T24. These residues occur on either side of a  $\beta$ -strand, which is predicted by CSI values to form in TFE. The residue G19 experiences a reasonably significant consistent, upfield amide proton shift with increasing TFE concentration. CSI values suggest that this residue changes from being part of a  $\beta$ -turn in the aqueous environment to become the first residue of a new  $\beta$ -strand formed in TFE.

Another notable observation is that virtually all the large and medium shifts with increasing TFE concentration are downfield. Furthermore, most of these occur in the regions of the connecting turn, S1 to V4, the connecting region, G20 to T24 and the  $\beta$ -strand of C27 to C29. These particular regions have been implicated as regions of structural change by CSI values. Other more minor observations are also evident.

The reasonably large, consistent upfield amide proton chemical shift involving C13 and the consistent downfield shifts involving T6, R7, L10, V12 and E15 with increasing TFE concentration could be explained by slight changes in the disordered region and the vicinity of the  $\beta$ -bulge. Lastly, minimal or no change is observed for residues of the  $\beta$ -turns centred on N8/G9 and P25/G26, indicating that the fold is not altered at these points.

Overall, it would appear from the CSI values and the data on amide proton chemical shifts that the fold of kalata-B1 remains largely intact in the TFE environment. A significant structural change seems to occur in the connecting region. A smaller change is evident around the  $\beta$ -bulge and disordered region, and the predominance of the proportion of downfield shifts with increasing TFE concentration could be indicative of a strengthening of the internal hydrogen bond network of kalata-E1 in TFE. To gain further insight into the structural changes that arise in TFE, the effect of temperature on amide proton chemical shifts was examined for evidence of the presence of hydrogen bonded structure.

#### 4.3.1.3 Effect of Temperature on Amide Proton Chemical Shifts

Temperature coefficients less than 3 ppb/K may indicate the presence of hydrogen bonded structure. Temperature coefficients from amide proton chemical shifts were calculated for each residue of kalata-B1 in 20% TFE using data from 280 K, 300 K and 333 K. These temperature coefficients are tabulated in Table 4.1, and a diagram illustrating the movement of amide protons chemical shifts with temperature is given in Figure 4.3.

Non-linear variation is obtained for almost half of the residues. Of the residues where the variation is linear, the following is indicated by the temperature coefficients. Only the V18 amide proton, with a low temperature coefficient, appears to be involved in hydrogen bonding or is protected from the solvent. High temperature coefficients greater than 6 ppb/K are obtained for V4 and C22, possibly suggesting that the amide protons of these residues are completely exposed to the solvent. The remaining residues with linear variation, namely N8, C13, T16, C17, G20, T24, G26, have intermediate temperature coefficients, implying that conformational equilibrium between hydrogen bonded or solvent exposed protons is occurring.

Although temperature coefficients are not a reliable measure of hydrogen bonding, there may be merit in these results. The extent of their accuracy may be only realised once the 3D structure of kalata-B1 in TFE is established. Having completed the analyses of the proton data, information from  $^{13}\text{C}$  data was examined and is described below.

#### 4.3.2 $^{13}\text{C}$ Chemical Shifts and Prediction of Secondary Structure.

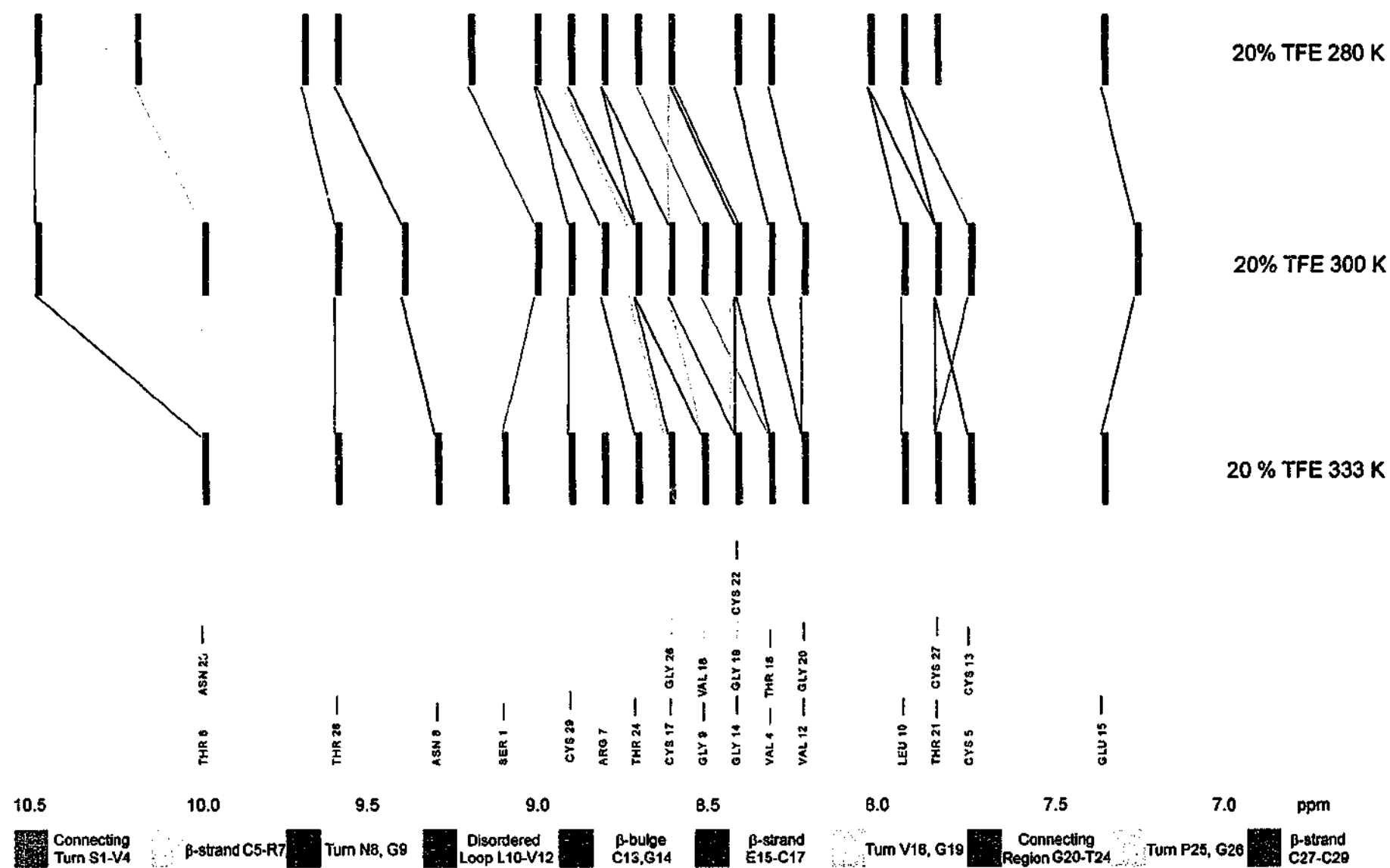
$^{13}\text{C}$  chemical shifts from  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$  and carbonyl  $^{13}\text{C}$  can give information on the presence of secondary structural elements with a fairly high degree of accuracy (Wishart and Sykes, 1994). In line with the objectives of this study,  $^{13}\text{C}$  chemical shifts of kalata-B1 in 6% and 20% TFE were assigned to determine how well the data from  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shift values

**Table 4.1.** Temperature Coefficients (ppb/K) measured for kalata-B1 in 20% TFE.

	Residue	Temperature Coefficient (ppb/K)
1	SER	*
2	TRP	-
3	PRO	-
4	VAL	8.0
5	CYS	*
6	THR	*
7	ARG	*
8	ASN	4.7
9	GLY	6.7
10	LEU	*
11	PRO	-
12	VAL	*
13	CYS	4.7
14	GLY	*
15	GLU	*
16	THR	4.9
17	CYS	4.1
18	VAL	2.1
19	GLY	-
20	GLY	4.3
21	THR	*
22	CYS	6.8
23	ASN	*
24	THR	6.1
25	PRO	-
26	GLY	5.3
27	CYS	*
28	THR	*
29	CYS	*

\* Denotes non-linear variation of NH chemical shift

- Denotes absent NH or proline (no NH)



**Figure 4.3.** Amide proton chemical shifts in ppm (thin black bars) of kalata-B1 at varying temperature. Colored lines link chemical shifts of the same residue. The colors of these lines are coded according to the region of the molecule.

predict secondary structure of kalata-B1 and if there are any changes in the TFE solvent environment.

Figures 4.4 and 4.5 show the assigned HMQC spectra of kalata-B1 in 6% TFE at 303 K and in 20% TFE at 333 K, respectively. Dispersion in the  $^{13}\text{C}$  dimension is generally good and, based on the proton assignments, it is possible to determine  $^{13}\text{C}$  shifts for most of the resonances. Support for many assignments were obtained from the HMQC-TOCSY spectrum of kalata-B1 in 20% TFE at 333. This spectrum is displayed with assignments in Figure 4.6. The HMQC-TOCSY spectrum is useful in resolving ambiguity due to multiple overlapping proton resonances. For example, the HMQC spectrum has five peaks with similar  $\alpha$ -proton shifts around 5.2 ppm, these being Leu-10, Pro-11, Thr-6, Cys-17 and Cys-5. The HMQC-TOCSY spectrum clearly distinguishes between the various possible assignments. Tables 4.2 and 4.3 summarise the  $^{13}\text{C}$  chemical shift assignments for kalata-B1 in 6% TFE at 303K and in 20% TFE at 333 K, respectively.

The method for determining secondary structure by chemical shift index was applied (Wishart and Sykes, 1994). The results are presented in the Figure 4.7. From the diagram, it can be seen than in the 6% TFE, only  $^{13}\text{C}_\beta$  values could be used for prediction, as many  $^{13}\text{C}_\alpha$  chemical shifts are not observed, probably due to bleaching by the water signal. Furthermore, of the available  $^{13}\text{C}_\alpha$  chemical shifts, several CSI values are not in accord with the known structure of kalata-B1. Although all care was taken in the assignment process, the potential for error with  $^{13}\text{C}_\alpha$  assignments is greater without confirmatory data specifically from an HMQC-TOCSY spectrum of kalata-B1 in 6% TFE. In this case,  $^{13}\text{C}_\alpha$  data has not been particularly useful. However, CSI values from  $^{13}\text{C}_\beta$  chemical shifts in 6% TFE were more informative.

The CSI values from  $^{13}\text{C}_\beta$  chemical shifts predict the presence of  $\beta$ -strands over residues V4 to R7 and G19 to C22, and a termination point at N8. These results are similar to those predicted by CSI values based on  $^1\text{H}_\alpha$ , including the formation of the  $\beta$ -strand in TFE not evident in the aqueous environment. The  $\beta$ -strands over residues L10 to V12 and E15 to C17,

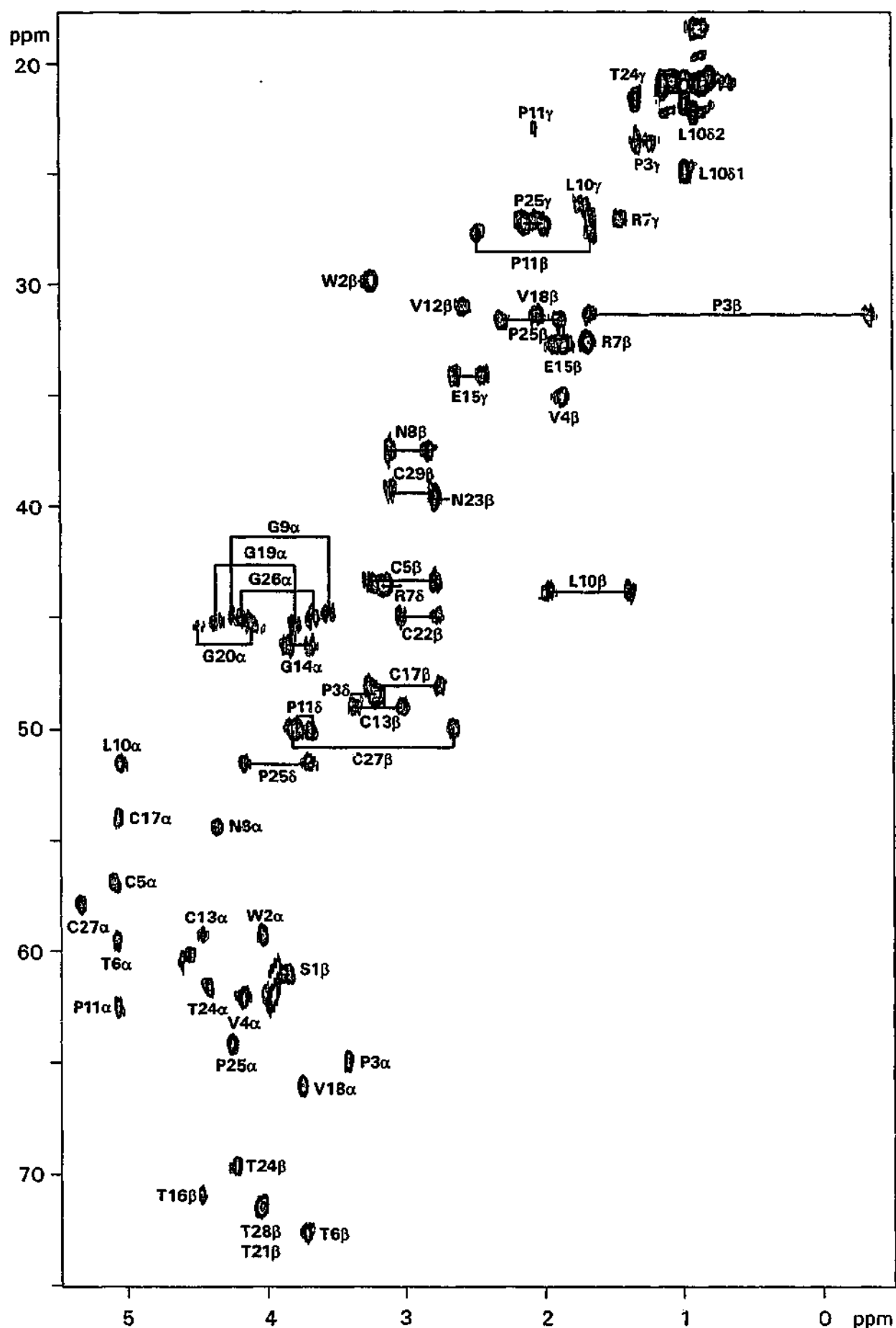


Figure 4.4. Assigned HMQC spectrum of kalata-B1 in 6% TFE at 303K.

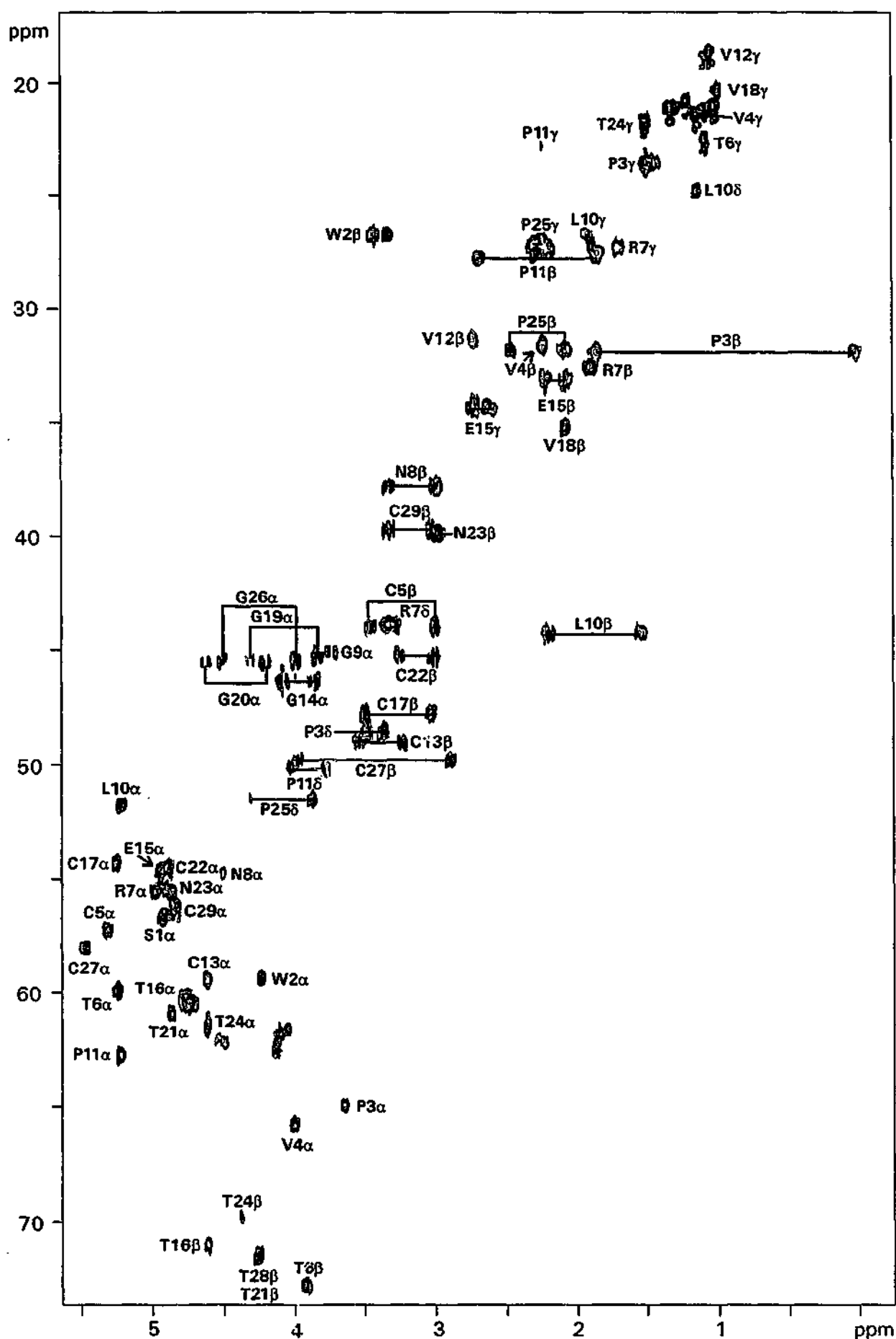


Figure 4.5. Assigned HMQC spectrum of kalata-B1 in 20% TFE at 333K.



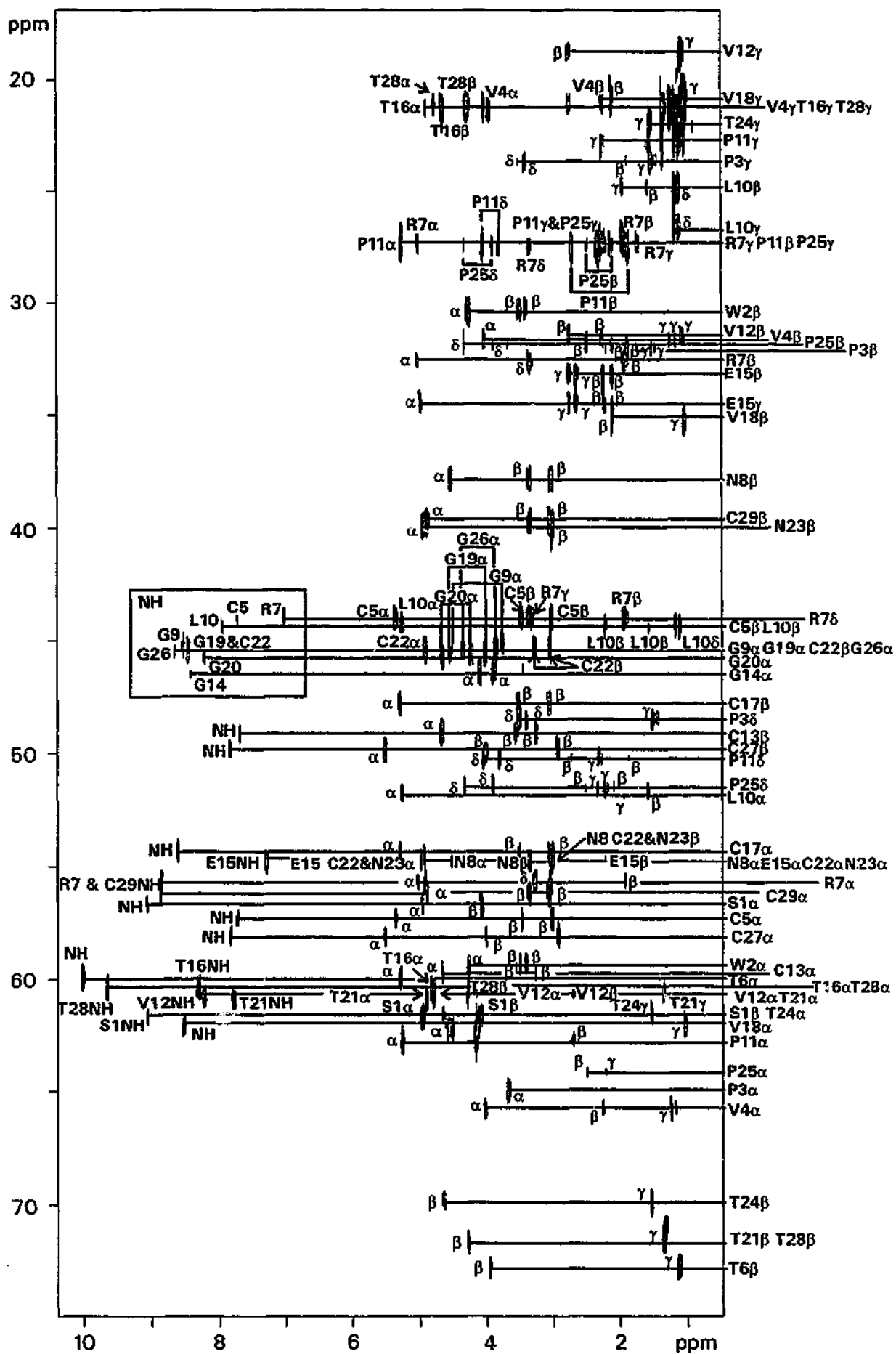


Figure 4.6. Assigned HMQC-TOCSY spectrum of kalata-B1 in 20% TFE at 333K.

Table 4.2.  $^{13}\text{C}$  chemical shift assignments (ppm) for kalata-B1 in 6% TFE at 303 K.

	RESIDUE	$^{13}\text{C}_\alpha$	$^{13}\text{C}_\beta$	OTHER	
1.	SER	*	61.0		
2.	TRP	59.1	30.0		
3.	PRO	64.7	31.4	$\text{C}_\gamma$ 23.7	$\text{C}_\delta$ 48.6
4.	VAL	62.0	35.0	$\text{C}_{\gamma 1}$ - $\text{C}_{\gamma 2}$ -	
5.	CYS	58.8	43.4		
6.	THR	59.4	72.7	$\text{C}_\gamma$ -	
7.	ARG	*	32.7	$\text{C}_\gamma$ 27.2	$\text{C}_\delta$ 43.7
8.	ASN	54.4	37.6		
9.	GLY	44.8			
10.	LEU	51.1	43.8	$\text{C}_\gamma$ 26.4	$\text{C}_{\delta 1}$ 25.0 $\text{C}_{\delta 2}$ 22.0
11.	PRO	62.8	27.8	$\text{C}_\gamma$ 23.0	$\text{C}_\delta$ 50.1
12.	VAL	*	31.1	$\text{C}_{\gamma 1}$ - $\text{C}_{\gamma 2}$ -	
13.	CYS	59.2	48.9		
14.	GLY	46.3			
15.	GLU	*	32.9	$\text{C}_\gamma$ 34.2	
16.	THR	60.0	70.8	$\text{C}_\gamma$ -	
17.	CYS	54.0	48.0		
18.	VAL	65.9	31.5	$\text{C}_{\gamma 1}$ - 22.0 $\text{C}_{\gamma 2}$ - 21.2	
19.	GLY	45.3			
20.	GLY	45.3			
21.	THR	*	71.4	$\text{C}_\gamma$ -	
22.	CYS	*	45.0		
23.	ASN	*	39.5		
24.	THR	61.5	69.4	$\text{C}_\gamma$ 21.8	
25.	PRO	64.1	31.7	$\text{C}_\gamma$ 27.4	$\text{C}_\delta$ 51.5
26.	GLY	45.1			
27.	CYS	57.9	49.9		
28.	THR	60.0	71.4	$\text{C}_\gamma$ -	
29.	CYS	*	39.4		

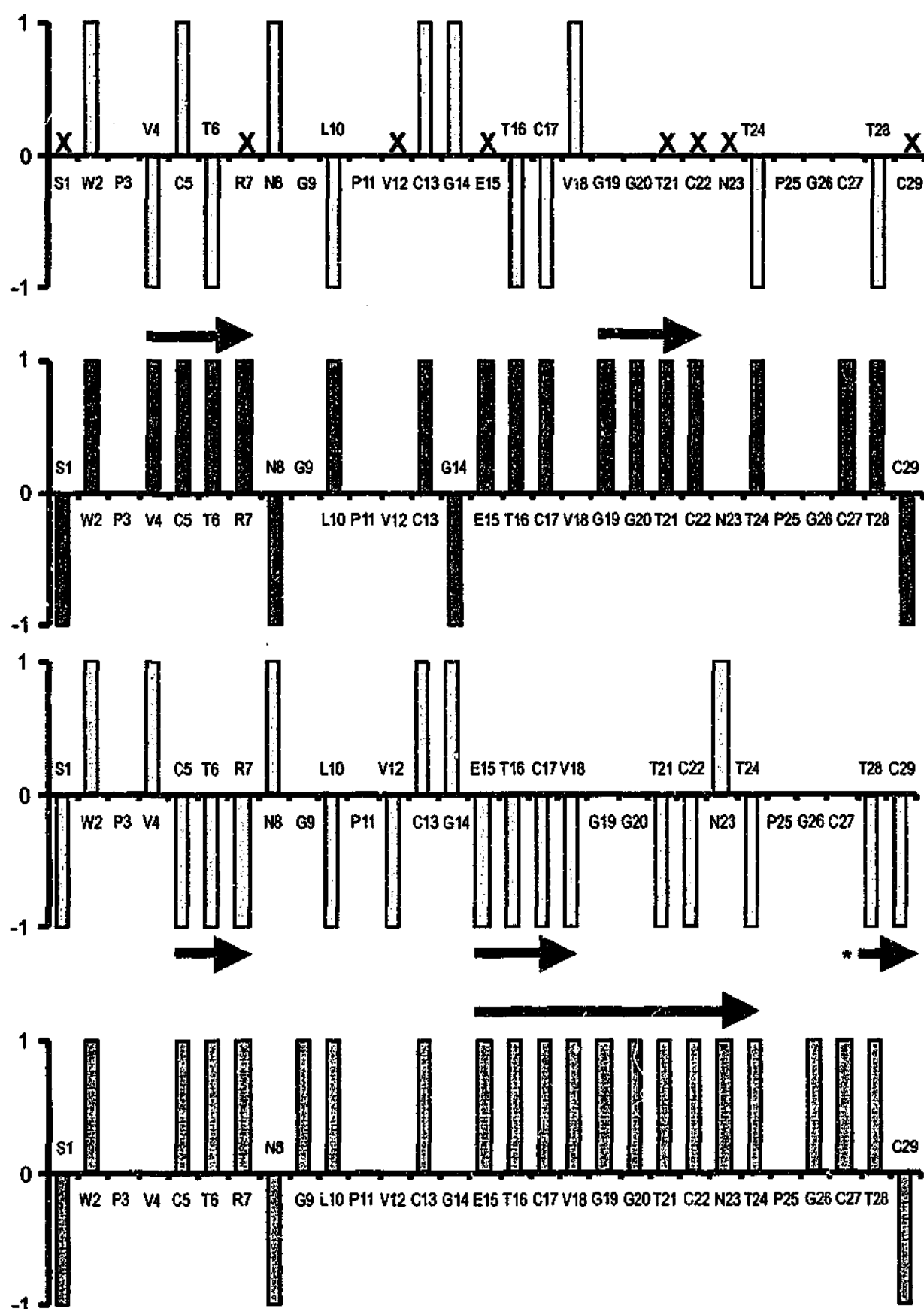
\* Denotes unassigned due to interference from water signal in  $^1\text{H}$  dimension

- Denotes omitted assignment due to ambiguity from overlap

Table 4.3.  $^{13}\text{C}$  chemical shift assignments (ppm) for kalata-B1 in 20% TFE at 333 K.

	RESIDUE	$^{13}\text{C}_\alpha$	$^{13}\text{C}_\beta$	OTHER	
1.	SER	56.6	61.6		
2.	TRP	59.4	30.2		
3.	PRO	65.0	31.9	$\text{C}_\gamma$ 23.5	$\text{C}_\delta$ 48.5
4.	VAL	65.8	31.8	$\text{C}_{\gamma 1}$ 21.2 $\text{C}_{\gamma 2}$ 20.8	
5.	CYS	57.2	44.2		
6.	THR	60.0	72.7	$\text{C}_\gamma$ 22.4	
7.	ARG	55.5	32.7	$\text{C}_\gamma$ 27.4	$\text{C}_\delta$ 44.9
8.	ASN	54.7	37.8		
9.	GLY	45.0			
10.	LEU	51.8	44.4	$\text{C}_\gamma$ 27.7	$\text{C}_\delta$ 24.9 deltas are coincident
11.	PRO	62.8	27.8	$\text{C}_\gamma$ 22.8	$\text{C}_\delta$ 50.2
12.	VAL	60.8	31.4	$\text{C}_\gamma$ 18.6 gammas are coincident	
13.	CYS	59.4	49.1		
14.	GLY	46.3			
15.	GLU	54.6	33.1	$\text{C}_\gamma$ 34.4	
16.	THR	60.3	71.0	-	
17.	CYS	54.3	47.8		
18.	VAL	62.1	35.3	$\text{C}_\gamma$ 20.3 gammas are coincident	
19.	GLY	45.5			
20.	GLY	45.5			
21.	THR	61.0	71.5	-	
22.	CYS	54.4	45.2		
23.	ASN	55.5	40.0		
24.	THR	61.5	69.7	$\text{C}_\gamma$ 21.6	
25.	PRO	64.1	31.8	$\text{C}_\gamma$ 27.2	$\text{C}_\delta$ 51.5
26.	GLY	45.4			
27.	CYS	57.9	49.7		
28.	THR	60.3	71.5	-	
29.	CYS	52.2	39.9		

- Denotes omitted assignment due to ambiguity from overlap



**Figure 4.7.** CSI values of carbon atoms for kalata-B1 in 6% TFE at 303 K (pink bars) and in 20% TFE at 333 K (purple bars). Those based  $^{13}\text{C}_\alpha$  are represented by bars of a lighter shade and those based on  $^{13}\text{C}_\beta$  by bars of a darker shade. The prediction of  $\beta$ -strand is indicated by solid arrows. Crosses represent absent CSI values. The  $\beta$ -strand marked with a star includes residue S1.

present in aqueous solution and predicted by  $^1\text{H}_\alpha$  CSI values, are not predicted by  $^{13}\text{C}_\beta$  CSI values. However, if the criterion for  $\beta$ -strand were to be altered to three or more "+1" CSI values, the  $\beta$ -strand from E15 to C17 would be predicted. The failure to indicate the  $\beta$ -strands may be due to some inaccuracy in prediction based on  $^{13}\text{C}_\beta$  chemical shifts and/or there may be some change arising in these regions of kalata-B1 in TFE. The later proposition is not unlikely, given that some change is indicated in the disordered region of L10 to V12 and around the  $\beta$ -bulge at C13/G14 by NH chemical shift movements in TFE. The CSI data on  $^{13}\text{C}$  nuclei for kalata-B1 in 20% TFE again presents some discord between  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$  and  $^1\text{H}_\alpha$  CSI values, although there is also some agreement.

In comparing the CSI values based for kalata-B1 in 20% TFE with  $^1\text{H}_\alpha$  CSI values, attention needs to be drawn to the fact that the temperature of the  $^{13}\text{C}$  CSI analysis is much higher at 333 K than the  $^1\text{H}_\alpha$  CSI values at 300K. As a consequence, it is difficult to ascertain whether differences in prediction are due to temperature or other factors. Nevertheless, there is more confidence in accuracy from the  $^{13}\text{C}$  analyses of kalata-B1 in 20% TFE as assignments in 20% TFE are largely supported by HMQC-TOCSY data. The CSI data based on  $^{13}\text{C}_\alpha$  predict  $\beta$ -strands over residues C5 to R7, E15 to V18 and T28 to S1, and termination points on either side of these strands at W2, V4, N8, G14, G19/G20 and G26/C27. The  $\beta$ -strand found over C28 to S1 is interesting as a  $\beta$ -strand does exist in kalata-B1 in aqueous solution in this region but is not detected by  $^1\text{H}_\alpha$  CSI values. The other predicted  $\beta$ -strands are reasonably concordant with  $^1\text{H}_\alpha$  CSI results. However, the  $\beta$ -strand predicted to form in TFE by  $^1\text{H}_\alpha$  CSI results in the connecting region is only partially indicated. Less information is obtained from the CSI values based on  $^{13}\text{C}_\beta$  chemical shifts, which do not indicate much known structure in 20% TFE at 333K.

The new  $\beta$ -strand of the connecting region and the  $\beta$ -strand over residues E15 to C17 is predicted and exaggerated by  $^{13}\text{C}_\beta$  CSI values as a long stretch of  $\beta$ -strand extending over residues E15 to T24. The  $\beta$ -turn at V18

indicated by  $^1\text{H}_\alpha$  CSI values is not present, and nor are the  $\beta$ -strands over C5 to R7 and L10 to V12. Furthermore, the  $\beta$ -strand over C28 to S1 predicted by  $^{13}\text{C}_\alpha$  CSI values is also not predicted. Once again, the  $\beta$ -strand C5 to R7 would be indicated if the definition of  $\beta$ -strand based on  $^{13}\text{C}_\beta$  CSI values were altered to 3 or more " $+1$ " CSI values. It is possible that in small polypeptides like kalata-B1 that consist of short sections of  $\beta$ -strand and several turns, it is more difficult to predict  $\beta$ -strand character by  $^{13}\text{C}_\beta$  CSI values, which require longer stretches of positive indices to prove the presence of  $\beta$ -strand structure.

In summary, the  $^{13}\text{C}$  CSI values don't seem to be as consistent as  $^1\text{H}_\alpha$  CSI values. They don't clearly indicate dramatic and consistent changes between the aqueous and TFE environments over C27 to P3, nor do they consistently predict the formation of  $\beta$ -strand in the connecting region by both  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  CSI values. However, the  $^{13}\text{C}_\beta$  CSI values do predict the formation and development of a  $\beta$ -strand in the connecting region. Also, both  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  CSI values indicate some change in the disordered region, which is also suggested by NH chemical shift movement but not by  $^1\text{H}_\alpha$  CSI values.

While indication of secondary structure from  $^{13}\text{C}$  chemical shifts has been difficult to clearly ascertain,  $^{13}\text{C}$  chemical shifts can be extremely useful in distinguishing the isomerisation of X-Pro bonds.

#### 4.3.3 Determination of *Cis/Trans* Proline Isomerisation

In the NMR study of kalata-B1 (Saether *et al.*, 1995), spectral overlap presented some difficulty in determining the state of isomerisation of X-Pro bonds. Alteration of temperature and solvent environment may result in shifting of proton chemical shift values, which could then permit observation of NOE cross peaks that indicate proline isomerisation. In addition,  $^{13}\text{C}$  chemical shift values of prolines may also assist in diagnosing *cis/trans* proline isomerisation. Both NOE data of kalata-B1 in different environments and  $^{13}\text{C}$  chemical shifts were used to further elucidate *cis/trans* proline isomerisation.

In the case of NOE data, *cis/trans* proline conformation may be determined by the presence and absence of certain NOE cross peaks between the proline residue and its previous residue (Wüthrich, 1986). The NOESY spectrum of kalata-B1 in 6% TFE at 303 K has informative NOE cross peaks for P3 and P25. Along the W2 proton alpha connectivity pattern, a strong cross peak attributable to the P3 alpha is present. Although the degenerate P3 delta cross peak would superimpose with one of the W2 betas, the identification of the Pro-3 alpha cross peak strongly indicates that P3 adopts the more unusual *cis* conformation. Similarly, along the T24 alpha connectivity pattern, two separate cross peaks corresponding to the P25 deltas are observed. This strongly indicates that P25 adopts the more common *trans* configuration. Unfortunately, no *cis/trans* assignment for P11 could be made, as no proline *cis/trans* information could be obtained from either the L10 NH or alpha connectivity patterns for the reason that the P11 and L10 alpha protons are coincident. Figure 4.8 shows the cross peaks that indicate the *cis/trans* conformations of P3 and P11 in the NOESY spectrum of kalata-B1 in 6% TFE at 303K.

The NOESY spectrum of kalata-B1 in 20% TFE at 300 K only indicates the isomerisation state of P25. The presence of two separate cross peaks, corresponding to P25 delta protons, are located along the T24 alpha connectivity pattern. The conformation of P11 and P3 could not be determined due to overlap between the alpha connectivity patterns of these prolines and the preceding residues. Data from  $^{13}\text{C}$  chemical shifts is more discriminating in the ability to assign peaks.

It is well established that the transition from *trans* to *cis* proline results in an upfield shift for the  $\beta$ -carbon and a downfield shift for the  $\gamma$ -carbon (Kessler and Seip, 1994). Therefore, the shift difference between  $\beta$ - and  $\gamma$ -carbons increases dramatically in the *cis* compared to *trans* forms, the difference generally being less than 5 ppm for *trans* prolines and approaching 10 ppm for *cis* prolines. The differences were measured for each proline of kalata-B1 in 6% and in 20% TFE. The results are tabulated in Table 4.4 and

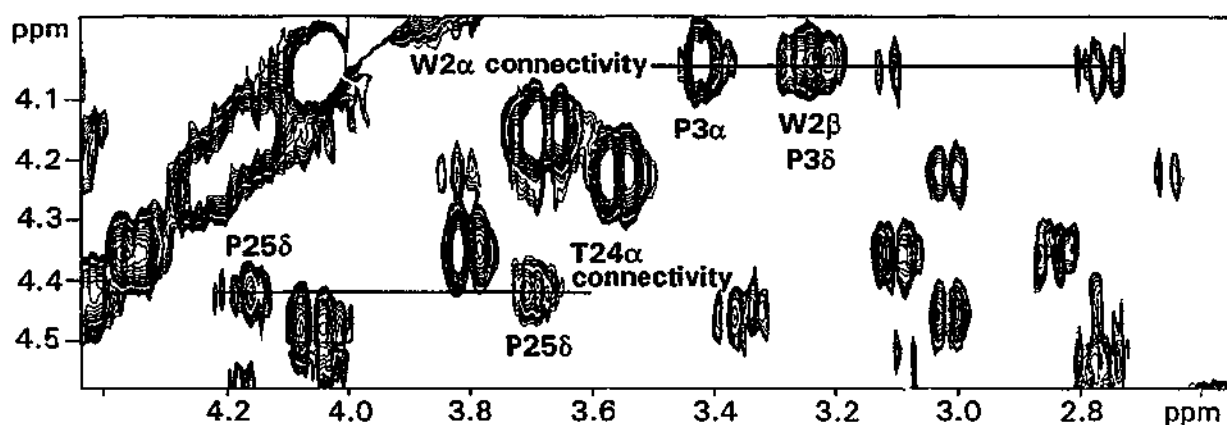


Figure 4.8. NOESY spectral data showing cross peaks associated with *cis/trans* conformation of P3 and P25 for kalata-B1 in 6% TFE at 303 K.

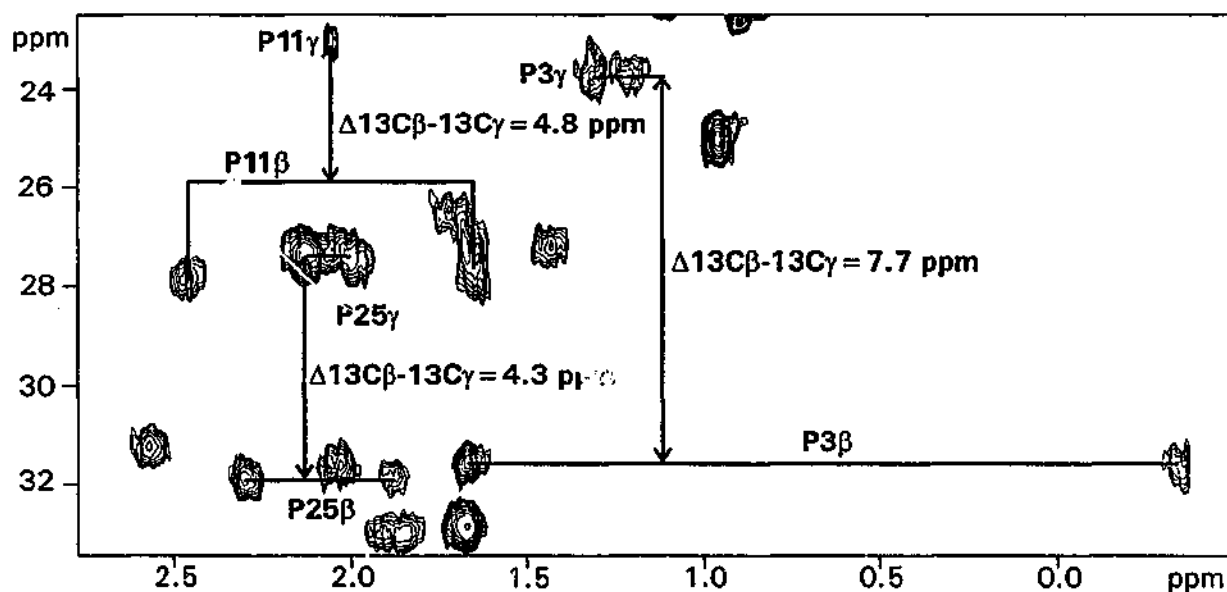


Figure 4.9. HMQC spectral data showing differences between  $\beta$ - and  $\gamma$ - carbons that indicate *cis/trans* conformation of prolines for kalata-B1 in 6% TFE at 303 K.



the HMQC spectrum showing the measurements are shown in Figure 4.9 for kalata-B1 in 6% TFE. The HMQC spectrum of the same region is almost identical for kalata-B1 in 20% TFE.

Proline	$C_{\beta} - C_{\gamma}$ (ppm) 6% TFE	$C_{\beta} - C_{\gamma}$ (ppm) 20% TFE
3	7.7	8.4
11	4.8	5.0
25	4.3	4.6

**Table 4.4.** Differences between  $C_{\beta}$  and  $C_{\gamma}$  carbons for proline residues of kalata-B1 in 6% and 20% TFE.

A difference of close to 10 ppm is obtained for P3 (7.7 ppm in 6% TFE; 8.4 ppm in 20 % TFE) and values of around 5 ppm are obtained for P11 (4.8 ppm in 6% TFE; 5.0 ppm in 20 % TFE) and P25 (4.3 ppm in 6% TFE; 4.6 ppm in 20 % TFE). These values are consistent with a *cis* conformation for P3 and a *trans* conformation for P11 and P25. The determination of X-Pro bonds for kalata-B1 from  $^{13}\text{C}$  chemical shifts are unequivocal, as assignments of  $\beta$ - and  $\gamma$ -carbons are not subject to coincidence with other cross peaks as occurs with the NOE cross peaks of NOESY spectra. For this reason, the use of  $^{13}\text{C}$  data has been highly successful for the conclusive establishment of *cis/trans* proline isomerism in kalata-B1. The known isomeric state of the proline bonds can be utilised in the final part of this study, which is the structure determination of kalata-B1 in 20% TFE at 300 K.

#### 4.3.4 Three Dimensional Structure of Kalata-B1 in the presence of TFE

##### 4.3.4.1 Quality of Structures

A set of 50 structures was generated using a total of 362 NOE distance restraints. A listing of these restraints is given in Appendix III. The 12 best structures in terms of total and NOE energies were examined in detail and used in comparison with the structure of kalata-B1 in the aqueous environment. Geometric and energetic statistics, which define these 12 structures, are given in Table 4.5. There appears to be no significant deviation from ideal covalent geometry. The structures fit the experimental restraints and there are no NOE violations greater than 0.3 Å. As expected the superimposition over the backbone atoms of the 12 structures in Figure 4.10 shows that all regions are mostly well defined.

The mean pairwise RMSDs over the whole kalata-B1 molecule are 0.96 Å for the backbone atoms and 1.62 Å for all non-hydrogen atoms. The overall structural statistics are of sufficiently good quality for the 12 best structures to be used in comparisons with the 3D structure of kalata-B1 in aqueous solution.

##### 4.3.4.2 Comparison of the Structures of Kalata-B1 in Different Solvents

Initially, the 3D structure of kalata-B1 in aqueous solution was recalculated using the original experimental restraints (Saether *et al.*, 1995) but including the *cis* proline configuration for P3 and applying the simulated annealing protocols identical to those utilised for calculating the structure of kalata-B1 in 20% TFE. To compare the 3D topologies of kalata-B1 in different solvent environments, the 12 best structures for both the 20% TFE and aqueous environments were used to produce single average structures for each solvent environment. These average structures were then compared as follows.

Table 4.5. Geometric and energetic statistics for the 12 final structures of kalata-B1 in 20% TFE.

<i>Mean RMS deviations from experimental restraints</i>	
Interproton distances (Å; 362) <sup>a</sup>	0.02 ± 0.001
Dihedral angles (deg.)	0.67 ± 0.44
<i>Mean RMS deviations from idealised geometry</i>	
Bonds (Å)	0.008 ± 0.0003
Angles (deg.)	2.32 ± 0.043
Impropers (deg.)	0.18 ± 0.03
<i>Mean Energies (kcal mol<sup>-1</sup>)</i>	
$E_{\text{NOE}}^b$	2.58 ± 0.36
$E_{\text{cdih}}^b$	0.04 ± 0.03
$E_{\text{LJ}}^c$	-106.9 ± 3.5
$E_{\text{bond}}$	4.02 ± 0.28
$E_{\text{improper}}$	0.55 ± 0.19
$E_{\text{angle}}$	53.5 ± 2.0
$E_{\text{total}}$	-41.2 ± 5.3

The values in the Table are the mean ± standard deviation.

<sup>a</sup> The number of restraints is shown in parentheses. None of the structures had distance violations >0.3 Å.

<sup>b</sup> Force constants for the calculation of square-well potentials of the NOE and dihedral angles were 50 kcal mol<sup>-1</sup> Å<sup>-1</sup> and 200 kcal mol<sup>-1</sup> rad<sup>-2</sup>, respectively.

<sup>c</sup> The Lennard-Jones van der Waals energy was calculated with the CHARMM empirical energy function.

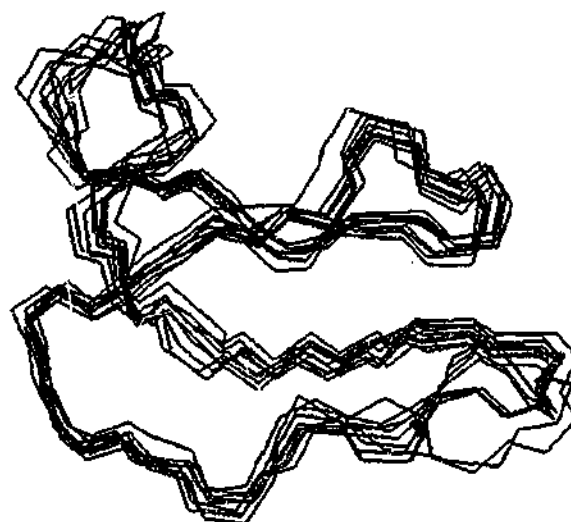


Figure 4.10. Superimposition of the 12 best NMR-derived structures over the backbone atoms (N, C $\alpha$ , C') of kalata-B1 in 20% TFE.

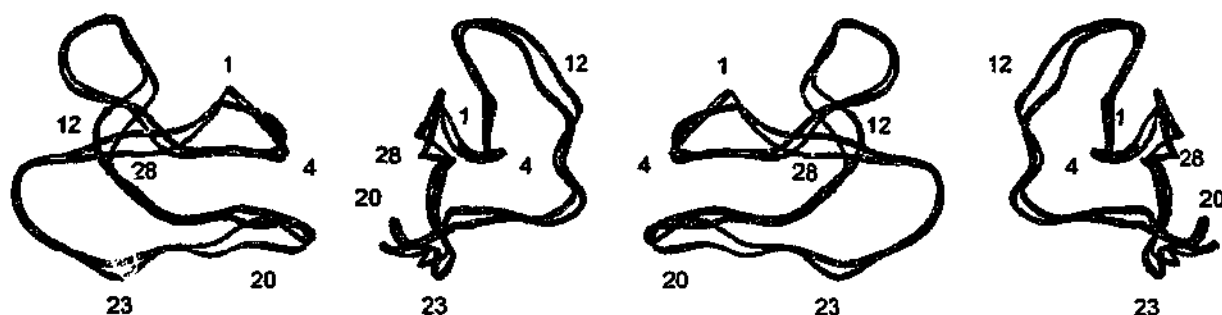
#### 4.3.4.2.1 Overall 3D Fold

In general, the overall fold has been largely maintained. The mean pairwise RMSDs over the whole of the average molecules is 0.99 Å for the backbone atoms and 1.71 Å for the non-hydrogen atoms. These values are close to those obtained for the 12 best structures for kalata-B1 in 20% TFE. In particular, the fold appears to be preserved much more closely over the knot motif, with the mean pairwise RMSDs for the regions incorporating residues 14-28 and 28-8 being 0.66 Å for backbone atoms and 1.06 Å for all non-hydrogen atoms.

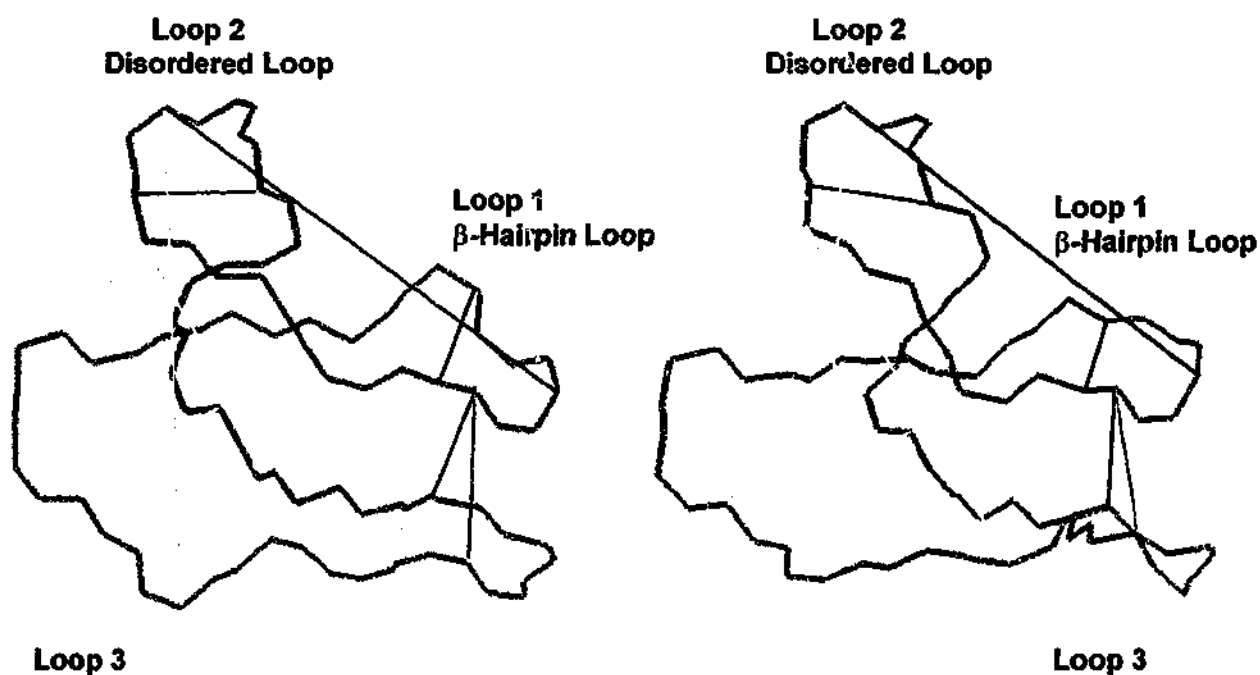
#### 4.3.4.2.2 Residual Deviations

A superimposition over the backbone atoms of the average structures is illustrated in Figure 4.11. A red ribbon represents the average structure of kalata-B1 in aqueous solution and a green ribbon represents that fold of kalata-B1 in 20% TFE. Four orientations involving 90° rotations in Y show that, while changes are minor, there are more noticeable deviations in local regions. The illustration shows deviations in the vicinity of residues 1, 4, 12, 20, 23, 28. The deviations are most visually notable at residues 12 and 20. To gauge the extent of deviation, the C $\alpha$  to C $\alpha$  distance between corresponding residues of the average structures was measured. The measurements are given in Table 4.6.

The greatest distances occur between corresponding C $\alpha$  atoms for residues 11, 12, 20 where the C $\alpha$  distance is > 2 Å apart. The next largest distance of ~1.5 Å occurs for residues 17 and 23 and distances of ~1.0 Å are found for residues 1, 2, 3, 4, 5, 8, 19, 21, 28, 29. In other words, the greatest deviations are found generally over residues 11 to 12, 17 to 23, 28 to 5. These regions have been implicated as regions of change due to the TFE solvent environment by previous results obtained throughout this chapter.



**Figure 4.11.** Four views of  $90^\circ$  rotations in Y of the superimposition of the backbone atoms (N, C $^\alpha$ , C') of the average structures of kalata-B1 in different solvent environments. The average structure of kalata-B1 in the 20% TFE is shown by the green ribbon. The average structure of kalata-B1 in aqueous solution is shown by the red ribbon. Residues where deviations between the two folds are especially noticeable are marked.



**Figure 4.12.** Display of backbone atoms (N, C $^\alpha$ , C') of the average structures of kalata-B1 in different solvent conditions showing distance changes. The green backbone represents the average structure in 20% TFE and the red backbone represents the average structure in aqueous solution. The three main loops of kalata-B1 are indicated. Lines for selected distances within and between the loops are shown colour coded according to the percentage change in the distance that arises in the 20% TFE environment. Dark blue represents  $> 30\%$  change, light blue  $> 20\%$  change, purple  $\sim 10\%$  change and yellow for  $< 5\%$  change.

**Table 4.6.** C $_{\alpha}$ -C $_{\alpha}$  distance between corresponding C $_{\alpha}$  atoms of the average structure in water and the average structure in TFE.

Residue No.	C $_{\alpha}$ -C $_{\alpha}$ (Å)	Residue No.	C $_{\alpha}$ -C $_{\alpha}$ (Å)	Residue No.	C $_{\alpha}$ -C $_{\alpha}$ (Å)
S1	1.10	P11	2.11	T21	0.90
W2	1.09	V12	2.10	C22	0.38
P3	1.11	C13	0.80	N23	1.46
V4	0.96	G14	0.72	C24	0.47
C5	0.94	E15	0.76	P25	0.50
T6	0.62	T16	0.65	G26	0.56
R7	0.72	C17	1.42	C27	0.49
N8	1.01	V18	0.70	T28	1.36
G9	0.84	G19	0.88	C29	1.22
L10	0.58	G20	2.11	-	

#### 4.3.4.2.3 Elements of Secondary Structure

To determine whether the residual changes have any effect on secondary structure, the average structures were examined using the program PROMOTIF (Hutchinson and Thornton, 1996). Several changes have indeed occurred. Foremost, neither of the anti-parallel sheet topology, the  $\beta$ -bulge (involving residues 5, 12, 13), the 2:4 hairpin (involving residues 3 to 17), the two  $\beta$ -strands (involving residues 4 to 6 and 11 to 16) that were classified for kalata-B1 in aqueous solution have been identified for kalata-B1 in 20% TFE. Moreover, a  $\beta$ -strand which was predicted to arise in 20% TFE by amide proton and  $^{13}\text{C}_{\beta}$  proton chemical shifts has not been classified. Although changes which could from some viewpoints be described as a straightening of the polypeptide chain this region (Figure 4.11) in 20% TFE, the formation of  $\beta$ -strand is not evident.

However, the  $\beta$ -bulge region in aqueous solution is described as an inverse gamma turn in 20% TFE involving residues 10 to 12. This change is

not surprising, as change in this region is pre-empted by amide NH,  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shift data described earlier. The 3D calculations fulfil the predictions and indicate the nature of the change. The  $\text{C}_\alpha\text{-C}_\alpha$  difference between corresponding  $\text{C}_\alpha$  atoms of the average structures in this region are a considerable  $\sim 2 \text{ \AA}$  (P11  $2.11 \text{ \AA}$ ; V12  $2.10 \text{ \AA}$ ) and a close view of the superimposed average structures (Figure 4.11) in this region shows a sharper twist of the backbone in 20% TFE.

In addition, two non-standard type IV  $\beta$ -turns involving residues 2 to 5 and 11 to 14 have been classified for kalata-B1 in 20% TFE. The changes are not unexpected, as amide proton chemical shift data, also suggested changes in this region. The calculated structures have substantial  $\text{C}_\alpha\text{-C}_\alpha$  difference between corresponding  $\text{C}_\alpha$  atoms average structures of  $\sim 1\text{-}2 \text{ \AA}$  for these residues. Once again, a close view of the superimposed average structures (Figure 4.11) in these regions shows the presence of sharper turns in 20% TFE.

Furthermore, the disulphide conformations differ between the average structures as shown in Table 4.7. It is also interesting to draw attention to the  $\text{C}_\alpha$  distance of the disulphides. This distance is the same in both solvent environments for the 5-22 disulphide but has decreased by  $0.4$  and  $0.7 \text{ \AA}$  for the 13-27 and 17-29 disulphides, respectively, in the 20% TFE environment, suggesting a possible change in the geometry in the TFE environment. However, comparing bond or individual distances must be viewed with caution, as compensating changes in other distances may also occur.

Explanations for the definitive loss of sheet topology, hairpin and strand elements are possible in terms of relative spatial distance changes, which are large enough to disable the application of formal classification. Whilst the overall fold is very similar in the TFE and aqueous environments, the combination of deviations from standard classifications indicate that the 3D structure of kalata-B1 has been significantly distorted in the TFE environment. Some of the spatial distortions are highlighted in the next section.

**Table 4.7.** Disulphide structure information for average structures.

Disulphide Connection	Structure in Water		Structure in TFE	
	Type	C <sub>α</sub> Distance (Å)	Type	C <sub>α</sub> Distance (Å)
5-22	Right Hand Hook	6.0	Left Hand Spiral	6.0
13-27	Left Hand Spiral	6.2	Right Hand Hook	5.8
17-29	Right Hand Spiral	6.4	-	5.7

#### 4.3.4.2.4 Distortion in Space

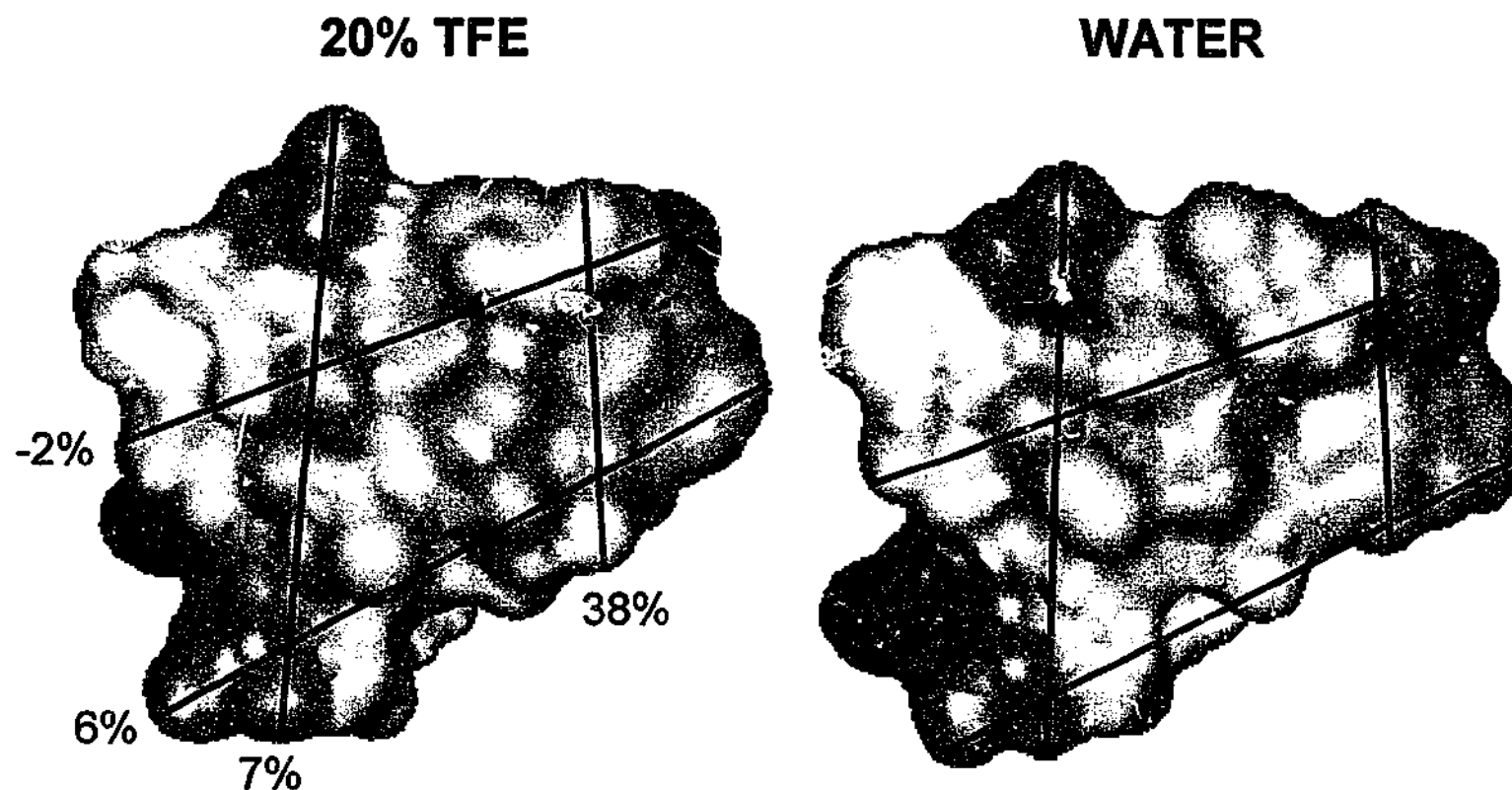
In consequence of having observed changes at particular residues and having obtained a loss of secondary structure elements, it was thought highly probable that distortions between key features in 3D space have occurred in the structure of kalata-B1 in the TFE environment. To examine this in detail, various distances between and within features of the kalata-B1 molecule were measured for each average structure and compared. Figure 4.12 displays the backbone atoms (N, C<sup>α</sup>, C') of the average structures labelled with the three main loops of kalata-B1 (Loop 1 - β-hairpin loop; Loop 2 - disordered loop; Loop 3 - third loop) and lines for distances measured between and within the loops. The distances are coloured according to the percentage change in distance that arises between the aqueous to the TFE environment, specifically, dark blue for > 30% change, light blue for > 20% change, purple for ~ 10% change and yellow for < 5% change. Distance changes of less than 5% were considered inconsequential. In this category, internal distance changes within Loop 3 and distance changes between Loop 3 and the disordered loop are apparently negligible. However, changes around 10% and greater could be of significance.



Increases in distance in the TFE environment of around 10% seems to occur within the disordered loop and between the extremities of the  $\beta$ -hairpin loop and the disordered loop. However, a decrease in distance of 13% in TFE is apparent between the 5 to 7  $\beta$ -strand of the  $\beta$ -hairpin loop and the 15 to 17  $\beta$ -strand of Loop 3. The latter change is accompanied by an increase in distance of 22% between the connecting region of Loop 3 with respect to the  $\beta$ -hairpin loop, indicating overall a tilting of Loop 3 in relation to the  $\beta$ -hairpin loop. Finally, a more pronounced distance increase of 33% arises within the  $\beta$ -hairpin loop across the connecting turn. Whilst the changes in distances are subtle and occur in some regions only, it can be envisaged there may be implications for changes with respect to surface exposure of kalata-B1 in the TFE environment.

#### 4.3.4.2.5 Hydrophobic Surface Area

One of the most striking and unusual features of the fold of kalata-B1 is the presence of a solvent exposed hydrophobic surface. There is considerable difference in the solvent hydrophobicity between the aqueous environment, which is a highly polar medium, and the TFE environment, which is far more hydrophobic. It was of interest to ascertain whether the presence of 20% TFE has any affect on the hydrophobic surface of kalata-B1. A Connolly algorithm supplied by the INSIGHT program suite (Biosym Technologies, San Diego, CA) was applied to both average structures to generate solvent accessible surfaces. The identical hydrophobic surface orientations for each average structure are shown in Figure 4.13. The surface is coloured according to the degree of hydrophobicity. Increased intensity of shading towards red is an indication of increased hydrophobicity. Cross-sectional measurements to encompass the area occupied by the surface were determined. The measurements are shown on the diagram labelled with the percentage change in length that occurs in 20% TFE relative to the original length in water. A positive value indicates an increase in length in the TFE environment and a



**Figure 4.13.** Hydrophobic solvent accessible surfaces for the kalata-B1 in different solvent conditions. The surfaces are coloured according to the degree of hydrophobicity. Increased intensity of shading towards red is an indication of increased hydrophobicity. Cross-sectional measurements to encompass the area occupied by each surface are shown labeled with the percentage change in the cross-section that arises in 20% TFE relative to the aqueous solution. A positive value indicates an increase in the cross-section size that occurs in 20% TFE.

negative value indicates a decrease in length in the TFE environment. The diagram illustrates that there are some differences of the hydrophobic surface in the different solvent environments.

In general, the overall red/pink intensity of the hydrophobic surface in the TFE environment is more intense than in aqueous solution, suggesting that perhaps structural changes of the fold have taken place to increase the exposure of hydrophobic residues on the surface. With respect to the orientation of the surfaces in the diagram, the cross-sectional measurements increase lengthwise by 7% and 38% in TFE. The width, however, increases by 6% in one measurement and decreases by what could be considered a negligible 2% in the other measurement.

Overall, this represents an increase in the hydrophobic surface area kalata-B1 exposed to solvent in 20% TFE which seems to vary in diameter across the surface from being negligible to ~38%. It would seem that in the TFE environment, the hydrophobic interactions between kalata-B1 and the solvent environment act to result in an increase in the hydrophobic surface area resulting in a loss of classical secondary structure definition. Nonetheless, the intramolecular forces and the 3D fold secured by the knot motif and cyclic nature are largely maintained, even in a disruptive solvent environment such as 20% TFE. In conclusion, while the results of this chapter suggest some conformational changes, they do further support the robust nature of kalata-B1, which as such could potentially serve as a design for a highly stable template in future drug design.

#### 4.4 Summary

The structure of kalata-B1 was explored under different temperature and solvent conditions by 2D homonuclear and heteronuclear NMR spectroscopy in combination with dynamical simulated annealing calculations. The main aims of the study were to ascertain the ability of  $^{13}\text{C}$  NMR data to predict secondary structural elements and to investigate the effect on structure of the addition of

the hydrophobic solvent, TFE, to the solution environment. This was done to test the hypothesis that kalata-B1 represents a stable molecular framework. If this is the case, it should be relatively impervious to the effects of external factors such as solvent.

The structure was examined in aqueous solutions containing 6% and 20% TFE and the 3D fold was determined from NMR distance data obtained on the 20% TFE solution. It is expected that the findings of this study are likely to have implications for cyclic peptides of the cystine knot family in general. Many of the 2D NMR parameters did predict the location of changes due to the solvent environment, while the structures generated from the dynamical simulated annealing calculations were able to describe the nature of the structural changes.

Alpha proton chemical shifts were used to calculate CSI indices. These indices predicted the formation of  $\beta$ -strand in the connection region of kalata-B1 and changes over residues T28 to P3 in the presence of TFE. Amide proton chemical shifts also indicated changes in these regions as well as changes in the disordered and  $\beta$ -bulge regions due to TFE. In accordance with the 2D NMR data, the 3D structures produced by dynamical simulated annealing calculations deviated more prominently from the original fold in aqueous solution in precisely these regions, although the  $\beta$ -strand structure is not apparent in the connecting region.

Mixed success was achieved with using  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shift data to ascertain the presence and location of secondary structure and to predict changes in this structure due to TFE through CSI indices. The  $^{13}\text{C}$  method is not as reliable as proton data and does not clearly indicate changes due to TFE. As such,  $^{13}\text{C}$  chemical shift data were not particularly useful for structure determination. However, full sets of  $^{13}\text{C}$  chemical shifts for kalata-B1 in 6% and 20% TFE were assigned and verified using 2D heteronuclear NMR spectroscopy. Additionally, an unexpected finding of a *cis* proline conformation for P3 was obtained from NOE data in 6% TFE which was unequivocally established from measurements on the magnitude of the

chemical shift difference between  $^{13}\text{C}_\beta$  and  $^{13}\text{C}_\gamma$ . The  $^{13}\text{C}$  data were highly reliable for the determination of *cis/trans* proline conformation of all the kalata-B1 prolines.

Analysis of the 3D structures for kalata-B1 in 20% TFE showed a loss of much of the secondary structure, including the anti-parallel sheet topology, the  $\beta$ -bulge (involving residues 5, 12, 13), the 2:4 hairpin (involving residues 3 to 17) and the two  $\beta$ -strands (involving residues 4 to 6 and 11 to 16). The  $\beta$ -bulge is replaced by a gamma inverse turn over residues 10 to 12. The disulphide core is slightly tighter. In terms of spatial distortion of the three main loops of kalata-B1, there is a tilting of the third loop with respect to the  $\beta$ -hairpin loop, an increase in distance between the extremities of the disordered loop and the  $\beta$ -hairpin and negligible distance changes between the disordered and third loops. There is also an increase within the diameters of disordered loop and the  $\beta$ -hairpin loop across the connecting turn, and negligible change within the third loop. The net effect of all the changes seems to be an increase in the hydrophobic surface area.

Despite these observations, all the data indicated that changes due to the presence of TFE are subtle and that the 3D fold is largely maintained in the environment of TFE. The hydrophobic solvent, TFE, which is known to induce formation of secondary structure, particularly  $\alpha$ -helices, seems to have a minimal effect on the structure of kalata-B1. This result further enforces the resilient and stable nature of the kalata-B1 structure, which may potentially be useful as a starting point for the future design of target specific drugs.

**CHAPTER FIVE**

**STUDIES OF ACYCLIC**

**PERMUTANT 6**

## 5.1 Introduction

This chapter presents the results of the study of acyclic permutant 6 proposed in Chapter 3. Acyclic permutant 6 is a key fragment, as it is the topological equivalent of non-cyclic members of the cystine knot family. Moreover, this particular fragment of kalata-B1 was of prime interest, with regard to activity, as the section missing from this acyclic permutant is considered highly probable for being important to receptor binding (Saether *et al.*, 1995). The missing section is the disordered loop, shown in perspective in Figure 5.1, and the acyclic permutant derived from kalata-B1, together with the sequence are also displayed in Figure 5.1. The numbering system used throughout this chapter refers to the order of the peptide synthesis rather than the equivalent residue number in kalata-B1 employed to date.

An additional kalata-B1 residue is included at the ends of the designed sequence, so that the sequence does not begin and finish with the all important cysteine residues. The acyclic permutant under study excludes Arg7 to Pro11 of the kalata-B1 sequence and includes Pro3 of the kalata-B1 sequence, which was found to be in the more unusual *c/s* proline conformation in the previous chapter. The oxidation, purification and 2D-NMR analysis of the acyclic permutant 6 will be described in detail in this chapter, commencing with materials and methods employed.

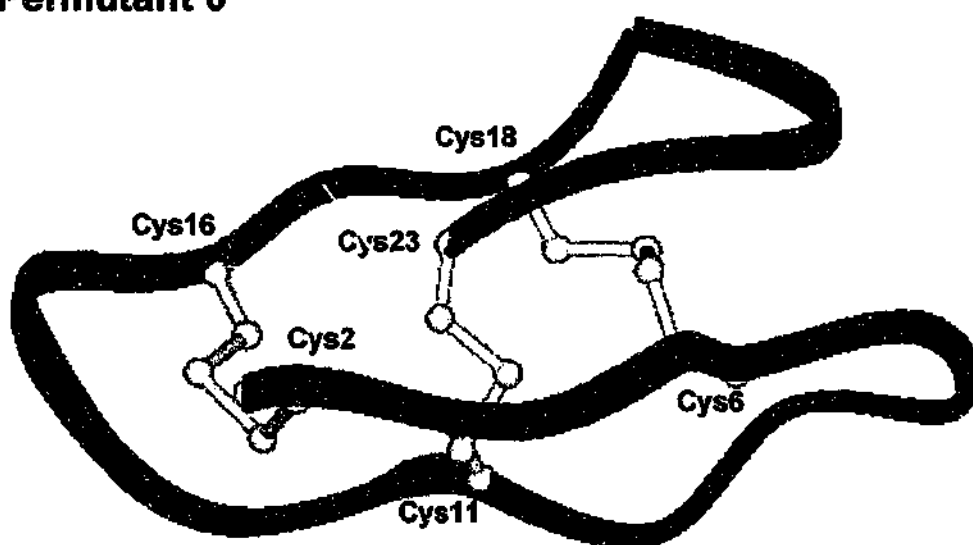
## 5.2 Materials and Methods

An overview of methods used in this thesis is given in Chapter 2. In the following section, specific details relevant to the study of acyclic permutant 6 are recorded.



**Disordered  
Loop**

**Acyclic  
Permutant 6**



**V C G E T C V G G T C N T P G C T C S W P V C T**  
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

**Figure 5.1.** Disordered loop, and representation and sequence of acyclic permutant 6 as derived from kalata-B1.



### 5.2.1 Synthesis and Oxidation

The linear sequence of acyclic permutant 6 was kindly synthesised by Dr. Dianne Alewood (Centre for Drug Design and Development, University of Queensland, Queensland, Australia). All subsequent experimental work was carried out as part of the current thesis project. The crude material was oxidised to form the disulphide linkages in a 0.1M  $\text{NH}_4\text{HCO}_3$  solution, at pH 8, with constant stirring overnight, at room temperature. These particular oxidation conditions are commonly used. Purification was then required to isolate the main component of oxidation, which was expected to be the acyclic permutant of interest with the most energetically favorable disulphide connections, presumably that of the native fold of kalata-B1.

### 5.2.2 Purification

The oxidised peptide mixture was purified using reverse phase high performance liquid chromatography (RP-HPLC). A Hewlett Packard Series 1050 HPLC system was used for purification. Several columns were tested with the best configuration for resolution being an Alltima C18 (5  $\mu\text{m}$  pore, size 300 Å particle size) 250 x 4.6 mm analytical column in series with an ICI Spherisorb C8 (5  $\mu\text{m}$  pore size, 300 Å particle size) 250 x 4.6 mm analytical column.

An acetonitrile (ACN) / water ( $\text{H}_2\text{O}$ ) / trifluoroacetic acid (TFA) mobile phase system and linear gradient elution was utilised (Eluent A: 90%  $\text{H}_2\text{O}$  + 0.1% TFA + ACN, Eluent B: 90% ACN + 0.1% TFA +  $\text{H}_2\text{O}$ ). Components were separated and collected at a flow rate of 1.0 ml/min, using slow gradient elution over the elution region of interest. UV detection at 214 nm was the wavelength used for detection of components.

Analytical columns are advantageous over semi-preparative columns due to improved resolution of components. However, a drawback of analytical columns for purification is the small column loading. The consequence of this

is that many consecutive and individual separations are required to isolate and concentrate the major component compared to a semi-preparative column. In the case of the oxidised peptide mixture, purification separations were carried out on  $\sim 50 \mu\text{l}$  injection volumes of  $\sim 10,000$  ppm solutions of oxidised peptide mixture. This amounts to a column loading of  $\sim 0.5$  mg of oxidised peptide mixture per separation for purification. The purified fractions were collected manually over defined time spans, concentrated and finally freeze dried using a DYNAVAC FD405 freeze drying unit to recover the pure, solid material.

### 5.2.3 Purity

In addition to purification separations, analytical HPLC was conducted to monitor and ascertain changes in composition from crude to oxidised peptide, to determine main fractions and elution regions and to establish purity of purified solids. Multivariate UV detection was applied for this chromatography which was performed on  $\sim 1000$  ppm solutions and  $10\text{--}25 \mu\text{l}$  injection volumes. Often only one column, namely the Alltima C18 column, was used and  $0.1\%$  TFA of the mobile phase was replaced by  $0.1\%$  phosphoric acid.

To further establish purity, a triethylammonium phosphate (TEAP) system was applied as the mobile phase on purified fractions (Eluent A:  $1\% \text{H}_3\text{PO}_4 + 1.2\%$  triethylamine (TEA) +  $\text{H}_2\text{O}$ , Eluent B:  $60\% \text{ACN} + 40\%$  Eluent A). Linear gradient elution with a flow rate of  $1.0 \text{ ml/min}$  was used. Given that the TEAP eluent system is significantly different from the TFA eluent system, it could be expected that components not separated by one mobile phase might be separated by the other. This served as a second check on purity, in addition to mass spectrometry.

Fractions that were collected were tested for purity by mass spectrometry kindly obtained by Dr. Dianne Alewood. Once the acyclic

permutant was obtained as a pure, solid material, NMR experiments could be carried out.

#### 5.2.4 NMR Spectroscopy

##### 5.2.4.1 Sample Preparation

The purified material was prepared in a 10%:90% by volume  $^2\text{H}_2\text{O}$ : $\text{H}_2\text{O}$  solution and placed in a 5 mm Wilmad NMR tube for NMR experiments. This solution had concentration of approximately 2.6 mM and pH 4.59, and was used throughout for all NMR data acquisition.

##### 5.2.4.2 NMR Experiments

The NMR spectra were recorded on a Bruker AVANCE DRX600 NMR spectrometer equipped with a shielded gradient unit. Two dimensional NMR spectra were acquired in phase sensitive mode, using time proportional phase incrementation for quadrature detection in the  $t_1$  dimension (Marion and Wüthrich, 1983). Spectra were recorded at 282 K and 275 K. Lower temperatures were maintained by a temperature controlled stream of cooled air, which was supplied by a Bruker BCU05 refrigeration unit and controlled by a B-VT 2000 control unit. Two dimensional homonuclear experiments were acquired over 7003 Hz. These will be now described in greater detail.

TOSCY spectra (Braunschweiler and Ernst, 1983) were recorded using an MLEV-17 spin lock sequence (Bax and Davis, 1985) and mixing times of 76 and 126 ms. NOESY spectra (Jeener *et al.*, 1979) were recorded with a mixing time of 300 ms. Solvent suppression for NOESY and TOSCY spectra was attained using a modified WATERGATE sequence (Piotto *et al.*, 1992), where two gradient pulses of 2 ms duration and 6  $\text{Gcm}^{-1}$  strength were applied on either side of the binomial 3-9-19 pulse. TOCSY spectra were acquired with 4096 complex data points in  $F_2$  and 521 increments in the  $F_1$

dimension, with 64 scans per increment. NOESY spectra were acquired with 4096 complex data points in  $F_2$  and 400 increments in the  $F_1$  dimension, with 150 to 160 scans per increment.

DQF-COSY (Rance *et al.*, 1983) and E-COSY (Griesinger *et al.*, 1987) spectra were acquired with 4096 complex data points in  $F_2$  and 600 increments in the  $F_1$  dimension for the DQF-COSY experiments and 640 increments in the  $F_1$  dimension for the E-COSY experiments. The number scans per increment was 64 for the DQF-COSY spectra and 84 for the E-COSY spectra. Solvent suppression was attained using selective low-power irradiation of the water resonance during a relaxation delay of 1.8 s.

The only 2D heteronuclear experiment performed was an HMQC (Bax and Subramanian, 1986). This particular spectrum was recorded with detection of  $^1\text{H}$  in  $F_2$  using a GARP composite pulse for decoupling of the  $^{13}\text{C}$  nuclei (Shaka *et al.*, 1985). Solvent suppression was attained using selective low-power irradiation of the water resonance during a relaxation delay of 1.8 s. The data were acquired with 4096 complex data points in  $F_2$  and 512 increments in the  $F_1$  dimension, with 352 scans per increment. The spectrum was collected over 6024 Hz in  $F_2$  and 10061 Hz in  $F_1$ .

All spectral data were processed on a Silicon Graphics Indy workstation using XWINMR (Bruker) software. The  $t_1$  dimension of 2D data were zero-filled to 2048 real data points and  $90^\circ$  phase-shifted sine-bell window functions were applied prior to Fourier transformation. Polynomial baseline correction was used in regions either side of the water resonance to improve the spectral appearance. For measurement of  $^3J_{\text{NH-H}\alpha}$  coupling constants, the DQF-COSY data were strip transformed to 16 K in both  $F_1$  and  $F_2$  dimensions with no baseline correction.

For proton data, chemical shifts were referenced to the  $\text{H}_2\text{O}$  resonance that had been externally calibrated over the temperature range of interest with respect to 4,4-dimethyl-4-silapentane-1-sulfonate (DDS) set to 0.00 ppm. Calibration with reference to the aqueous solvent was used as the peptide degraded before calibration relative to a known reference could be measured.

In particular, this affected the HMQC data where the  $^{13}\text{C}$  chemical shifts could not be calibrated at all.

The processed NMR data provided results for a qualitative structural analysis of separated isolated peptide. These results are presented in the following section.

### **5.3 Results and Discussion**

#### **5.3.1 Separation and Purity**

##### **5.3.1.1 Oxidation**

The HPLC traces of the crude and oxidised peptide mixtures using a single C18 column are shown in Figure 5.2. Peaks and retention times differ between the crude and oxidised peptide, although the major peaks elute at similar retention times. For instance, two significant peaks at retention time 21.35 and 21.66 (main peak) minutes were obtained for the crude peptide. However, for the oxidised peptide, four peaks at retention times 20.62 (main peak), 21.14, 21.06 and 22.51 minutes were obtained. These differences indicate that changes have occurred upon oxidation to produce more than one species that requires separation.

##### **5.3.1.2 Separation**

Resolution for separation improved significantly when a C18 column was placed in series with a C8 column. The improvement can be seen in Figure 5.3, in which expansions over the elution region of major peaks for single and two columns are displayed. When two columns are used, five larger peaks at 24.97 (main), 26.52, 27.10, 27.78, 28.92 minutes, and five minor peaks are evident as opposed to three larger and four minor peaks when a single C18 column is used. Three regions can be defined for separation. These

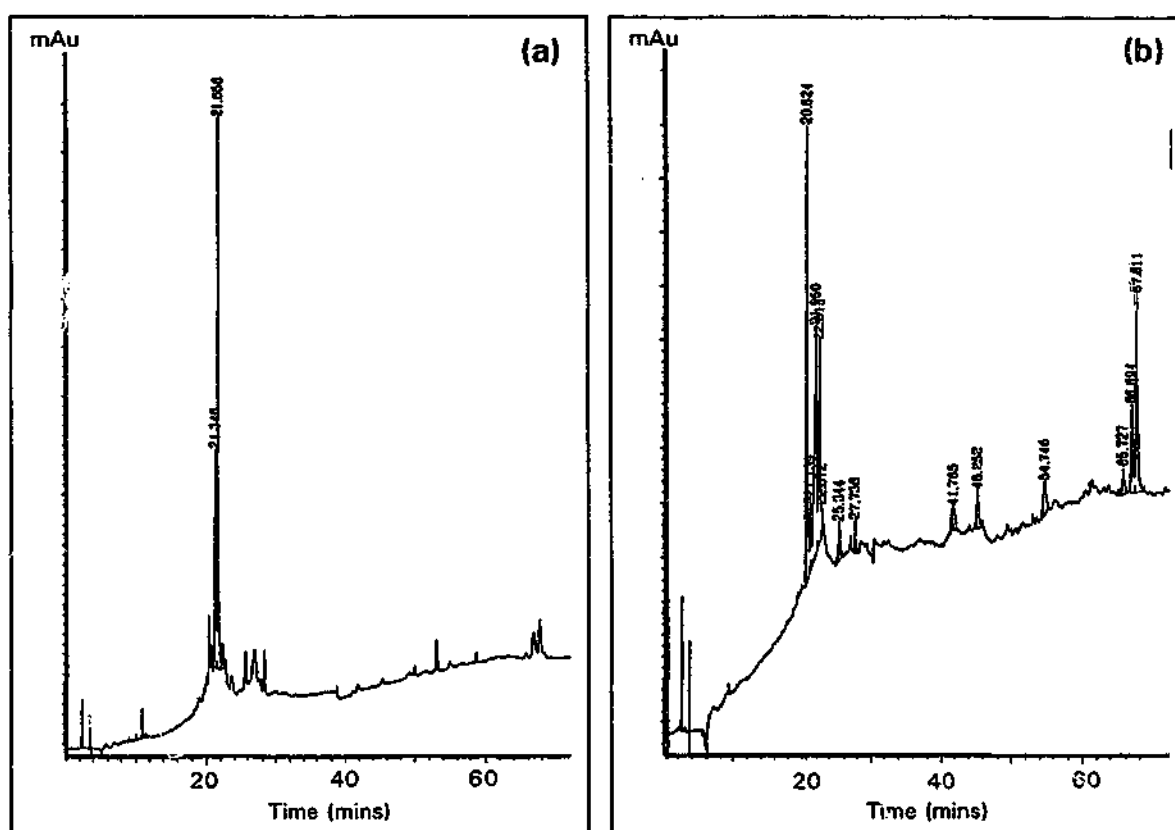


Figure 5.2. HPLC traces using a single C18 reverse phase column of (a) crude peptide (b) oxidised peptide.

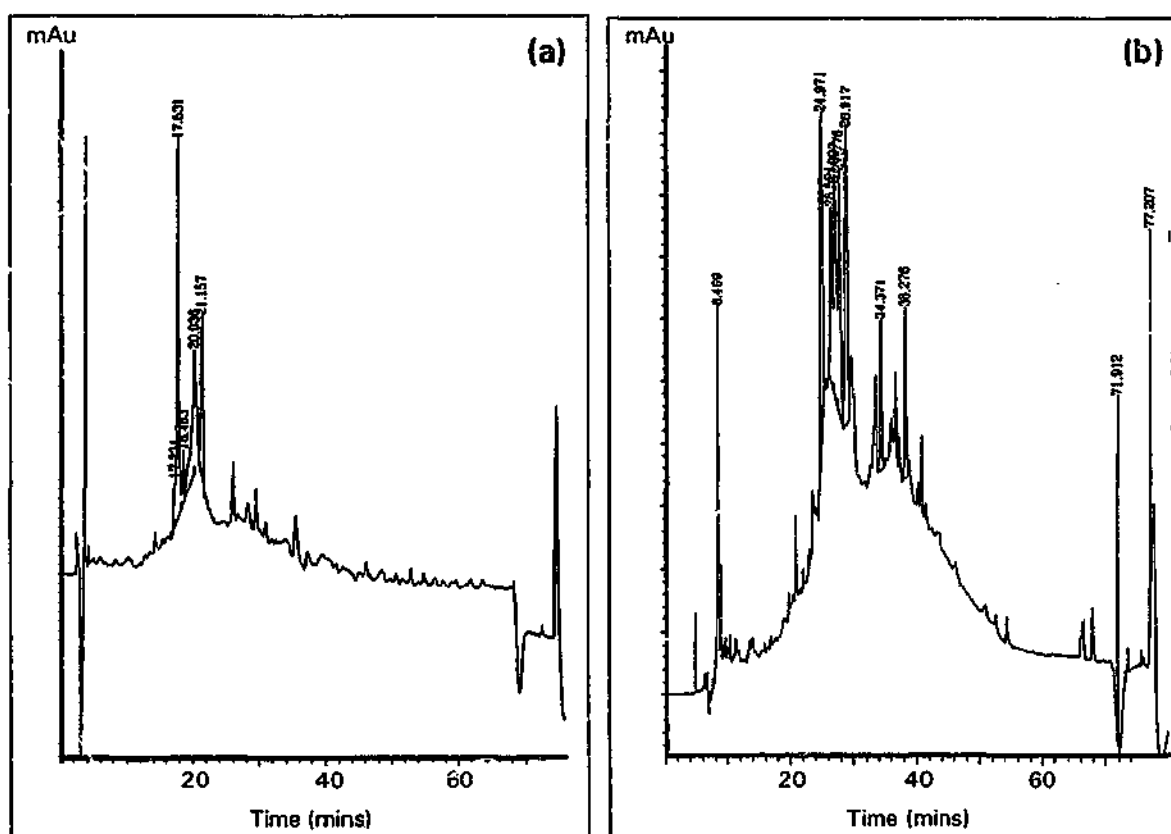


Figure 5.3. HPLC traces of oxidised peptide using (a) a single C18 reverse phase column (b) a C18 phase column in series with a C8 reverse phase column.

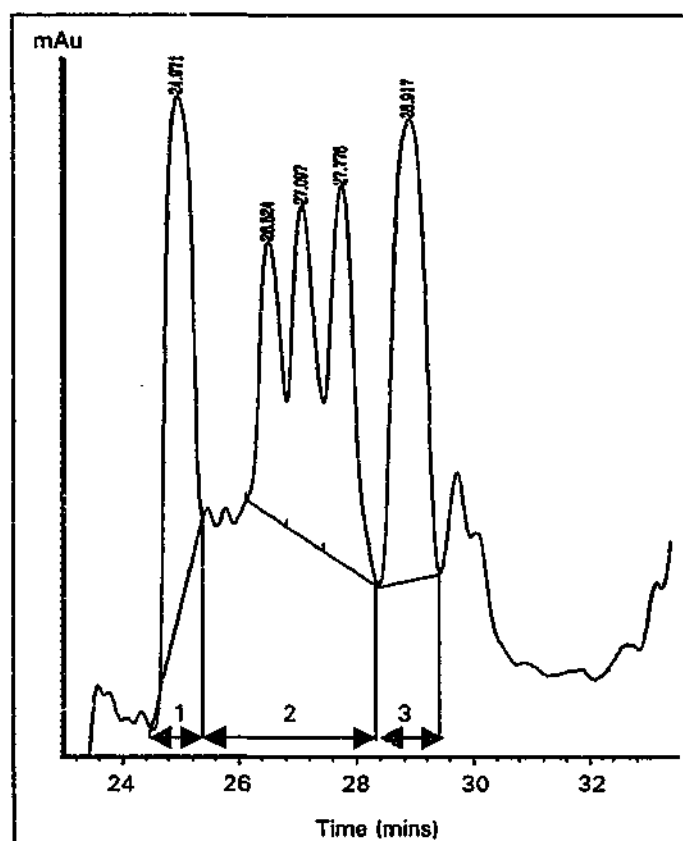
are shown in Figure 5.4 and material was collected over time spans corresponding to the trough points observed in the HPLC trace.

#### 5.3.1.3 Purity

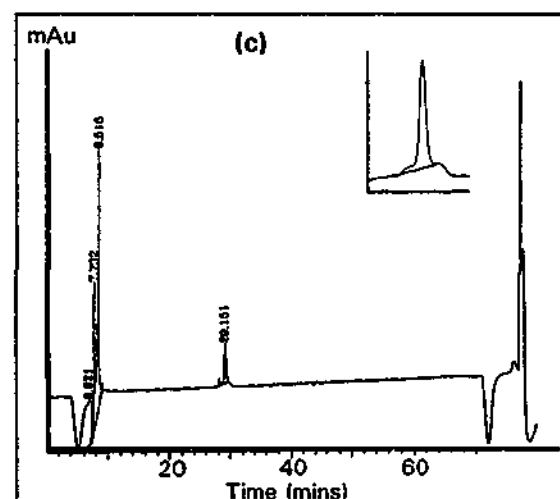
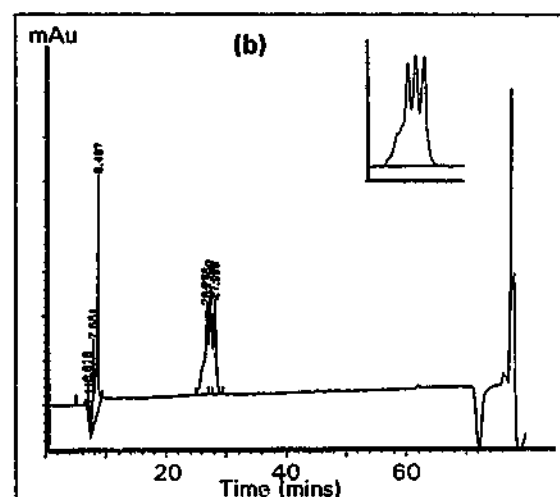
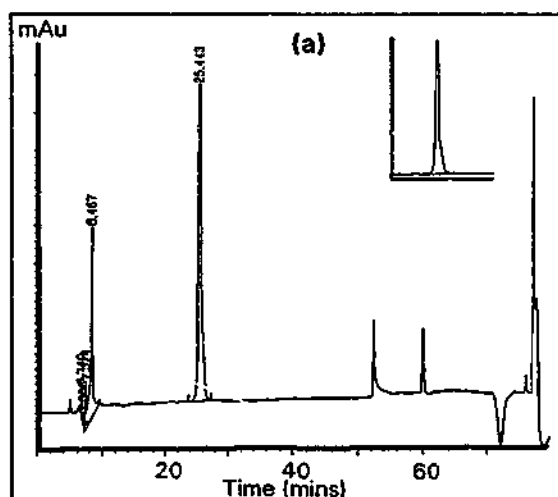
All three fractions were tested for purity by mass spectrometry and HPLC using the HPLC configuration of two columns in series. The resulting HPLC traces are given in Figure 5.5. A single peak was obtained for the first fraction, three peaks and two shoulders for the second fraction and a large peak with a couple of minor peaks for the third fraction. Data from the mass spectrometer indicated that all the three fractions are comprised of species consistent with the fully oxidised, synthesised sequence shown in Figure 5.1.

The first fraction appears to be the purest, as well as containing the highest amount of peptide, as indicated by the integration area, which is much larger than that of the other two fractions in the HPLC trace of Figure 5.4. For these reasons, the first fraction became the focus of subsequent NMR work. However, as it is well known that in certain instances, a single peak may be composed of more than one component, a second HPLC elution system was applied, where TEAP replaced TFA as the organic component of the mobile phase.

A TEAP elution system is driven by different chemical interactions from that of TFA (Chapter 2). It was speculated that this difference might be sufficient to separate multiple components should they be present in the first fraction. The result was that a single major peak was obtained with several additional minor peaks. As each of the minor peaks is present in the TEAP mobile phase, the minor peaks can be attributed to contamination, probably degradation products, of the TEAP mobile phase. The HPLC traces of the first fraction in the TEAP mobile phase and the TEAP mobile phase itself are shown in Figure 5.6. This outcome supports the previous finding that the first fraction is reasonably pure and satisfactory for NMR studies.



**Figure 5.4.** HPLC trace showing the three time regions over which material was collected.



**Figure 5.5.** HPLC traces of purified fractions (a) fraction from region 1 (b) fraction from region 2 (c) fraction from region 3. The HPLC trace expanded over the main peak is displayed in the right hand corner of the entire HPLC trace for each fraction.



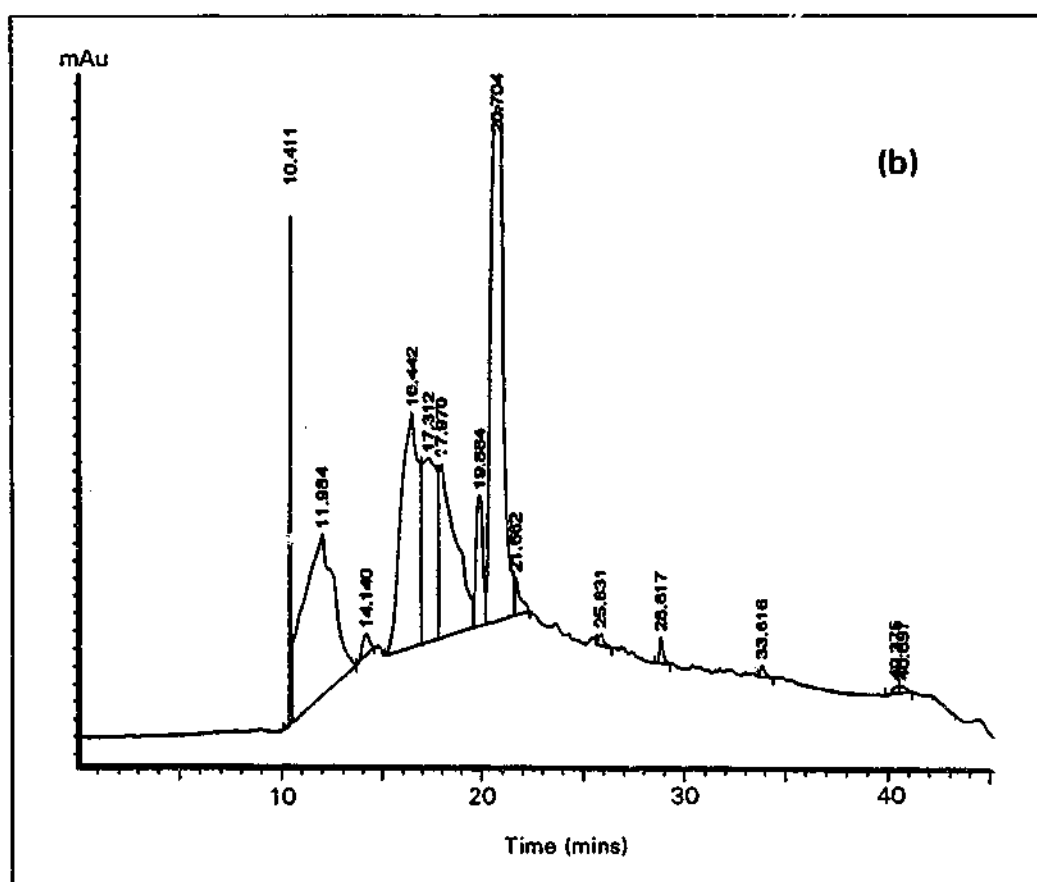
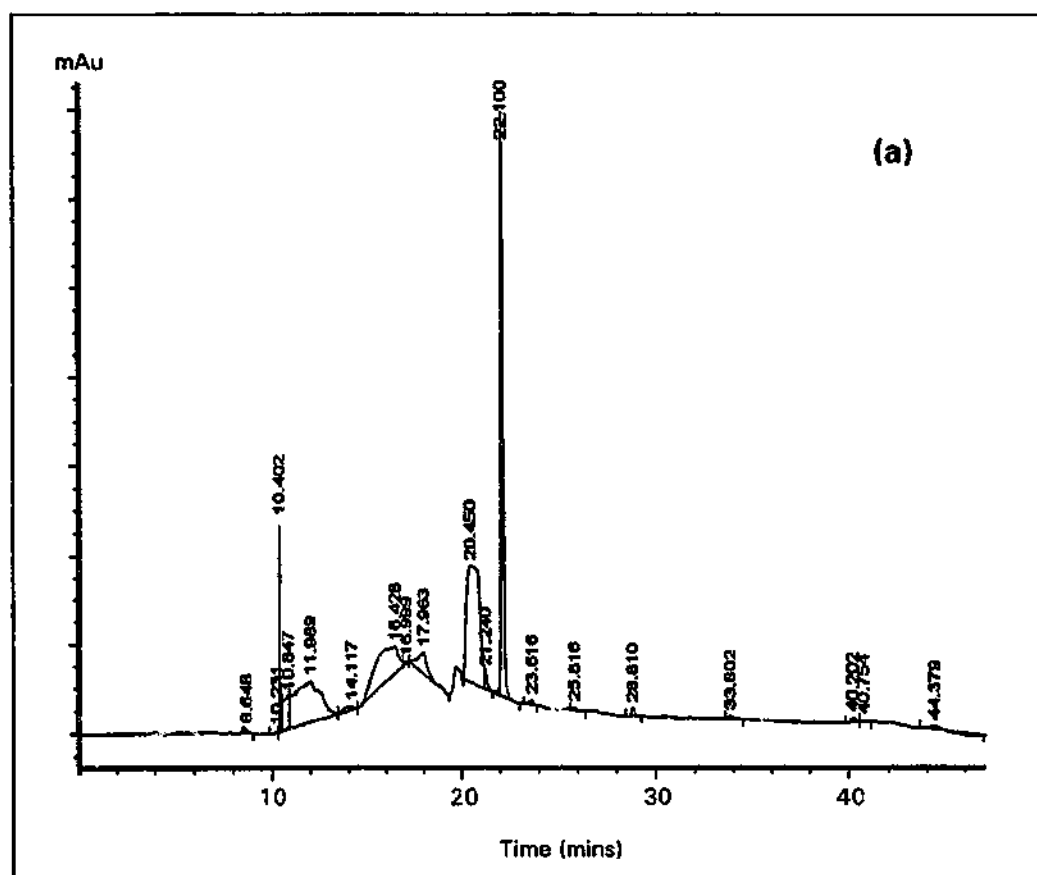


Figure 5.6. HPLC traces of (a) first fraction in TEAP mobile phase  
(b) TEAP mobile phase without first fraction.

### 5.3.2 Assignment of Proton Chemical Shifts and Detection of Isomers

The signal spread in the 1D spectrum as shown in Figure 5.7 is reasonable, although it can be described as crowded. The number of signals is in excess of that expected for a single species in solution according to the sequence. This implies the presence of more than one species. Intraresidue connectivities were neatly resolved in the 2D data, so that proton assignment of each signal was possible.

The intraresidue connectivities were determined from the TOCSY and COSY spectra. Characteristic spin systems (Wüthrich, 1986) were identified for each residue type. However, due to the complexity of the data, sequential information from NOESY spectra was required to assign the connectivities to separate species. Two distinct species are evident. Assignments of the TOCSY fingerprint regions are displayed in Figures 5.8. COSY data supported all amide, alpha and beta proton assignments.

The sequential assignments of the NOESY spectra for each individual species are shown in Figure 5.9. The sequential assignments are in line with the expected sequence, with the complete series of  $\alpha\text{H}$ -NH sequential connectivities unbroken except at the prolines. Nonetheless, the connectivities are completed as NOE cross peaks are present between the  $\alpha\text{H}$  protons of the preceding residue and the  $\alpha\text{H}$  or  $\delta\text{H}$  proline protons. This confirms that the product of synthesis and oxidation is correct with respect to the peptide sequence. Furthermore, the presence of two distinct species related by some common connectivity patterns suggests that conformational isomerism is taking place.

The intensities of the cross peaks are very similar between the conformers, indicating that the conformers are present in approximately equal proportion. However, one set of cross peaks due to one conformer is more easily recognised and was designated as isomer 1 and the other by default, as isomer 2. There was only one troublesome assignment, which was that of P21 of isomer 2.

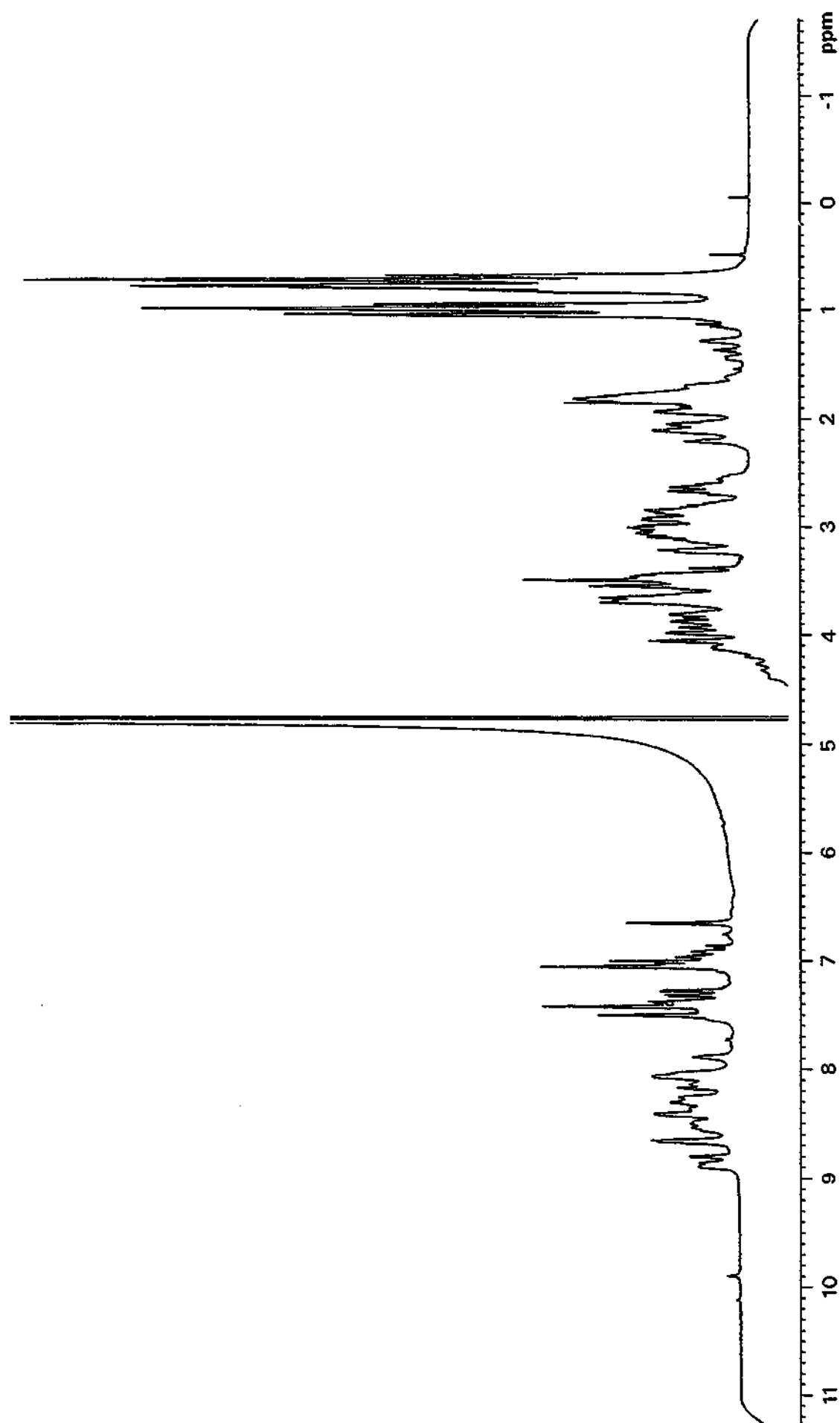
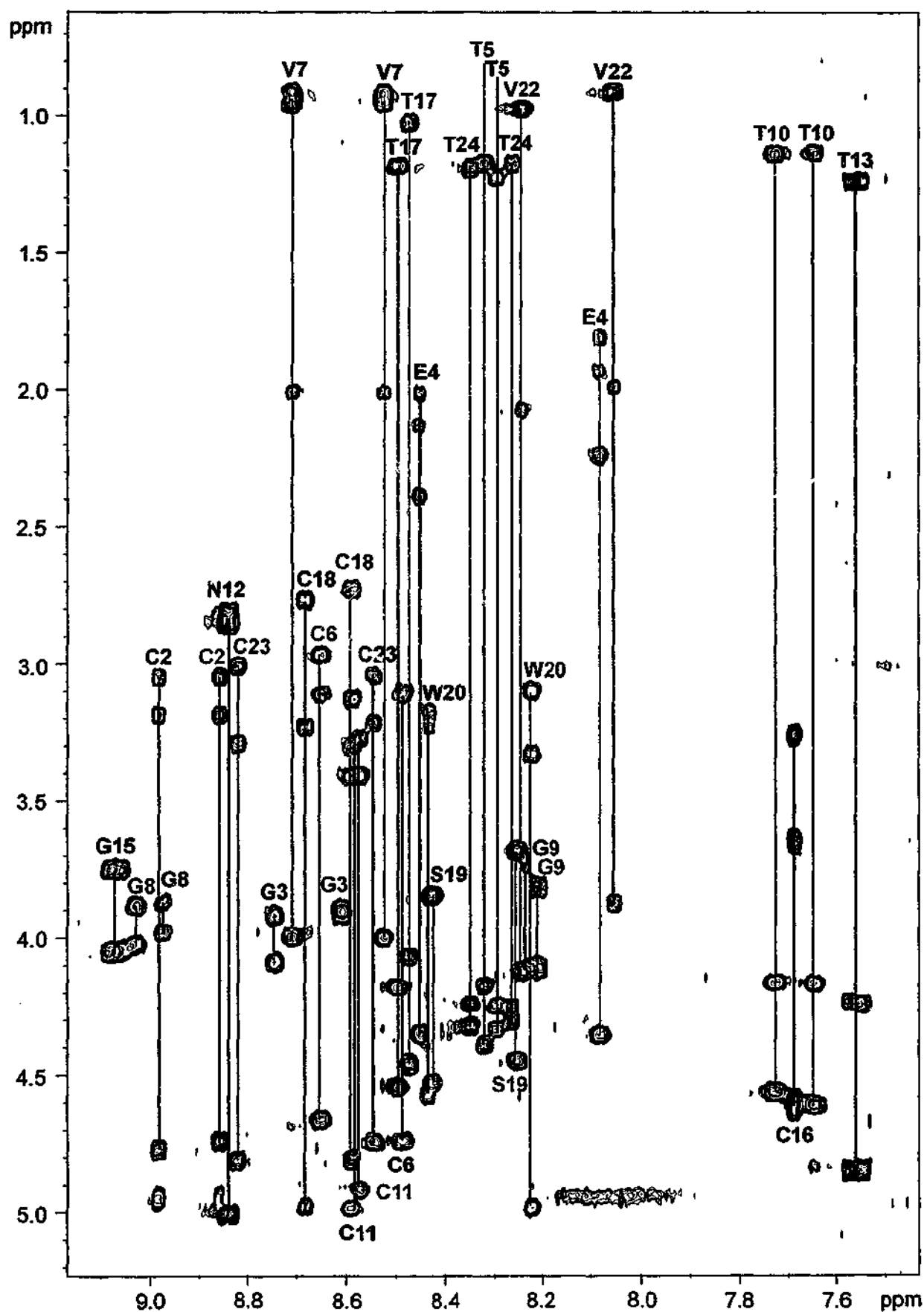
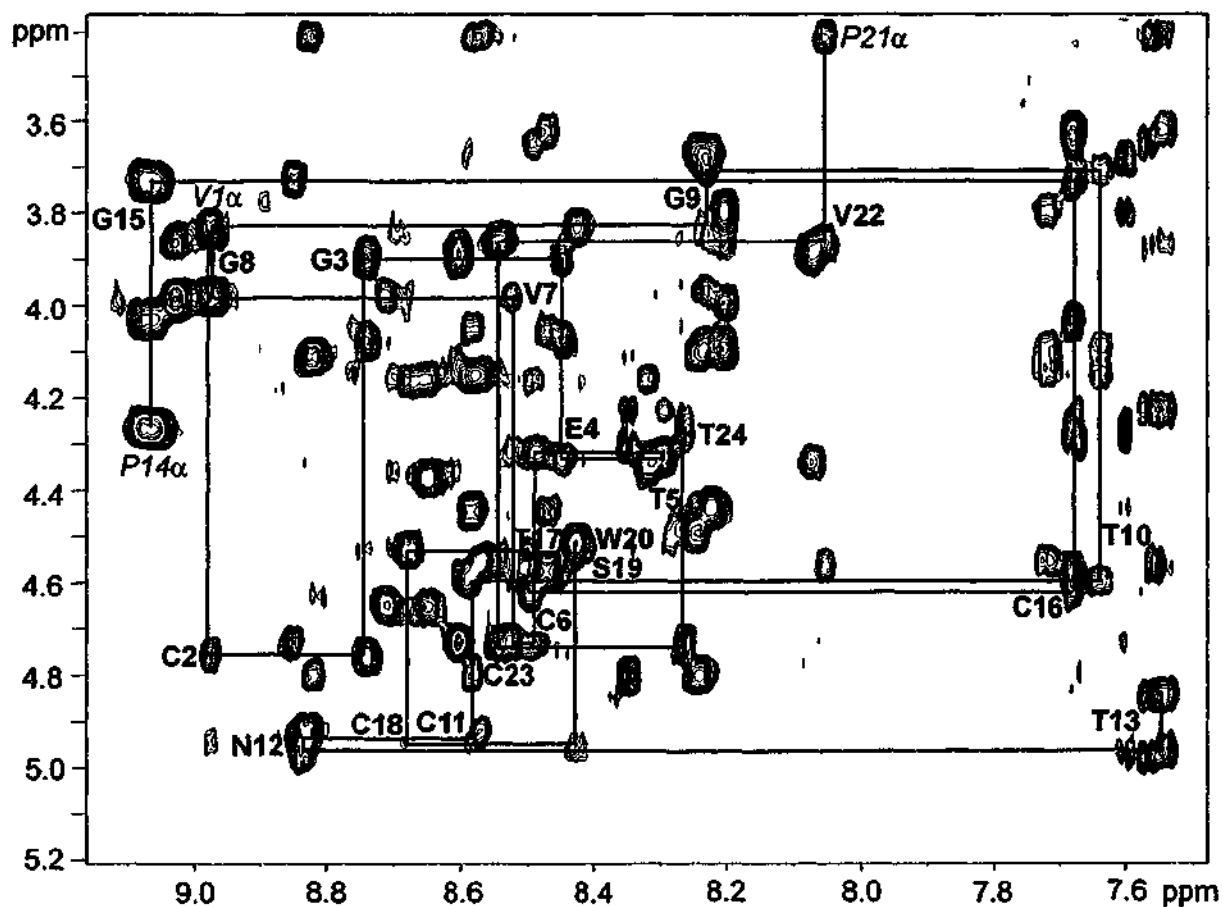
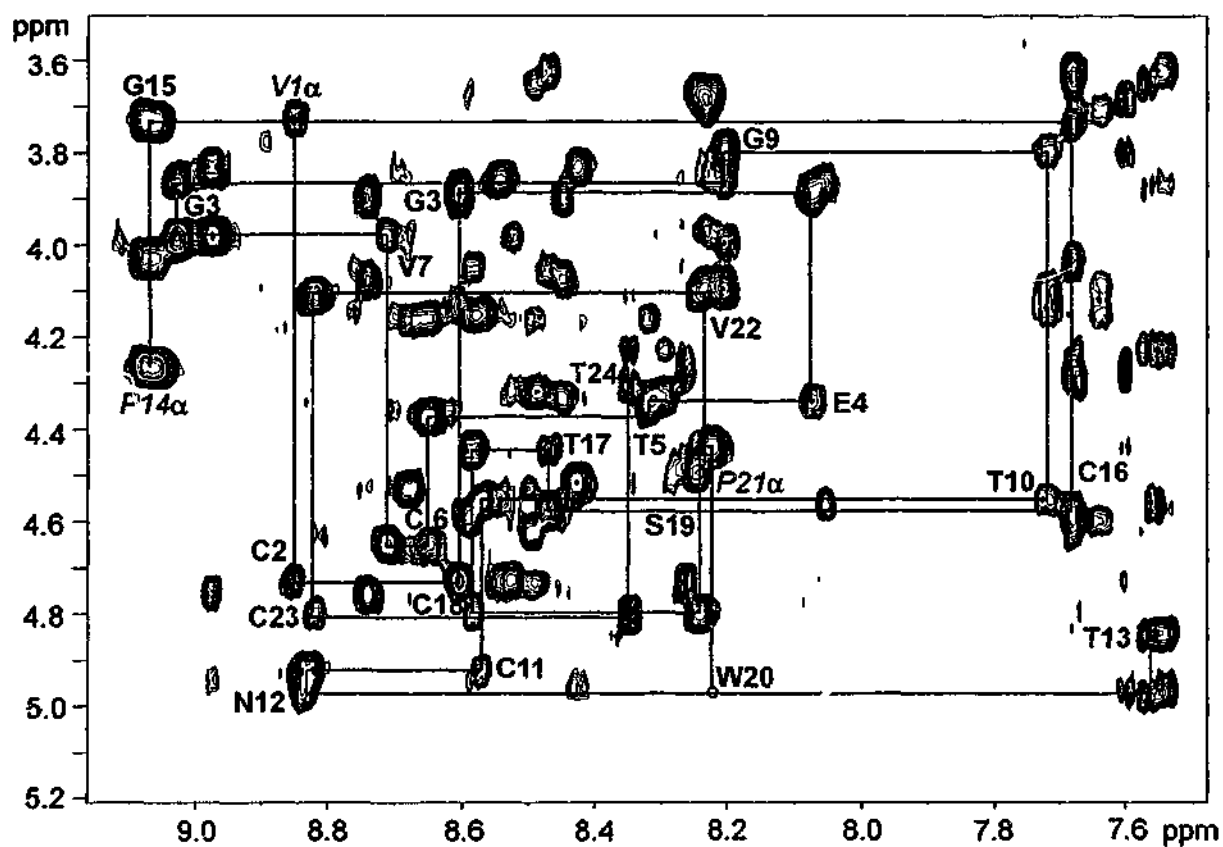


Figure 5.7. 1D Spectrum of acyclic permutant 6.



**Figure 5.8.** TOCSY fingerprint region of acyclic permutant 6. Connectivity patterns for isomer 1 are given in red and for isomer 2 are given in blue. Connectivity patterns common to both isomers are given in black.



**Figure 5.9.** NOESY fingerprint region showing the sequential connectivity pattern for isomer 1 (red) and isomer 2 (blue).

Assignment of the P21 alpha proton of isomer 2 was perplexing in that the alpha proton is virtually coincident with the P21 downfield delta proton. This was initially apparent by the slight shift of the centre of the cross peak due to degenerate gamma protons relative to the centres of the beta proton cross peaks along the practically superimposed alpha/downfield delta connectivity pattern in both the TOCSY and NOESY spectra.

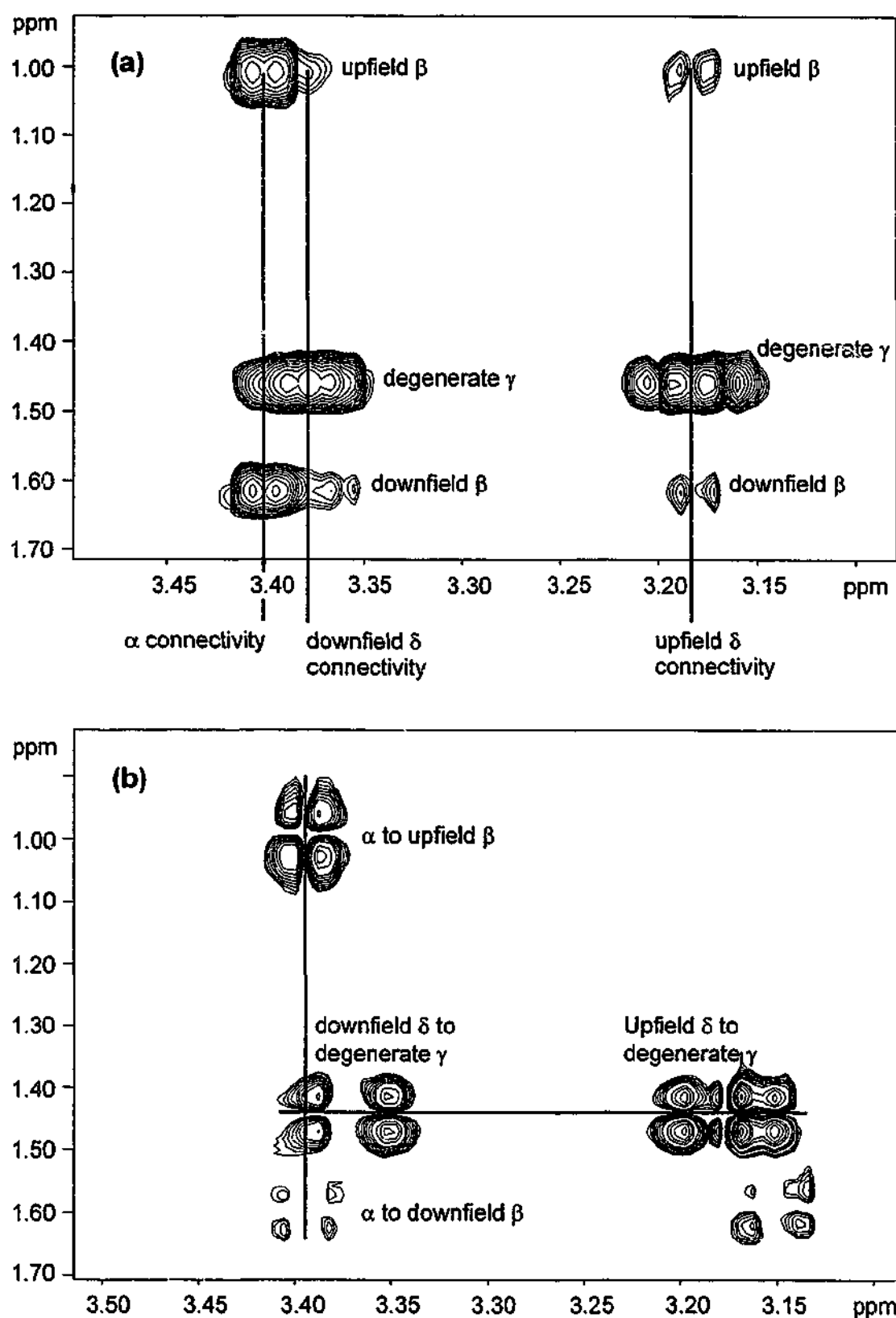
Confirmation for the overlap of the alpha and downfield delta protons was obtained from the DQF-COSY spectrum. Two cross peaks due to beta protons are present at the position of the alpha connectivity and there are strong cross peaks due to degenerate gamma protons along the two delta connectivities. Figure 5.10 shows the evidence for these assignments. With this final assignment confirmed, the chemical shift values for both isomers could be measured.

### 5.3.3 Proton Chemical Shifts and Structural Similarity

The  $^1\text{H}$  chemical shifts for each individual isomer are given in Tables 5.1 and 5.2. It is well known that chemical shift values, particularly those of alpha and amide protons, can give information on secondary structure and hydrogen bonding, respectively. Consequently, interesting observations on structural similarity can be obtained from chemical shift overlap between the isomers.

#### 5.3.3.1 Information from Alpha Protons

To illustrate the degree of chemical shift overlap between two isomers, the alpha protons were plotted on a graph in Figure 5.11. In the case of non-degenerate alpha protons of glycine, the average value is plotted. Bars below the graph represent differences between analogous alpha chemical shift values of the isomers. Corresponding alpha proton chemical shifts for kalata-B1 (Saether *et al.*, 1995) were also plotted on the graph for comparison. Several observations can be made from Figure 5.11.



**Figure 5.10.** Assignment of P21 alpha and deltas of isomer 2 in (a) NOESY (b) DQF-COSY spectra.

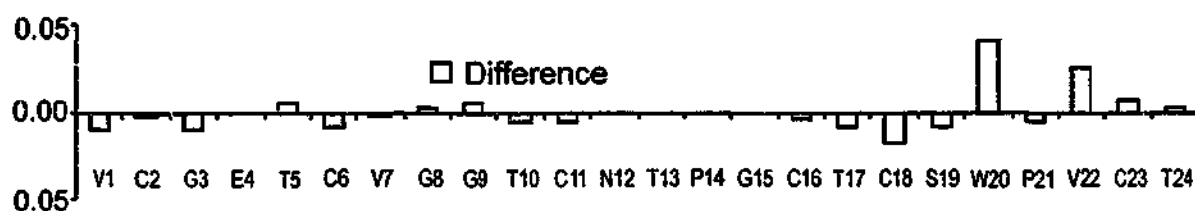
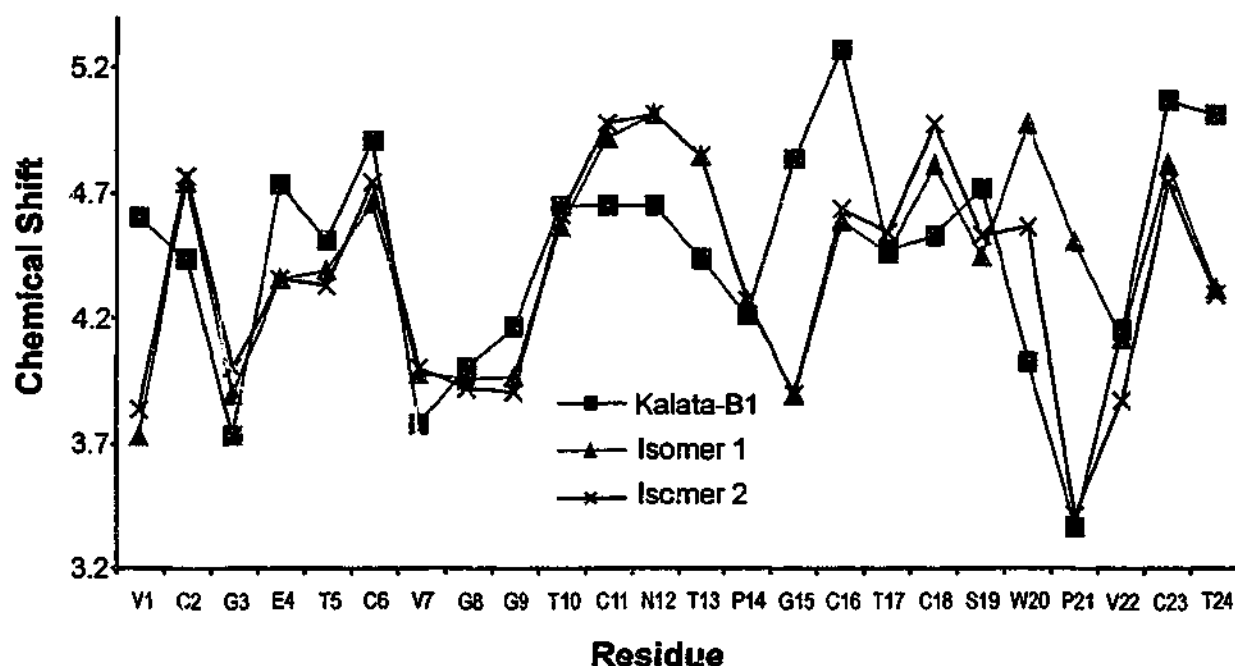
**Table 5.1.** Proton chemical shifts (ppm) for isomer 1 of acyclic permutant 6 at 282 K. Shifts are measured at 600 MHz and referenced to water externally calibrated against 2,2-dimethylsilapentane-5-sulfonate (DDS).

	RESIDUE	HN	$\alpha$ H	$\beta$ H	OTHER	
1.	VAL	-	3.73	2.00	$\gamma$ CH <sub>3</sub> 0.85	
2.	CYS	8.86	4.74	3.19, 3.05		
3.	GLY	8.61	3.90			
4.	GLU	8.08	4.35	1.93, 1.81	$\gamma$ CH <sub>2</sub> 2.24	
5.	THR	8.32	4.39	4.17	$\gamma$ CH <sub>3</sub> 1.17	
6.	CYS	8.65	4.66	3.11, 2.96		
7.	VAL	8.71	3.98	2.01	$\gamma$ CH <sub>3</sub> 0.96, 0.91	
8.	GLY	9.02	4.03, 3.88			
9.	GLY	8.20	4.10, 3.82			
10.	THR	7.73	4.56	4.16	$\gamma$ CH <sub>3</sub> 1.14	
11.	CYS	8.57	4.92	3.41, 3.27		
12.	ASN	8.84	5.01	2.84		
13.	THR	7.58/7.54	4.85	4.24	$\gamma$ CH <sub>3</sub> 1.24	
14.	PRO		4.27	2.30, 1.88	$\gamma$ CH <sub>2</sub> 2.12, 1.96	$\delta$ CH <sub>2</sub> 3.88, 3.62
15.	GLY	9.07	4.05, 3.75			
16.	CYS	7.69	4.59	3.62, 3.25		
17.	THR	8.47	4.46	4.07	$\gamma$ CH <sub>3</sub> 1.02	
18.	CYS	8.59	4.81	3.13, 2.73		
19.	SER	8.25	4.45	3.68		
20.	TRP	8.22	4.98	3.33, 3.10	2H 7.18 4H 7.61 5H 7.09 6H 7.19 7H 7.45 NH 10.07	
21.	PRO		4.51	2.28, 1.95,	$\gamma$ CH <sub>2</sub> 1.99	$\delta$ CH <sub>2</sub> 3.81, 3.70
22.	VAL	8.24	4.12	2.08	$\gamma$ CH <sub>3</sub> 0.96	
23.	CYS	8.82	4.81	3.29, 3.01		
24.	THR	8.35	4.32	4.24	$\gamma$ CH <sub>3</sub> 1.19	



**Table 5.2.** Proton chemical shifts (ppm) for isomer 2 of acyclic permutant 6 at 282 K. Shifts are measured at 600 MHz and referenced to water externally calibrated against DDS.

	RESIDUE	HN	$\alpha$ H	$\beta$ H	OTHER	
1.	VAL	-	3.83	2.21	$\gamma$ CH <sub>3</sub> 0.96	
2.	CYS	8.98	4.77	3.19, 3.05		
3.	GLY	8.74	4.09, 3.92			
4.	GLU	8.45	4.35	2.13, 2.01	$\gamma$ CH <sub>2</sub> 2.39	
5.	THR	8.30	4.33	4.25	$\gamma$ CH <sub>3</sub> 1.23	
6.	CYS	8.49	4.74	3.10		
7.	VAL	8.52	4.00	2.00	$\gamma$ CH <sub>3</sub> 0.95, 0.92	
8.	GLY	8.98	3.98, 3.86			
9.	GLY	8.23	4.10, 3.72			
10.	THR	7.65	4.61	4.16	$\gamma$ CH <sub>3</sub> 1.13	
11.	CYS	8.59	4.98	3.41, 3.31		
12.	ASN	8.84	5.01	2.84		
13.	THR	7.58/7.54	4.85	4.24	$\gamma$ CH <sub>3</sub> 1.24	
14.	PRO		4.27	2.30, 1.88	$\gamma$ CH <sub>2</sub> 2.12, 1.96	$\delta$ CH <sub>2</sub> 3.88, 3.62
15.	GLY	9.07	4.05, 3.75			
16.	CYS	7.69	4.63	3.65, 3.22		
17.	THR	8.49	4.54	4.18	$\gamma$ CH <sub>3</sub> 1.18	
18.	CYS	8.68	4.98	3.23, 2.76		
19.	SER	8.42	4.53	3.84		
20.	TRP	8.43	4.57	3.24, 3.17	2H 7.22 4H 7.55 5H 7.15 6H 7.23 7H 7.49 NH 10.29	
21.	PRO		3.41	1.62, 1.02	$\gamma$ CH <sub>2</sub> 1.47	$\delta$ CH <sub>2</sub> 3.40, 3.19
22.	VAL	8.06	3.87	1.99	$\gamma$ CH <sub>3</sub> 0.91	
23.	CYS	8.54	4.74	3.21, 3.04		
24.	THR	8.26	4.30	4.26	$\gamma$ CH <sub>3</sub> 1.17	



**Figure 5.11.** Graph of alpha proton chemical shifts of each residue for isomer 1 (line of red filled triangles), isomer 2 (line of blue crosses) and corresponding residues of kalata-B1 (line of green filled squares). For non-degenerate glycines, average values are plotted. Bars for each residue below the graph represent differences between analogous alpha proton chemical shifts of the isomers. A positive bar indicates that the alpha proton resonates downfield in isomer 1 from that proton in isomer 2.

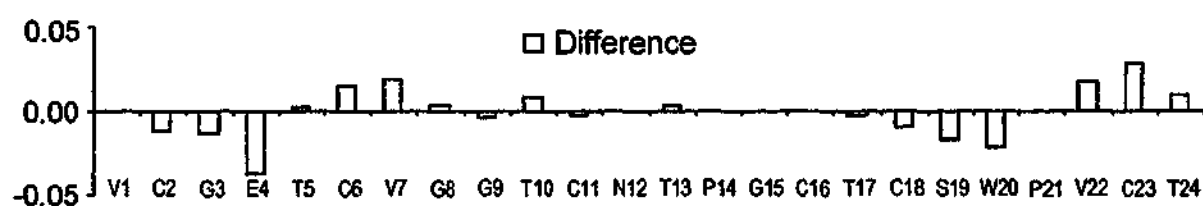
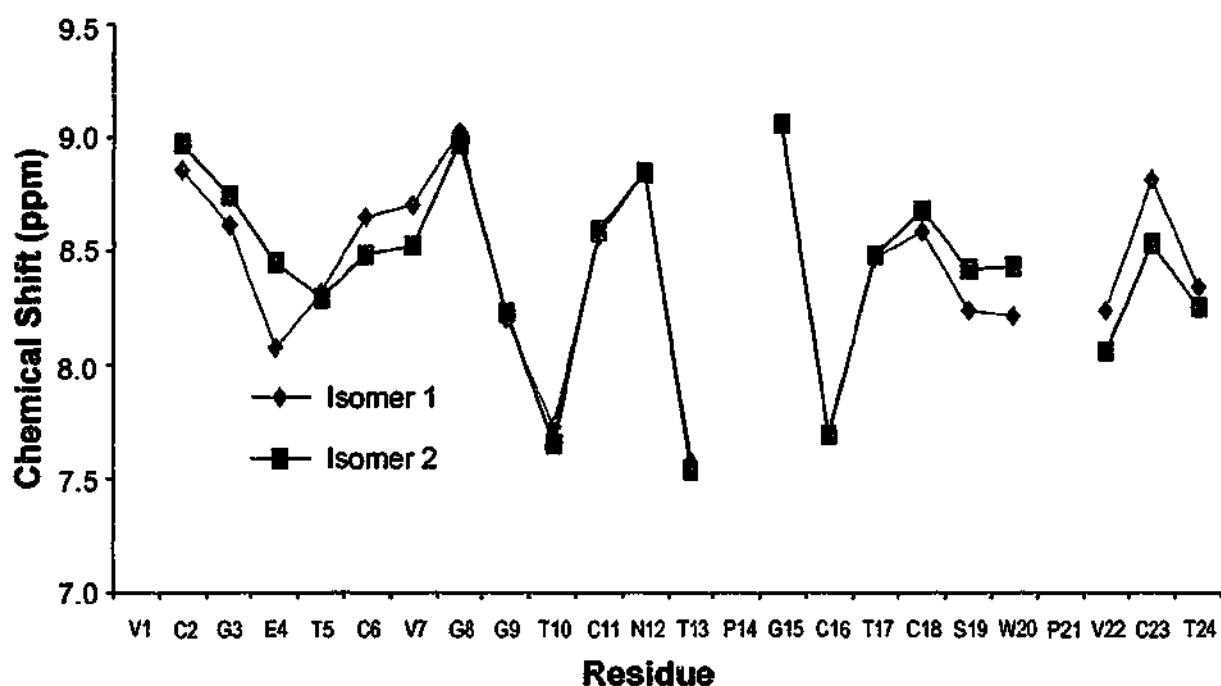
Firstly, the alpha protons for both isomers are identical over residues N12, T13, P14, G15. It seems that in this region of the peptides, the isomers are structurally very similar, if not identical. Secondly, the largest difference between alpha chemical shift values of the isomers occurs over residues W20 (0.41 ppm), P21 (1.1 ppm) and V22 (0.25 ppm). In this region of the sequence, it appears that there is likely to be a notable difference in structure. Thirdly, the corresponding chemical shift values vary in synchrony along the sequence between the isomers.

For the most part, the corresponding alpha proton chemical shifts across the sequence are close in value except for the region near residues including W20, P21 and V22. This would imply that overall the isomers are quite similar in structure with some important structural difference occurring around P21. This finding is in contrast with comparison to alpha chemical shifts of kalata-B1.

Whilst in part there is similar variance of alpha chemical shifts of kalata-B1 with equivalent values in the isomers, the differences are much larger and more pronounced especially for certain residues. Generally, the alpha chemical shifts of kalata-B1 differ from those of the isomers over the entire sequence, except for residues G8, T10, P14 and T17, where the alpha chemical shifts coincide in value. This would suggest that the isomers have deviated significantly in structure from that of the parent, native peptide. Amide proton chemical shifts gave further insight into the differences between the isomers.

#### 5.3.3.2 Information from Amide Protons

The hydrogen bonding in peptides and proteins is sensitive to changes in structure and although there may not be a major secondary structure difference between the isomers, any slight change may disrupt the hydrogen bonding. These changes are manifest in the amide proton chemical shifts. The amide proton chemical shifts for each isomer were plotted on a graph given in Figure 5.12. Bars below the graph represent differences in amide



**Figure 5.12.** Graph of amide proton chemical shifts of each residue for isomer 1 (line of red filled triangles) and isomer 2 (line of blue filled squares). Bars for each residue below the graph represent differences between analogous amide proton chemical shifts for the isomers. A positive bar indicates that the proton chemical shift resonates downfield in isomer 1 from that value in isomer 2.

proton chemical shifts between analogous amide protons of the isomer. The following observations can be made from Figure 5.12.

The central region over residues N12, T13, G15, C16 is overlapped, while over residues S19, W20, V22, C23 large differences occur, especially at W20. This is in accord with results from alpha chemical shift data. In addition, however, the ends of the sequence, also have large differences over C2, G3, E4, C6, V7 at the beginning of the sequence and over V22, C23, T24 at the end of the sequence. It is probable that the larger differences observed for the end of the sequence are due to mechanisms operating at P21, while the larger differences observed at the beginning of the sequence may simply be the effect of increased mobility often found at the ends of peptides and proteins.

Combining all the chemical shift data gives some indication of the structural similarity between isomers. In summary, the isomers seem to preserve structure in the central region of the sequence over R8 - T17, where the structures appear to be identical. On one side of this central region, towards the end of the sequence, there is evidence for significant structural difference over C18 - C23. This difference is most pronounced at P21. At the beginning of the polypeptide chain, on the other side of the central region, there is also subtle structure difference. It is speculated that there may be increased mobility at the beginning of the sequence. The final and most interesting finding concerns comparison of the isomers to kalata-B1. It seems that the structures of the isomers are considerably different from that of kalata-B1. To obtain more specific information on the presence of secondary structure, CSI indices based on alpha proton chemical shifts were calculated.

#### **5.3.4 Secondary Structure Identification from CSI Values**

A method developed to identify  $\alpha$ -helix and  $\beta$ -strand secondary structure from trends of chemical shifts deviating from random coil values was applied (Wishart *et al.*, 1992). CSI values of the alpha protons for each

isomer were calculated and graphically presented in Figure 5.13. CSI values for corresponding alpha chemical shifts (Saether *et al.*, 1995) in kalata-B1 were also calculated and graphically presented in Figure 5.13, in order to compare structure between kalata-B1 and its fragment isomers. Several interesting comparisons can be made between the isomers, as well as between the isomers and kalata-B1.

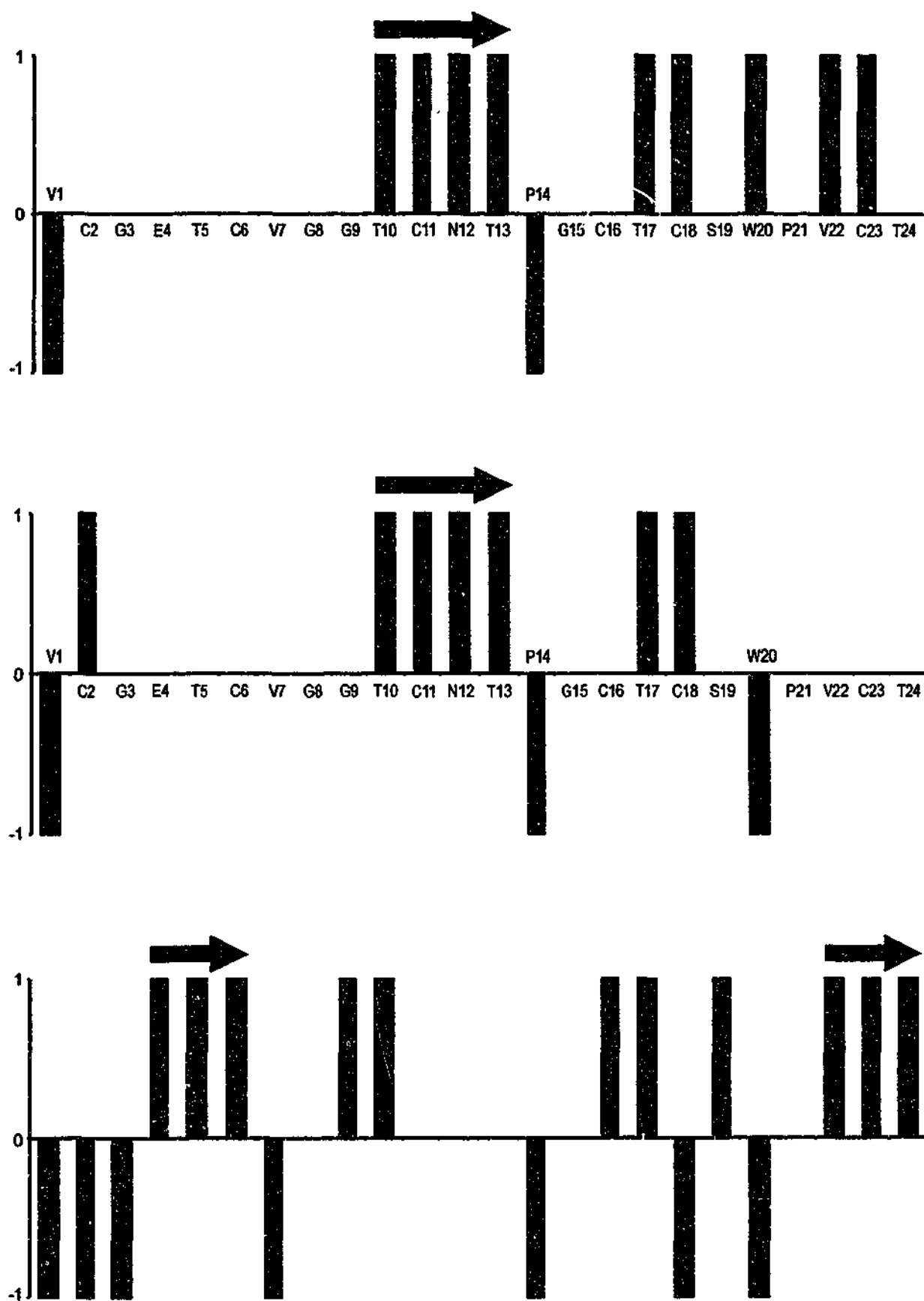
#### 5.3.4.1 CSI Values of Isomers and Comparison of their Structures

In comparing the isomers, the CSI diagrams show that CSI values are very similar for the isomers for most of the polypeptide chain. The alpha proton shifts at the beginning of the sequence, G3 - G9, are very close in value to their random coil values giving a consecutive series of zero CSI values which is associated with coil structure.

Four consecutive positive CSI values for T10, C11, N12 and T13, and a negative CSI value at P14 then follows, indicating that the central part of the isomers has  $\beta$ -strand character and that there are termination points on either side of the  $\beta$ -strand. The change in sign from T13 to P14 could indicate the presence of a reverse-turn or  $\beta$ -turn of some sort. This central section is the only part that has any defined secondary structure in the isomers.

Over the next few residues from G15 - C18, the CSI values for both isomers are identical but do not display any trends characteristic of secondary structure. A significant difference occurs at W20, where for isomer 1 the CSI value is positive, and for isomer 2 it is negative. It is only at this one residue, W20, where the CSI values are opposite in sign. This extreme change could imply that the difference between the isomers originates in this region. The remainder of the polypeptide chain, P21 - T24 is also different between the isomers and this may be the result of the difference manifested at W20.

While the structures of the fragment isomers are generally similar, further differences are observed in comparing the CSI values of the isomers with those of kalata-B1.



**Figure 5.13.** CSI values of alpha protons for corresponding residues of isomer 1 (red bars), isomer 2 (blue bars) and kalata-B1 (green bars). The prediction of the presence of  $\beta$ -strand is indicated by solid arrows.

#### 5.3.4.2 Comparison of Kalata-B1 with Fragment Isomer Structures

The CSI diagrams show that the CSI values are considerably altered between kalata-B1 and its fragment isomers. Unlike the fragment isomers, kalata-B1 has a large proportion of residues with negative and positive CSI values over the entire sequence, consistent with it being a highly structured molecule.

Two  $\beta$ -strand regions and turn features are evident in kalata-B1. One of these  $\beta$ -strands occurs in a region designated as coil in the isomers at the beginning of the isomer sequence. The other  $\beta$ -strand occurs at the end of the sequence of the isomers adjacent to the position where the major difference between the isomers has been identified. Mobility in this region of the isomers would explain the lack of defined structure. The one defined region of structure of the isomers, being the centrally located  $\beta$ -strand, occurs in a region corresponding to the connecting region of kalata-B1, which is a fairly featureless region for kalata-B1. However, there are some similarities in CSI values between the isomers and kalata-B1.

As would be expected from the alpha proton chemical shifts, coincident CSI values between the isomers and kalata-B1 occur for G8, T10, P14 and T17. Of these residues, P14 is prominent. In the isomers P14, according to the CSI analysis, could be a reverse-turn or  $\beta$ -turn at one end of the centrally located  $\beta$ -strand. The equivalent residue in kalata-B1 is involved in a type II  $\beta$ -turn (Saether *et al.*, 1995). It seems that a proline-driven turn occurs at the same point for the isomers and kalata-B1. Less subtle, but structurally noteworthy, is the coincident CSI value for G8. In the isomers, G8 forms part of a termination point at the other end of the centrally located  $\beta$ -strand. In kalata-B1, the equivalent residue is involved in a type I  $\beta$ -turn. It is possible that the turning points in the isomers and kalata-B1 are retained.

Overall, the CSI analysis indicates that the isomers are mainly coil in structure with a central region from T10 - T13 of defined  $\beta$ -strand, which is preserved between the isomers. Termination points occur on either side of the



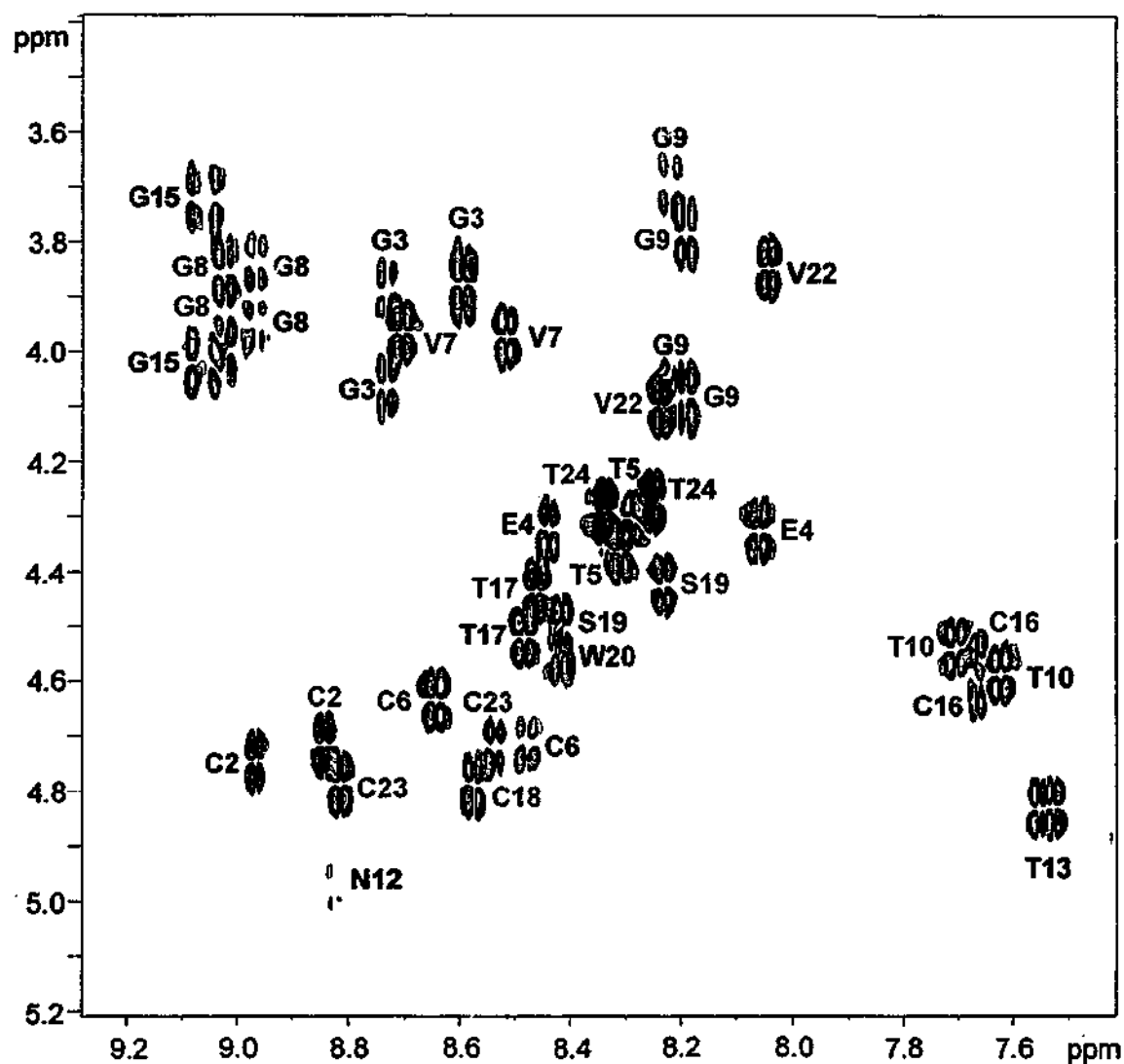
$\beta$ -strand and it is possible that there may be a reverse-turn or  $\beta$ -turn in the vicinity of P14. A major difference occurs at the latter section of the polypeptide chain around W20. These results are in agreement with structural similarities pre-empted by raw chemical shift data and are not generally, in synchrony, with those of kalata-B1.

In fact, the implications are that kalata-B1 and its fragment isomers have different topologies. Kalata-B1 is a more highly structured molecule, while the isomers display far less structure and also some mobility at the C-terminus. The most striking differences seem to be the existence of secondary structure in the isomers in a relatively featureless region of kalata-B1 and no defined structure in the isomers in regions of secondary structure in kalata-B1. Coupling constants, the magnitude of which are related to secondary structure can give further information on secondary structure.

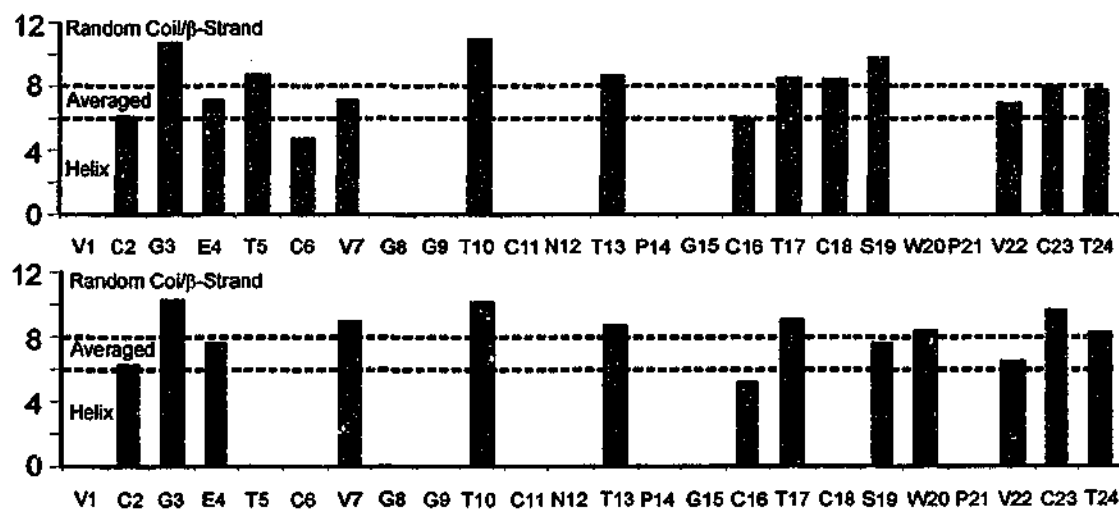
### 5.3.5 $^3J_{\text{NH-}\alpha\text{H}}$ Coupling Constants

$^3J_{\text{NH-}\alpha\text{H}}$  coupling constants were measured for both isomers from a DQF-COSY spectrum. Assignments of the NH- $\alpha$ H region of the DQF-COSY spectrum are shown in Figure 5.14. Separate displays of the  $^3J_{\text{NH-}\alpha\text{H}}$  coupling constants values for every residue for each of the isomers are given in Figure 5.15. Coupling constants that could not be measured due to interference from spectral overlap or noise, or could not be measured at all, such as those for prolines, glycines and the N-terminus, are omitted. The values provide further support for previous observations from other spectral parameters.

The displays indicate that  $^3J_{\text{NH-}\alpha\text{H}}$  coupling constants, overall, are very similar for both the isomers, supporting the finding that the isomers are structurally similar. Much of the isomers seem to have  $^3J_{\text{NH-}\alpha\text{H}}$  coupling constant values tending towards random coil/ $\beta$ -strand structure. Although a complete picture of the region of the central  $\beta$ -strand is not possible due to missing values, of those  $^3J_{\text{NH-}\alpha\text{H}}$  coupling constants that could be measured, the values are in accord with those of  $\beta$ -strand structures. At the beginning



**Figure 5.14.** Fingerprint region of DQF-COSY spectrum of acyclic permutant 6. HN- $\alpha$ H cross peaks for isomer 1 are labelled in red and for isomer 2 are labelled in blue. Those common to both the isomers are labelled in black.



**Figure 5.15.**  $^3J_{\text{NH}-\alpha\text{H}}$  coupling constant values for each residue of isomer 1 (blue bars) and isomer 2 (red bars). Those values that could not be measured due to overlap or absence are omitted.

and the end of the polypeptide chain, there is alternation between values consistent with substantially averaged helical conformation and random coil/ $\beta$ -strand. This could be a consequence of less definite structural definition in this region.

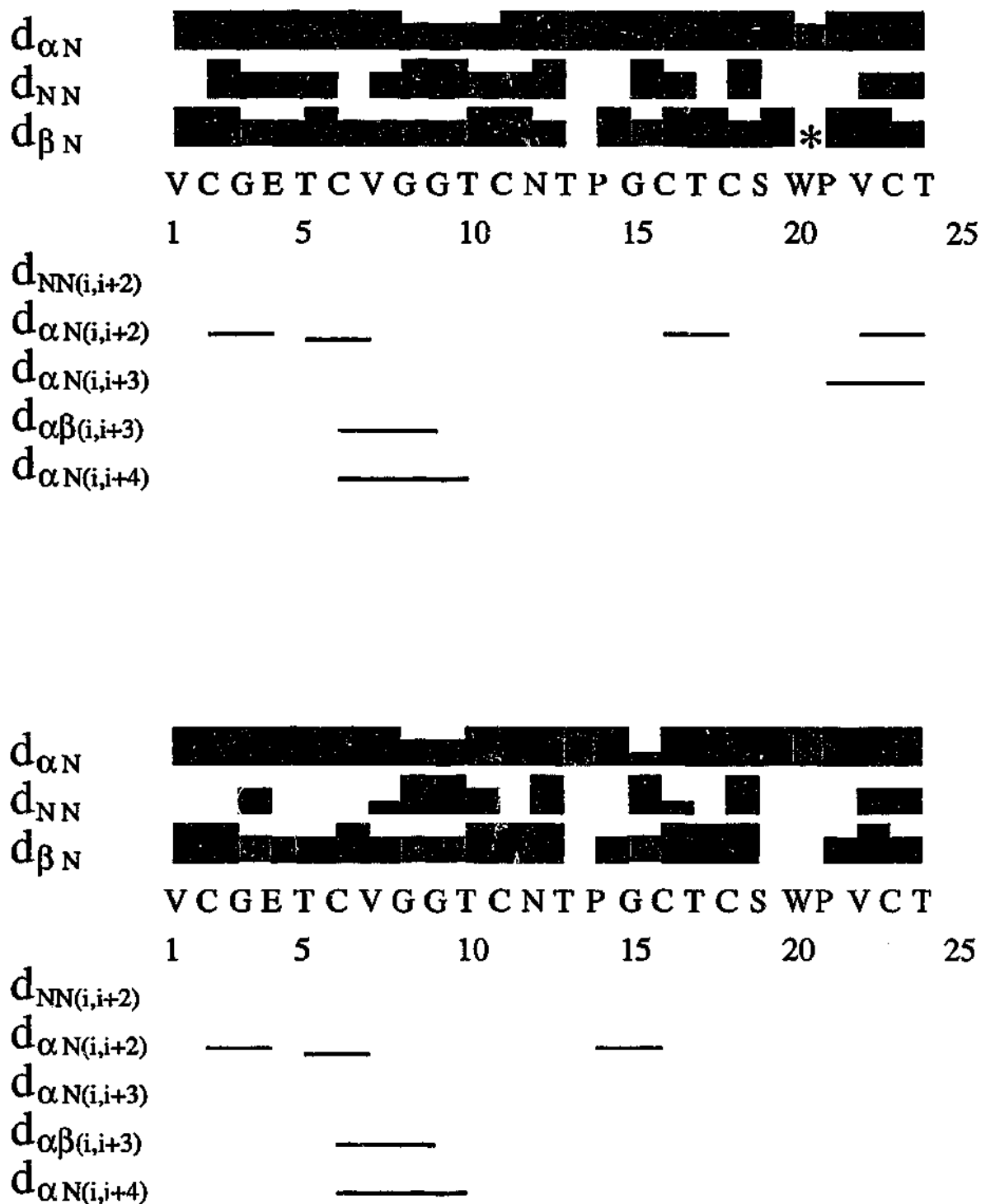
Overall, the  $^3J_{\text{NH}-\alpha\text{H}}$  coupling constants reinforce the similarity in structure between the isomers and are consistent with the presence of  $\beta$ -strand structure in the central region of the peptide. A final analysis of structure was obtained from NOE data.

### 5.3.6 Nuclear Overhauser Enhancements

Insufficient NOE connectivities were obtained from 2D Nuclear Overhauser experiments for serious simulated annealing molecular dynamic calculations. However, a variety of medium to long range connectivities were identified, in addition to intra and neighboring NOE connectivities. These NOE connectivities could be utilised in a qualitative discussion on structure. In first instance, it is instructive to analyse short and medium range NOE connectivities.

#### 5.3.6.1 Short and Medium Range NOE Connectivities

A summary of short to medium range NOES connectivities is graphically presented in Figure 5.16. As pre-empted by results thus far, the data are similar for both isomers. There are sequential NOE connectivities observed for most residues. However, little can be said concerning the presence of secondary structure, as most of these are of similar medium to strong intensity. The only notable difference between the isomers in the sequential NOE connectivities occurs around C2 - C6. In this region, for isomer 1  $d_{\text{N,N}}(i,i+1)$  connectivities are of medium intensity from G3 - C6 and of strong intensity around G3 - C2, but for isomer 2,  $d_{\text{N,N}}(i,i+1)$  connectivities, are absent except around E4 - G3 where the NOE is of medium intensity. This



**Figure 5.16.** Summary of short and medium range NOE connectivities measured for isomer 1 (top) and isomer 2 (bottom). NOES between neighboring residues are shown as filled blocks and their intensities are indicated by the height of the block, being higher for more intense cross peaks. Medium range NOES are shown by a bar drawn between the connected residues. Overlapping or ambiguous NOES are indicated by \*. NOES that are not expected to be present due to the lack of specified protons (ie Gly  $\beta$ H, Pro NH) are shaded.

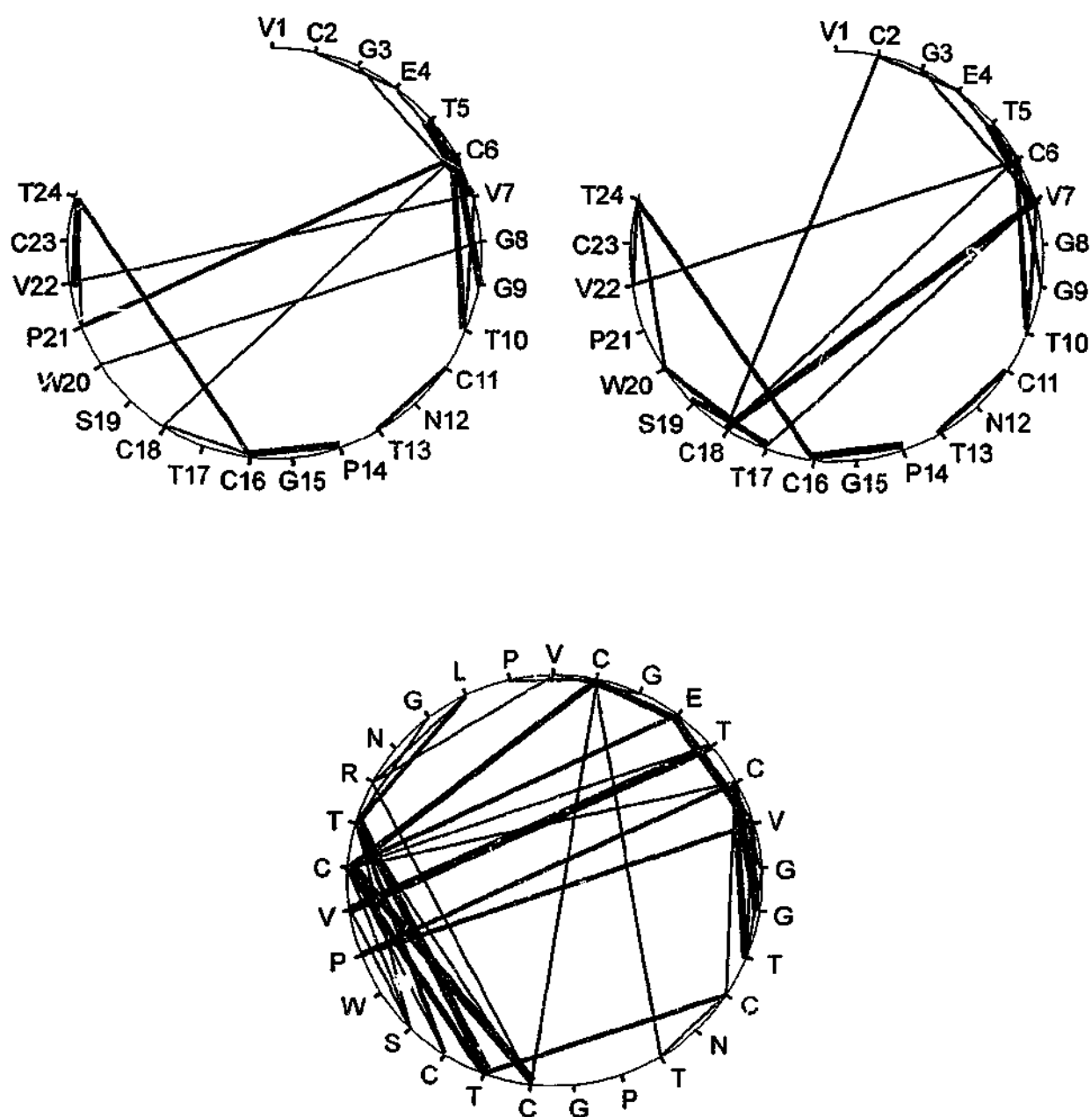
could suggest some possible structural difference in the region of C2 - C5 between the isomers. Medium range NOE connectivities give other indications.

The absence of any series of (i, i + 2) or (i, i + 3) NOE connectivities confirms the lack of  $\alpha$ -helix character for the isomers. However, tight turns are perhaps indicated by the presence of  $d_{\alpha,N}(i, i + 2)$  at E4 - C2, V7 - T5 for both isomers, and at C18 - C16 for isomer 1, and at C16 - P14 for isomer 2. These turns are located on either side of the  $\beta$ -strand character identified over T10 - T13 through the CSI analysis. The presence of some sort of a turn at around V7 - T5 for both isomers is reinforced by  $d_{\alpha,\beta}(i, i + 3)$  NOE connectivity at G9 - C6 and  $d_{\alpha,N}(i, i + 4)$  NOE connectivity at T10 - C6.

Finally, there appears to an additional turn in isomer 1, not evident in isomer 2 that is located at the C-terminus. NOE connectivities  $d_{\alpha,N}(i, i + 2)$  at V22 - T24 and  $d_{\alpha,N}(i, i + 3)$  at P21 - T24 occur near the C-terminus for isomer 1. The combination of the slightly different location of the third tight turn at C18 - C16 for isomer 1, the impression of a turn at the C-terminus, and the large chemical shift data differences imply a substantial local perturbation in structure in this region. These differences between the isomers may be attributable to *cis/trans* proline isomerisation involving P21. This is highly likely given that proline isomerisation is the most common cause of conformational change (Kessler *et al.*, 1996). Examination of medium and long range NOE connectivities and their composition with those of kalata-B1 gives more idea on overall structure.

#### 5.3.6.2 Medium and Long Range NOE Connectivities

The spatial distributions of medium to long range NOE connectivities for each isomer are illustrated in Figure 5.17. The thickness of the lines are proportional to the number of NOE connectivities linking the particular residues, so that, the thicker the line, the more NOE connectivities were observed between the connected residues. For comparison to the parent peptide, kalata-B1, a similar corresponding illustration is included in Figure



**Figure 5.17.** Spatial distribution of corresponding medium and long range NOE connectivities for isomer 1 (diagram with red lines), isomer 2 (diagram with blue lines) and kalata-B1 (diagram with green lines) (Craig, 1994). The thickness of the lines is proportional to the number of NOE connectivities between the residues, so that, the thicker the line, the more NOE connectivities exist between the residues.

5.17 (Craik, 1994). Immediately apparent from Figure 5.17 is, once again, the similarity between the isomers.

Several similar NOE connectivities between distant residues, in addition to the medium range NOE connectivities were observed for both isomers. The distribution pattern further emphasises the overall  $\beta$ -character of the isomers in that there are several long range NOE connectivities typical of strands lying in close proximity and also many medium NOE connectivities characteristic of turns. The region T5 - T10 is the most well defined region, with many medium range NOE connectivities found in the region.

Differences between the isomers relate to mainly the exact location of the long range NOE connectivities. Most of these emanate from the T5 - T10 region but the connected residue tends to be spread over C18 to V22 for isomer 1 and localised around C18 for isomer 2. This serves as further evidence that a significant difference between the isomers originates in the region near W20. The other main difference is a long range NOE connectivity between C2 and C18 which is present for isomer 2 but absent for isomer 1.

Lastly, the presence of structure implicated at the end of the polypeptide chain for isomer 1 in the previous section seems to also be present for isomer 2, though possibly more subtle in the case of isomer 2. Albeit, the spatial distribution of medium and long range NOE connectivities for the isomers is different to that of kalata-B1.

The prime difference is that there are fewer NOE connectivities for the isomers than for kalata-B1. Evidently, kalata-B1 has considerably more defined structure. For instance, the isomers do not display NOE connectivities between P21 - C23 to C2 - V7 and have very few NOE connectivities between C16 - T24. These regions are abundant with NOE connectivities in kalata-B1. However, the isomers do have NOE connectivities linking residues around C18 to residues around V7 that are not evident in kalata-B1. One region in which the isomers may have retained structure from kalata-B1 is in the region of T5 - T10, perhaps extending to C2, where there are many NOE connectivities in the isomers as well as kalata-B1.

The similarity in NOE connectivities between the isomers and kalata-B1 could suggest that the native disulphide bonding of kalata-B1 in this region is retained in this form of acyclic permutant 6. With regard to the representation of acyclic permutant 6 as derived from kalata-B1 (Figure 5.1), the disulphide bond that is significant in the region is Cys6-Cys18. Further evidence for the possible disulphide bonding pattern in acyclic permutant 6 was obtained from the examination of NOE connectivities between cysteine residues. In a study on the ability of NMR data to predict disulphide bonding patterns, NOE cross peaks between beta to beta protons and alpha to beta protons of linked disulphides are strongly predictive for the identification of the bond. By contrast, NOE cross peaks between amide to alpha protons of two cysteines are an indication that the cysteines are not bonded (Klaus *et al.*, 1993).

In line with these findings, a couple of NOE cross peaks from the amide proton of Cys18 to alpha proton of Cys16 and from the amide proton of Cys18 to the alpha proton of Cys23 infer that Cys18 is not linked to either Cys16 nor Cys23. However, Cys16 and Cys23 may be connected, as there are NOE cross peaks between each of the beta protons of Cys16 and the alpha proton of Cys23. With four of the six cysteines paired, a link between Cys2 and Cys11 is implied by default. Using the labeling of the cysteines of the global alignment of members of the inhibitor cystine knot family, the apparent disulphide bonding pattern of acyclic permutant 6 in this study is Cys(1)-Cys(3), Cys(2)-Cys(5) and Cys(4)-Cys(6). It seems that only one of the native disulphide bonds of kalata-B1 is retained, while the other two disulphides are interchanged. This particular disulphide bonding pattern explains many of the observations concerning the isomers.

The non-native pairing of Cys2-Cys11 explains the loss of NOE connectivities between C2 to T13 and C2 to C16, which are prominent for kalata-B1. The Cys2-Cys11 disulphide bond could be expected to result in an increased distance between C2 and T13, and C2 and C16, as C2 is now closer to C11. Moreover, the disulphide bonding pattern no longer forms the threaded, restrictive cystine knot. The ring of kalata-B1 is broken so that, the



backbone is easier to unravel. The 3D structure would be more open and it may be envisaged that the resulting structure could accommodate *cis/trans* isomerisation of P21.

Overall, these observations from the NOE data, suggest that the isomers have lost structure associated kalata-B1, including the  $\beta$ -hairpin and the disulphide bonding pattern, but may have retained at least, partially, some of the structure of the connecting region, as well as forming additional  $\beta$ -strand structure within that region, not obvious in kalata-B1. A consequence of this could be that the overall fold of the fragment isomers differs somewhat from that of the parent biomolecule, kalata-B1. Furthermore, this new fold, by virtue of the existence of isomeric forms, must inherently be less stable than kalata-B1. The cause of this instability is the subject of the next section.

#### 5.3.7 *Cis/Trans* Proline Isomerism

There are several phenomena which lead to isomerism in proteins and peptides. One of the most well established is *cis/trans* isomerism. The isomers of acyclic permutant 6 have two prolines, P14 and P21, which could potentially be involved in isomerism. The determination of proline isomerism from spectral data, as discussed in Chapter 2, was applied to the prolines of the isomers.

##### 5.3.7.1 P14

P14 is located in the best defined and probably the most stable structural region of the isomers. The residue appears to partake in  $\beta$ -strand secondary structure for the both isomers and the chemical shifts of the protons of each isomer are identical. Given this stability, it is not surprising that NOESY and HMQC data indicate that P14 adopts the more common and more stable *trans* configuration.

In NOESY spectra, NOE cross peaks of the two delta prolines of P14 are evident along the T13 alpha and NH connectivity patterns and these assignments are unique and consistent with the *trans* form. HMQC data also infers P14 is *trans*, with a difference of 4.15 ppm found between the  $^{13}\text{C}_\beta$  and  $^{13}\text{C}_\gamma$  chemical shifts, although, the HMQC data is less certain, as assignment of  $^{13}\text{C}_\gamma$  occurs in a region of spectral overlap. The determination of configuration of P21 is more interesting.

#### 5.3.7.2 P21

P21 is located in a less defined region of the isomers in terms of secondary structure. Much of the data analysed and reported in the previous sections of this chapter suggests that a major difference occurs between the isomers in the vicinity of P21. Unlike P14, the proton and carbon chemical shifts of P21 are different between the isomers. It is interesting to note that P21 of acyclic permutant 6 corresponds to the *cis* proline of kalata-B1, established in Chapter 4. P21 is a most likely candidate for *cis/trans* isomerism and indeed, it was found that *cis/trans* isomerism is occurring at P21.

In the NOESY spectra, two NOE cross peaks due to delta protons of P21 of isomer 1 are definitely observed along the W20 alpha connectivity pattern of isomer 1. This is strong evidence supporting that P21 of isomer 1 adopts the *trans* configuration. In line with this finding, although less convincing due to spectral overlap, the difference in  $^{13}\text{C}_\beta$  and  $^{13}\text{C}_\gamma$  chemical shifts is around 4.30 ppm for isomer 1, which is a value typically obtained for *trans* prolines.

The *cis* configuration of P21 of isomer 2 is unequivocally indicated by HMQC data. A difference of 9.28 ppm between  $^{13}\text{C}_\beta$  and  $^{13}\text{C}_\gamma$  chemical shifts, which is a characteristic of *cis* prolines, is obtained. No spectral overlap or artifacts interfere with the  $^{13}\text{C}_\beta$  and  $^{13}\text{C}_\gamma$  chemical shift assignments and so, there is a high degree of confidence in the determination of the proline

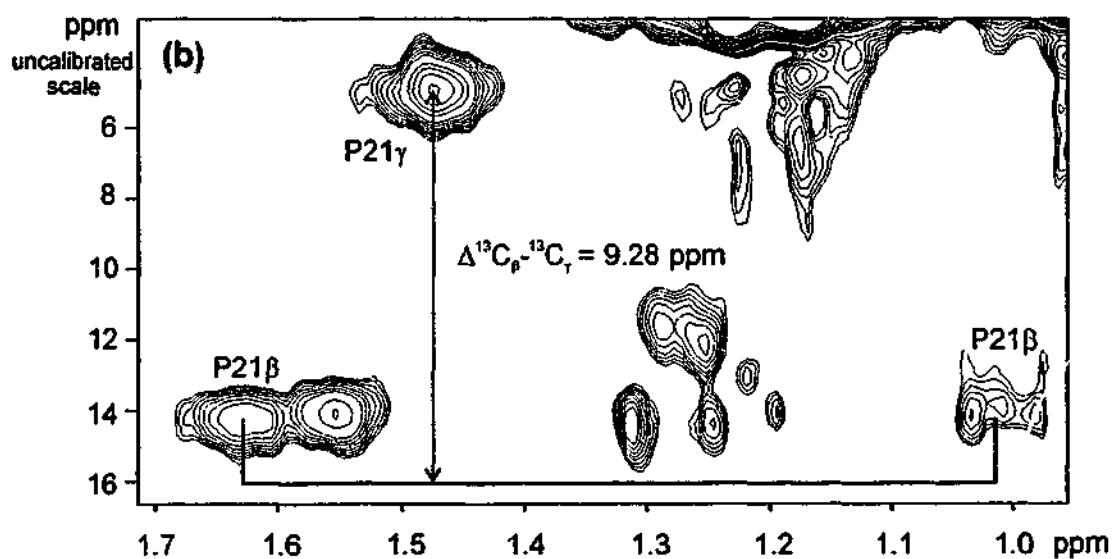
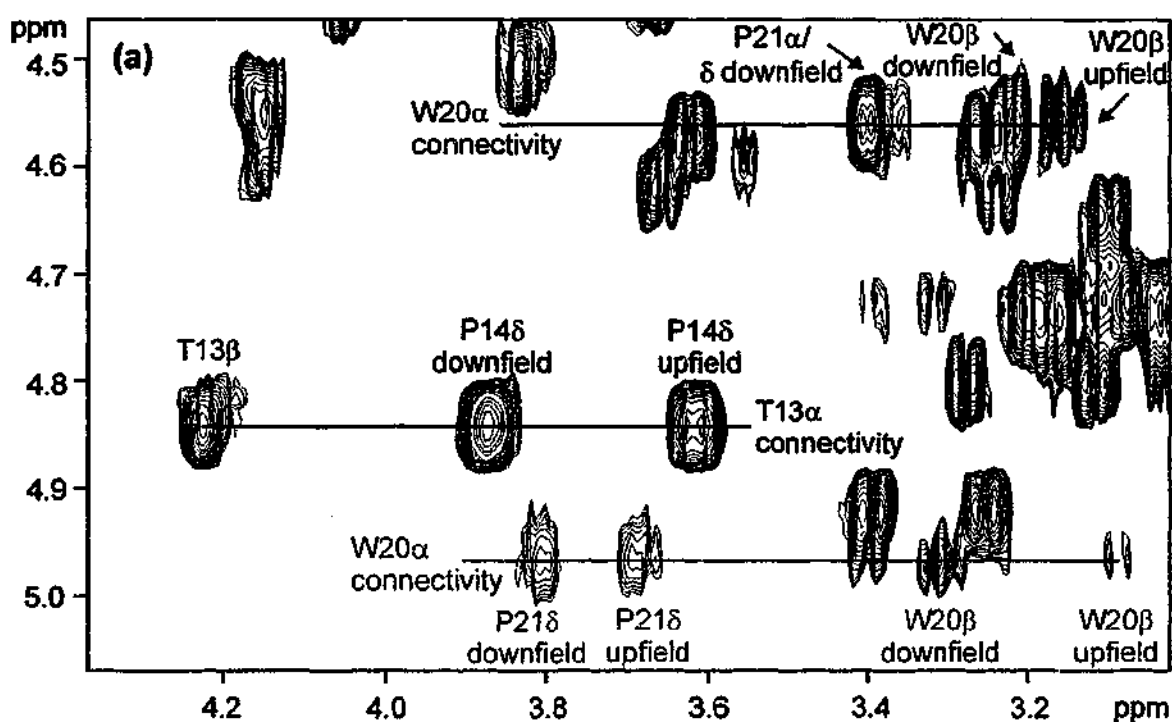
conformation for P21 of isomer 2. The NOESY data also indicated that P21 is a *cis* proline as only the overlapped alpha/downfield delta of P21 is clearly observed along the alpha connectivity pattern of W20 of isomer 2. The evidence from the NOESY data is less certain due to the potential for overlap of the W20 betas with the upfield P21 delta along the W20 alpha connectivity pattern. Nonetheless, overall, there is conclusive spectral data to show that *cis/trans* proline isomerism is taking place at P21. These data are displayed in Figure 5.18.

The occurrence of *cis/trans* proline isomerism of P21 explains the existence of the two distinct species, their general similarity and the difference identified in the region of P21. This completes the results obtained for acyclic permutant 6. The following and final section of this chapter summarises these results in light of the work carried out on acyclic permutant 6.

#### 5.4 Summary

A synthesised peptide with the sequence of acyclic permutant 6 proposed in Chapter 3 was obtained and oxidised. The predominant oxidation product was isolated, concentrated and analysed by 2D NMR spectroscopy. The main fraction was found to consist of two distinct species, which are the result of *cis/trans* proline isomerism taking place at one of the two prolines, namely P21.

The existence of *cis/trans* isomerism of P21 was indicated by several measured NMR spectral parameters including, alpha and amide proton chemical shift data, CSI values, NOE connectivities, and was confirmed by the observation of definitive cross peaks in HMQC spectra. The <sup>1</sup>H chemical shifts were completely assigned to two separate isomers. The slightly more prominent isomer was labeled isomer 1. Isomer 1 adopted the *trans* conformation at P21, while isomer 2 adopted the *cis* conformation at P21. It is interesting to draw attention to the fact that P21 corresponds to the *cis*



**Figure 5.18.** Spectral data showing cross peaks associated with *cis/trans* conformation of prolines of acyclic permutant 6 (a) NOESY spectra (b) HMQC spectra. Assignments common to both isomers are written in black, while those of isomer 1 are written in red and those of isomer 2 are written in blue.

proline of the parent cyclic peptide, kalata-B1. In terms of 3D structure, the acyclic permutant 6 studied seems to be notably different from kalata-B1, but the isomeric forms themselves are similar.

The combination of chemical shift data, CSI values and NOE connectivities indicated that the isomers are overall similar in structure. This structure consists of a  $\beta$ -strand, which is centrally located over T10 - T13 and possibly several turns either side of this  $\beta$ -strand at E4 - C2, V7 - T5 on one side and near C18 - C16 for isomer 1 and C16 - P14 for isomer 2 on the other side. Structural differences between the isomers seem to prevail around the P21 region, where *cis/trans* isomerism is taking place, and also to some extent at the beginning of the sequence. The structures of the isomers have little in common with kalata-B1.

Kalata-B1 is a significantly more well defined structure and is featureless in the region of the  $\beta$ -strand of the isomers. The only common region between the isomers and kalata-B1 is over T5 - T10, where there are several similar NOE connectivities. Kalata-B1 has  $\beta$ -strand structure in regions not evident in the isomers. It is apparent that the isomers may have retained the structure of the connecting region of the parent over T5 - T10, but have lost the structure of the other regions, including that of the all important  $\beta$ -hairpin, which defines the class of inhibitor cystine knots.

The similarity of the isomers over T5 - T10 and specific NOE data, known to be indicative of the cysteine bonding pattern, were used to infer the disulphide bonds in acyclic permutant 6. It seems that the native cysteine bond, Cys(2)-Cys(5), has been retained and non-native cysteine bonds of Cys(1)-Cys(3) and Cys(4)-Cys(6) have been formed. This particular disulphide bonding pattern would be expected to be more unstable than that of the kalata-B1 and would produce a relatively more open structure that would allow *cis/trans* isomerisation to proceed. The fact that this is indeed the case, confirms that acyclic permutant 6 with two non-native disulphide bonds forms a less well defined 3D structure. This result indicates the importance of the cystine knot in maintaining a robust structure.

In summary, the fold of the fragment isomers studied in this chapter, are significantly different from the parent peptide, kalata-B1. The difference is likely to be due to an altered disulphide bonding pattern. Future studies should be directed towards the confirmation of this disulphide bonding pattern and elucidation of the 3D structure.

**CHAPTER SIX**

**STRUCTURE DETERMINATION**

**OF CIRCULIN A**

## 6.1 Introduction

This chapter describes the structure determination of the plant peptide, circulin A, which has been found to have anti-HIV activity (Gustafson *et al.*, 1994). The sequence of circulin A is given below in Figure 6.1. The numbering of residues used throughout this chapter will be that according to the alignment of Figure 1.1.10 (Chapter 1.1), with C1 of the cystine knot designated as residue 1.

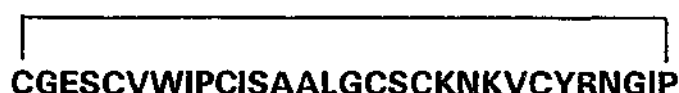


Figure 6.1. Sequence of circulin A (Gustafson *et al.*, 1994).

Circulin A shares around 50% sequence homology with kalata-B1. This is a much lower percentage than that shared with kalata-B1 by many of the other macrocyclic peptides. As a consequence, there is potential for significant structural dissimilarity due to the differences in sequence, spacing of the loops and also in the distribution of hydrophobic residues. These differences coupled with the fact that the mechanism of activity is to date unknown, renders the determination of the 3D structure of interest and importance. Through communication with the research group associated with discovery, purification and characterisation of the circulins (Gustafson *et al.*, 1994), circulin material was kindly provided to enable determination of their 3D structures. The determination is presented in the following pages, beginning with the experimental methodology applied for the data acquisition on circulin A.

## 6.2 Materials and Methods

An overview of methods used in this thesis is given in Chapter 2. In the following section, specific details relevant to the study of circulin A are recorded.



### 6.2.1 NMR Experiments

Circulin A material was prepared for NMR experiments by dissolving the purified material in either 10%:90% by volume  $^2\text{H}_2\text{O}$ : $\text{H}_2\text{O}$  or 99.99%  $^2\text{H}_2\text{O}$  to give solutions of 1mM concentration. These solutions were adjusted to pH 3.5 and placed in 5 mm Wilmad NMR tubes for NMR data acquisition.

To assist in resolving spectral ambiguities and to optimise the amount of information, data were collected under varying conditions, including different fields, temperatures and pulse sequence parameters. The NMR data were recorded at 285 and 300 K on Bruker ARX 500 and Bruker AVANCE DRX600 spectrometers, each of which is equipped with a shielded gradient unit. The temperature was maintained using a temperature controlled stream of air which was supplied by a Bruker BCU05 refrigeration unit and controlled by a B-VT 2000 control unit. Two dimensional NMR spectra were acquired with a 5 mm inverse broadband probe in the phase sensitive mode using time proportional phase incrementation for quadrature detection in the  $t_1$  dimension (Marion and Wüthrich *et al.*, 1983) over 6024 Hz. The following NMR data described in greater detail was used for the structural analysis.

TOCSY spectra (Brauschweiler and Ernst, 1983) were recorded using an MLEV-17 spin lock sequence (Bax and Davis, 1985) with a mixing period of 80 ms. NOESY spectra (Jeener *et al.*, 1979) were recorded with mixing times of 100, 200 and 250 ms. Solvent suppression for NOESY and TOCSY spectra was achieved using a modified WATERGATE sequence (Piotto *et al.*, 1992) in which two gradient pulses of 2 ms duration and 6 G cm $^{-1}$  strength were applied either side of a binomial 3-9-19 pulse. The spectra were acquired with 4096 complex data points in  $F_2$  and 400 or 512 increments in the  $F_1$  dimension with 32 or 48 scans per increment for the TOCSY spectra, and 512 or 600 increments in the  $F_1$  dimension with 64 or 96 scans per increment for the NOESY spectra.

DQF-COSY (Rance *et al.*, 1983) and E-COSY (Griesinger *et al.*, 1987) spectra were acquired with 4096 complex data points in  $F_2$  and 600

increments in the  $F_1$  dimension for the DQF-COSY spectra and 862 increments in the  $F_1$  dimension for the E-COSY spectra. The number of scans per increment was 40 for the DQF-COSY spectra and 60 for the E-COSY spectra. Solvent suppression for the COSY-type spectra was attained using selective low-power irradiation of the water resonance during a relaxation delay of 1.8 s.

Slowly exchanging NH protons were identified by acquiring a series of 1D and TOCSY spectra of the fully protonated peptide immediately following dissolution in  $^2\text{H}_2\text{O}$ . The  $^3J_{\text{NH-H}\alpha}$  coupling constants were measured from DQF-COSY spectra. The  $^3J_{\text{H}\alpha\text{-H}\beta}$  coupling constants were measured from E-COSY spectra for non-valine residues. The DQF-COSY and E-COSY data were strip transformed to 13 K in both  $F_1$  and  $F_2$  dimensions for coupling constant measurement. A deconvolution program which fits Lorentzian line shapes to 1D rows was used to improve the accuracy of  $^3J_{\text{NH-H}\alpha}$  coupling constants measured from DQF-COSY spectra as this procedure minimises error due to finite line widths of anti-phase peak components.

All spectral data were processed on a Silicon Graphics Indy workstation using XWINMR (Bruker) software. The  $t_1$  dimension of 2D data was zero-filled to 2048 real data points and  $90^\circ$  phase-shifted sine-bell window functions were applied prior to Fourier transformation. Polynomial baseline correction was used in regions either side of the water resonance to improve the spectral appearance of the 2D data. The 1D data were multiplied by an exponential function with line broadening of 0.3 prior to Fourier transformation. Chemical shifts were referenced to DDS set at 0.00 ppm. The processed data were then used to determine distance restraints for structure calculations, which were conducted as follows.

### 6.2.2 Structure Calculations

Distance restraints were determined primarily from the 250 ms NOESY spectrum recorded at 285K on the 500 MHz spectrometer. Peak volumes

were visually classified as strong, medium, weak or very weak and upper bound interproton distance restraints of 2.7, 3.5, 5.0 and 6.0 Å, respectively, were assigned (Williamson *et al.*, 1985; Clore *et al.*, 1986a). Pseudoatom corrections were applied to non-stereospecifically assigned methylene and methyl protons (Wüthrich *et al.*, 1983), so that for distances involving methylene protons, 1.0 Å was added to the upper limit, and for distances involving methyl protons, 1.5 Å was added to the upper limit.

Backbone dihedral restraints were inferred from  $^3J_{\text{NH-H}\alpha}$  coupling constants, with  $\phi$  restrained to  $-120 (\pm 30)^\circ$  for  $^3J_{\text{NH-H}\alpha}$  in the range of  $8.5 \pm 0.5$  Hz,  $-120 (\pm 15)^\circ$  for  $^3J_{\text{NH-H}\alpha}$  greater than 9.5 Hz and  $-65 (\pm 15)^\circ$  for  $^3J_{\text{NH-H}\alpha}$  less than 5.8 Hz (Pardi *et al.*, 1984; Wüthrich, 1986). Additional  $\phi$  restraints of  $100 (\pm 80)^\circ$  were applied to residues where the intraresidue  $\text{H}^\alpha$ - $\text{H}_\text{N}$  NOE was weaker than the sequential  $\text{H}^\alpha$ - $\text{H}_\text{N}$  NOE (Clubb *et al.*, 1994). Residues with intense intraresidue  $\text{H}^\alpha$ - $\text{H}_\text{N}$  NOEs and a  $^3J_{\text{NH-H}\alpha}$  of  $\sim 7$  Hz were restrained to  $\phi$  of  $50 (\pm 40)^\circ$  (Ludvigsen and Poulsen, 1992; Omecinsky *et al.*, 1996). Finally,  $\chi_1$  dihedral angle restraints and stereospecific assignment of methylene protons were derived from  $^3J_{\text{H}\alpha\text{-H}\beta}$  coupling constants and intraresidue NOE intensities (Zuiderweg *et al.*, 1985; Wagner *et al.*, 1987). Ranges of  $\pm 30^\circ$  were used for all  $\chi_1$  dihedral angle restraints.

Hydrogen bonds were also included in the final 3D calculations. Hydrogen-deuterium exchange experiments permitted the identification of the amide proton hydrogen-bond donors. The hydrogen-bond acceptors for the amide protons were ascertained through spatial examination of the initial structures generated. Hydrogen bonds (*i-j*) were then expressed as distance restraints of  $1.58 \leq d \leq 2.3$  Å for  $\text{HN-O}_j$  and  $1.58 \leq d \leq 3.2$  Å for  $\text{N-O}_j$ .

Three dimensional structures were generated using simulated annealing and energy minimisation protocols in the program X-PLOR Online 3.851 (Brünger, 1992). The calculations were carried out on a Silicon Graphics (4D/30) computer. Initially, an *ab initio* simulated annealing protocol (Nilges *et al.*, 1988) was used to generate a set of 50 structures, starting from template structures with randomised  $\phi$  and  $\psi$  angles and extended side-chains. The

backbone was cyclised throughout the calculations and all peptide bonds were defined as *trans*. At this stage, the disulphide bonds were included as pseudo-NOE restraints and the simulated annealing protocol consisted of 20 ps of high temperature molecular dynamics at 1000 K with a low weighting on the repel force constant and the NOE distance restraints.

This was followed by a further 10 ps of high temperature dynamics with an increased force constant on the NOE distance restraints. The disulphide bonds were formally bonded and the dihedral force constant increased prior to cooling the system to 0 K and increasing the repel force constant over 15 ps of dynamics. The NOE restraints were checked for violations and ambiguous cross peaks were resolved on the basis of interproton distances in the family of structures. A further 50 structures were calculated and these structures were refined using the conjugate gradient Powell algorithm with 1000 cycles of energy minimisation (Clore *et al.*, 1986b) and a refined force field based on the program CHARMM (Brooks *et al.*, 1983).

Structures were examined using PROMOTIF (Hutchinson and Thornton, 1996), checked using PROCHECK (Laskowski *et al.*, 1993) and displayed using INSIGHT (Biosym Technologies, San Diego, CA). The resulting set of structures, together with the analysis of the NMR data, allowed for a detailed insight into the 3D topology of circulin A.

## 6.3 Results and Discussion

### 6.3.1 Proton Resonance Assignments

By the application of standard resonance-assignment strategies to the 2D NMR data (Wüthrich, 1986; King and Mackay, 1996), the  $^1\text{H}$  chemical shifts for all backbone and side chain protons were identified. The full set of  $^1\text{H}$  chemical shifts is given in Table 6.1. The  $^1\text{H}$  resonances in the TOCSY fingerprint region are well dispersed, allowing classification of the amino acid

**Table 6.1.** Proton chemical shifts (ppm) for Circulin A at 285 K. Shifts are measured relative to DDS set at 0 ppm.

	RESIDUE	HN	$\alpha$ H	$\beta$ H	OTHER	
1.	CYS	8.51	4.57	3.25, 3.06		
2.	GLY	8.96	4.10, 3.70			
3.	GLU	7.63	4.65	2.04, 1.74	$\gamma$ CH <sub>2</sub> 2.46, 2.37	
4.	SER	8.79	4.72	4.04, 3.94		
5.	CYS	8.08	5.35	3.29, 3.18		
6.	VAL	7.93	3.44	1.65	$\gamma$ CH <sub>3</sub> 0.93, 0.15	
7.	TRP	8.21	4.87	3.50, 3.15		
8.	ILE	7.52	4.77	1.96	$\gamma$ CH <sub>2</sub> 1.02 $\gamma$ CH <sub>3</sub> 0.91	$\delta$ CH <sub>3</sub> 0.91
9.	PRO		4.48	2.43, 1.87	$\gamma$ CH <sub>2</sub> 2.00	$\delta$ CH <sub>2</sub> 3.80, 3.64
10.	CYS	8.50	4.31	2.98, 2.72		
11.	ILE	8.13	4.13	1.94	$\gamma$ CH <sub>2</sub> 1.27 $\gamma$ CH <sub>3</sub> 0.95	$\delta$ CH <sub>3</sub> 0.95
12.	SER	9.07	4.41	4.41, 4.07		
13.	ALA	7.86	4.50			
14.	ALA	8.44	4.34	1.41		
15.	LEU	7.50	4.34	1.72	$\gamma$ H 1.72	$\delta$ CH <sub>3</sub> 0.98, 0.91
16.	GLY	8.28	4.33, 3.77			
17.	CYS	7.67	4.80	3.28, 2.62		
18.	SER	9.53	4.79	3.89		
19.	CYS	8.94	4.73	3.14		
20.	LYS	9.60	4.62	1.87	$\gamma$ CH <sub>2</sub> 1.29 $\delta$ CH <sub>2</sub> 1.51	$\epsilon$ CH <sub>2</sub> 2.83
21.	ASN	9.39	4.54	3.06, 2.76		
22.	LYS	8.25	3.52	2.27, 1.91	$\gamma$ CH <sub>2</sub> 1.28 $\delta$ CH <sub>2</sub> 1.63, 1.55	$\epsilon$ CH <sub>2</sub> 2.92
23.	VAL	7.97	4.24	2.05	$\gamma$ CH <sub>3</sub> 0.98, 0.56	
24.	CYS	7.69	5.16	3.11, 2.68		
25.	TYR	9.73	5.14	2.84, 2.69		
26.	ARG	9.34	4.79	1.87, 1.76	$\gamma$ CH <sub>3</sub> 1.66, 1.47	$\delta$ CH <sub>2</sub> 3.20
27.	ASN	9.56	4.39	3.12, 2.86		
28.	GLY	8.59	4.18, 3.50			
29.	ILE	8.07	4.83	2.20	$\gamma$ CH <sub>2</sub> 1.33 $\gamma$ CH <sub>3</sub> 1.05	$\delta$ CH <sub>3</sub> 0.92
30.	PRO		4.20	2.22, 1.55	$\gamma$ CH <sub>2</sub> 2.02, 1.90	$\delta$ CH <sub>2</sub> 4.20, 3.87

spin systems. These are shown in Figure 6.2. Specific assignment of the amino acid spin system was enabled through the analysis of the fingerprint region of the NOESY spectrum.

The complete cycle of  $\alpha\text{H}$ -NH sequential connectivities was established in the fingerprint region of NOESY spectrum. Only at the proline residues was the cycle broken. However, NOE cross peaks between  $\alpha\text{H}$  of proline residues and the subsequent amide protons permitted completion of the cycle. The completed cycle is illustrated in Figure 6.3. This result provides unequivocal evidence for the cyclic nature of circulin A, confirming previous findings from mass spectral and enzymatic cleavage studies (Gustafson *et al.*, 1994; Derua *et al.*, 1996).

The NOESY spectra also gave evidence for the isomerisation of the amide bonds of the two proline residues of circulin A. The presence of NOE cross peaks between  $\alpha\text{H}_{i-1}$  -  $\delta\text{H}_i$  protons, coupled with the absence of NOE cross peaks between  $\alpha\text{H}_{i-1}$  -  $\alpha\text{H}_i$  protons for the proline residues, ascertained that the amide bonds of the proline residues adopt the more common *trans* conformation. Having assigned the NMR data, secondary structure features were then identified.

### 6.3.2 Secondary Structure

Elements of secondary structure can be deduced from CSI values, NOE connectivities,  $^3J_{\text{NH-H}\alpha}$  coupling constants and slow exchanging NH protons, or from a combination thereof. Initially, CSI values were calculated for the backbone  $\alpha$  protons (Wishart *et al.*, 1992). The results are graphically displayed in Figure 6.4. As an initial observation, many of the residues have  $\alpha$  proton chemical shifts, which differ from their random chemical coil values (Merutka *et al.*, 1995) by more than 0.1 ppm. As a consequence, most of the residues have CSI values of " $\pm 1$ ". This strongly suggests that circulin A is a highly structured molecule.

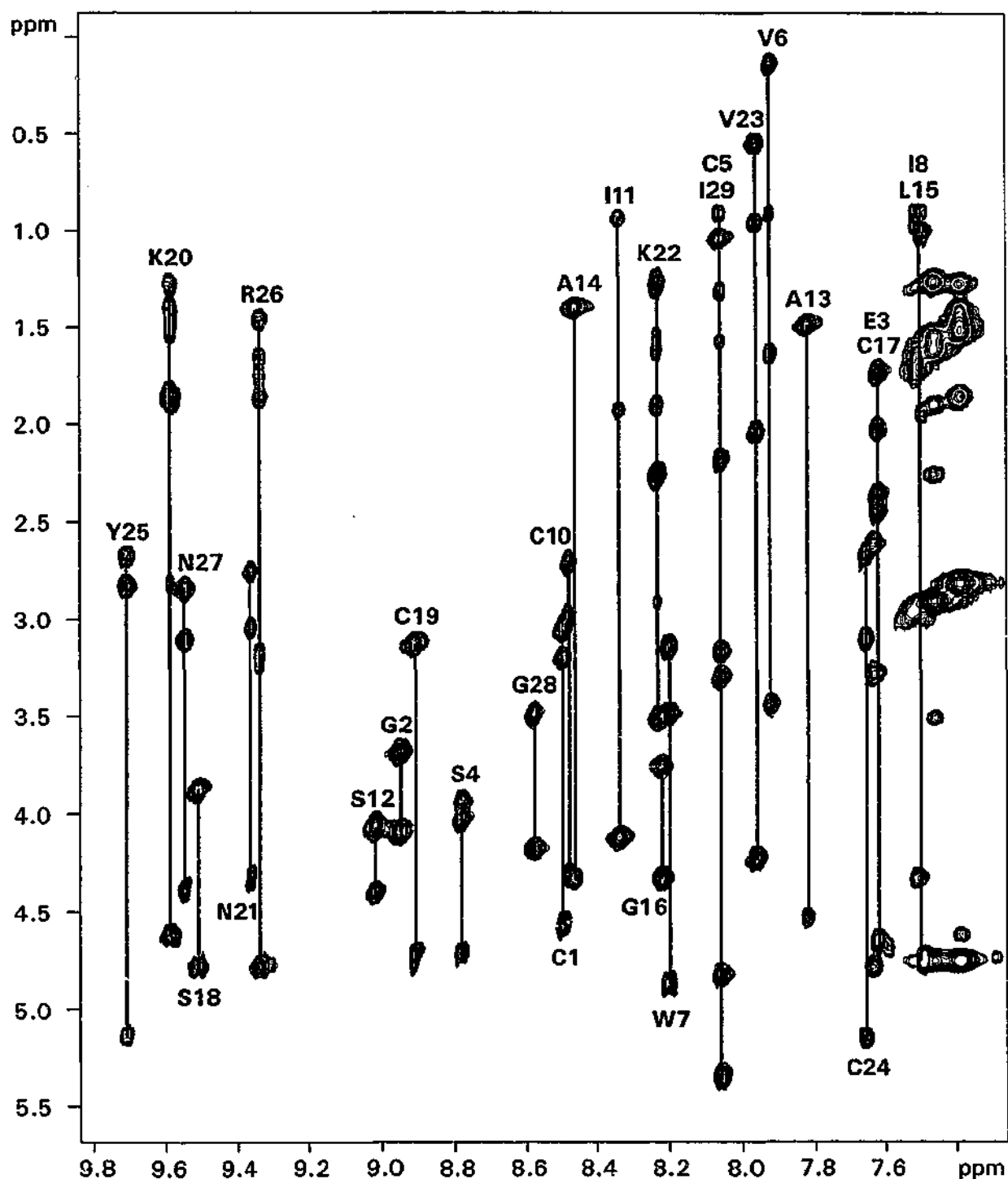
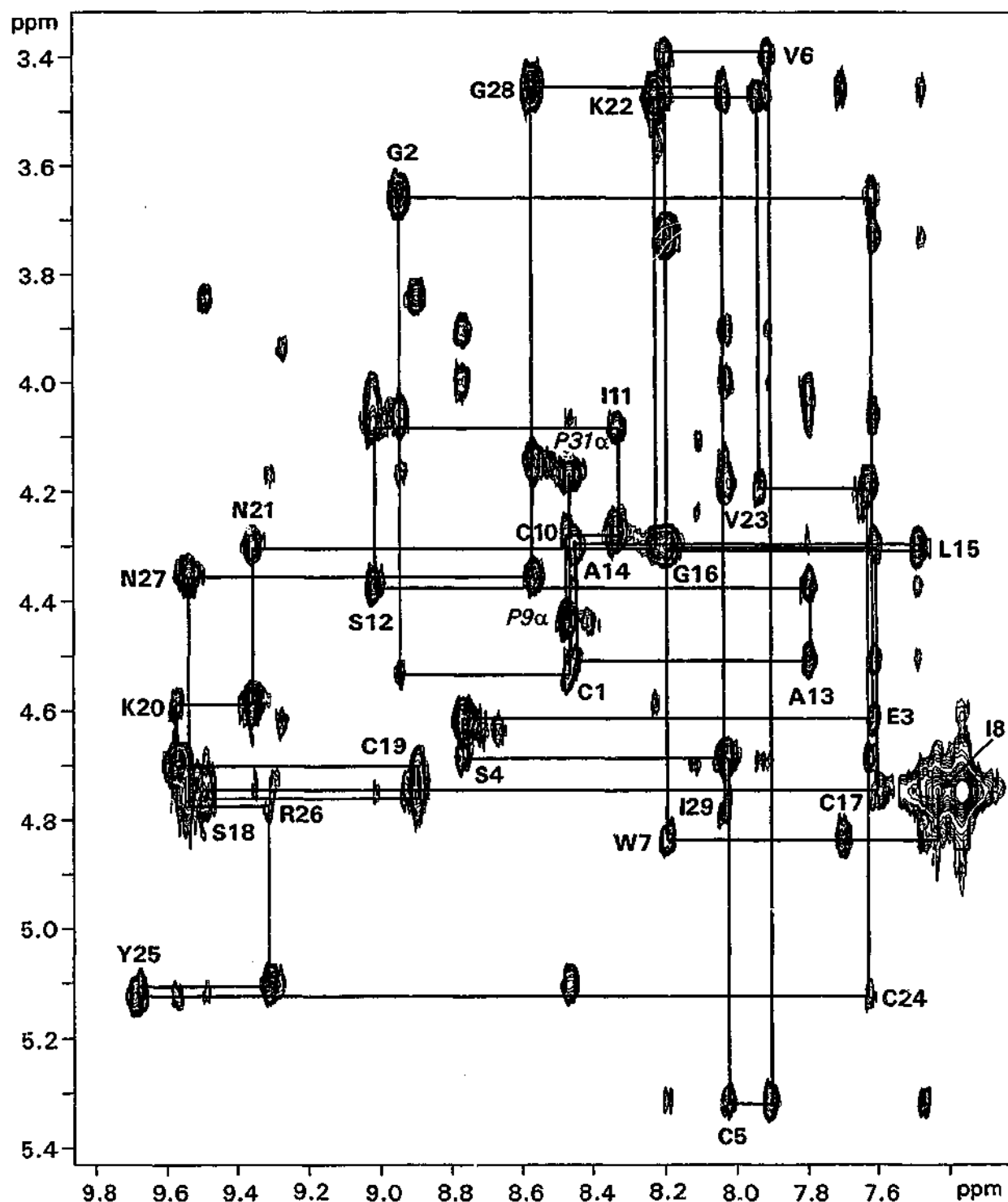
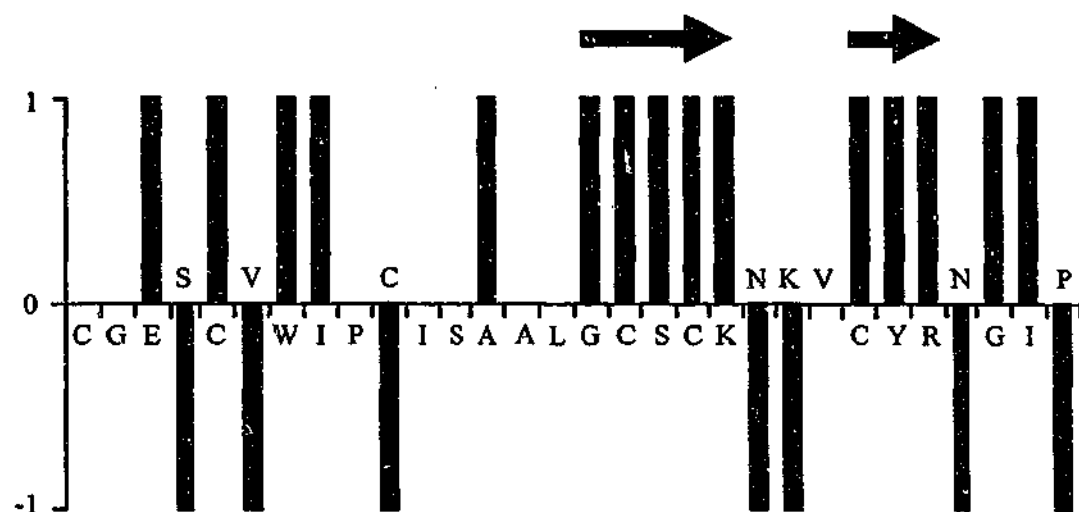


Figure 6.2. Fingerprint region of the 600 MHz TOCSY spectrum showing assignments of side chain spin systems linked with their corresponding amide protons for circulin A at 300 K in 90% H<sub>2</sub>O / 10% <sup>2</sup>H<sub>2</sub>O.

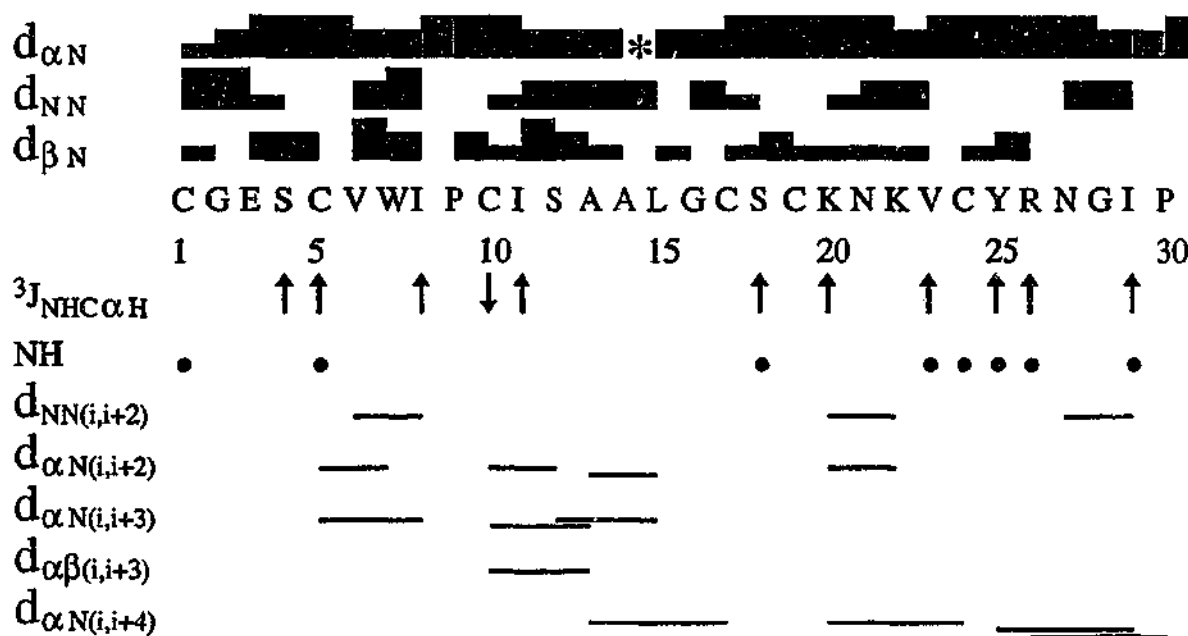


**Figure 6.3.** Fingerprint region of the 600 MHz NOESY spectrum for circulin A at 300 K in 90% H<sub>2</sub>O / 10% <sup>2</sup>H<sub>2</sub>O. The sequential connectivity pattern is shown confirming the cyclic nature of the polypeptide backbone.





**Figure 6.4.** CSI values derived from alpha proton chemical shifts of circulin A. The prediction of  $\beta$ -strand character is indicated by solid arrows.



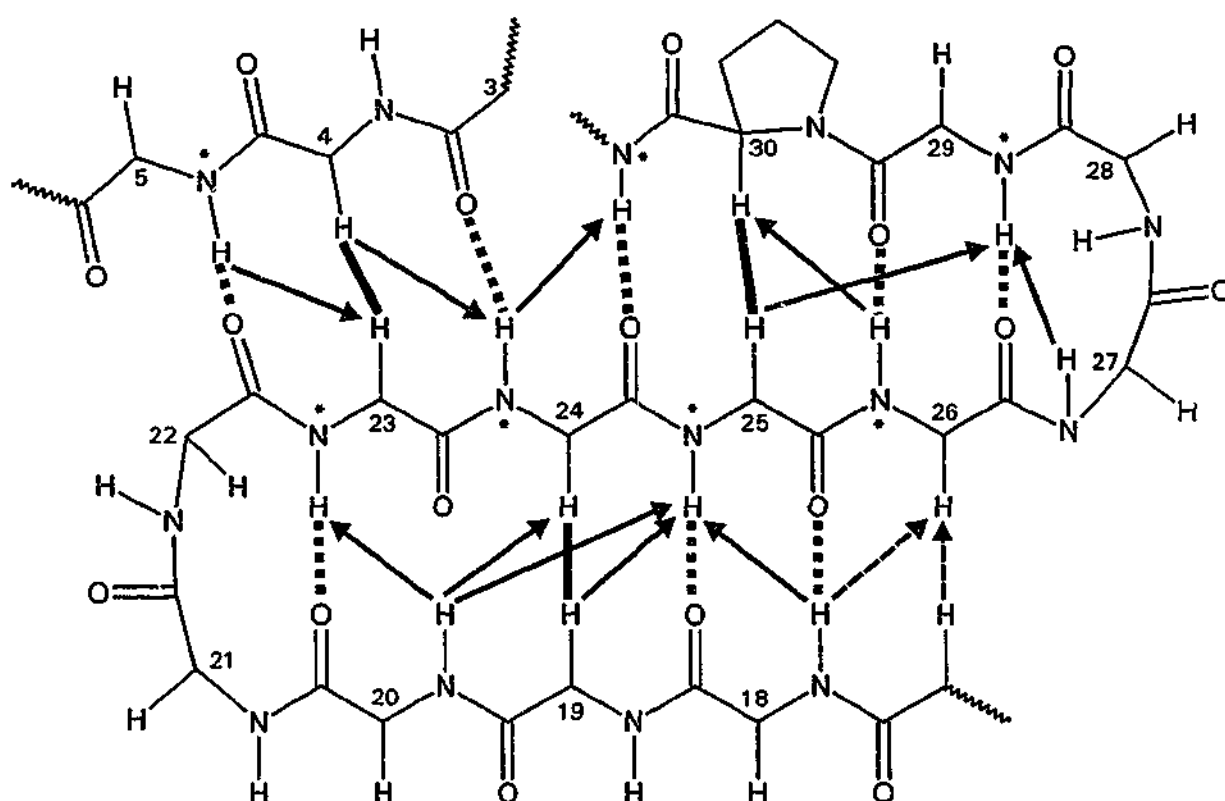
**Figure 6.5.** Summary of the sequential and medium-range NOE connectivities,  $^3J_{\text{NHCOH}}$  coupling constants and slowly exchanging NH protons observed for circulin A. Filled bars indicate sequential connectivities observed in a 250 ms NOESY spectrum at 12°C, pH 3.5. Shaded bars correspond to sequential  $\text{H}^\alpha\text{-H}^{\alpha+1}$  and  $\text{NH-H}^{\alpha+1}$  connectivities for proline residues. The height of the bar indicates the strength of the NOE. Overlapping NOEs are indicated by an asterisk; ( $\uparrow$ ) indicate  $^3J_{\text{NHCOH}}$  coupling constants  $\geq 8.5$  Hz. Slowly exchanging backbone NH protons observed in a TOCSY spectrum recorded in  $^2\text{H}_2\text{O}$  are indicated by filled circles.

Examination in finer detail reveals that there are two regions of consecutive positive CSI values consistent with the presence of extended structure. These include residues G16 to K20 and C24 to R26. The remainder of the molecule displays an irregular pattern of positive and negative CSI values, suggesting the occurrence of turns. Further evidence is obtained from analysis of short and medium range NOE data, which is collated in Figure 6.5 together with  $^3J_{\text{NH-H}\alpha}$  coupling constant and slow exchange data.

It may be deduced from this diagram that  $d_{\text{NN}}(i, i+2)$  and  $d_{\alpha\text{N}}(i, i+2)$  NOE data also indicates the presence of turns (Wüthrich, 1986) throughout the molecule. Furthermore, although there is no contiguous stretch of negative CSI values that is characteristic of  $\alpha$ -helix structure, the  $d_{\alpha\text{N}}(i, i+4)$  NOE between residues S12 and G16 suggests there may indeed be a small stretch of  $\alpha$ -helix (Wüthrich, 1986) in this region. Such NOEs are also evident in the region between K20 and P30. However, they are associated with large  $^3J_{\text{NH-H}\alpha}$  coupling constants and positive CSI values, rendering it unlikely that they reflect helices in this region.

Lastly slowly exchanging amide protons marked in Figure 6.5, give a little more information on the general nature of the structure of circulin A. Eight of the 28 amide protons are present more than one hour after dissolution in  $^2\text{H}_2\text{O}$  and five of these are still present at least 20 hours later. This indicates that these residues are protected from the solvent environment and that well-defined secondary structure features in circulin A. Combining all the data from the sequential and long range NOE cross peaks associated with the backbone, slow exchange and coupling patterns, an interpretation of major secondary structure can be derived. This is schematically illustrated in Figure 6.6.

The preeminent feature in circulin A is a small triple stranded  $\beta$ -sheet, which is comprised of residues S18 to P30 and E3 to C5, and contains a  $\beta$ -hairpin. The turn of the  $\beta$ -hairpin seems to be centered around residues N21 to K22 based on the negative CSI values and the presence of  $d_{\text{NN}}(i, i+2)$  and  $d_{\alpha\text{N}}(i, i+2)$  NOE cross peaks, which are indicative of turns (Wüthrich, 1986).



**Figure 6.6.** Schematic diagram of the  $\beta$ -sheet region of circulin A showing the interstrand NOEs (arrows), potential hydrogen bonds (broken lines) and slowly exchanging amide NH bonds (\*). For clarity, sequential NOEs are not shown. The broken arrows corresponds to NOEs which are ambiguous due to overlap.

The extended regions of the  $\beta$ -hairpin involve residues G16 to K20 and C24 to R26 determined from positive CSI values, slow exchange amide protons and large coupling constants typical of extended conformations. The strand adjacent to the hairpin is substantially distorted and the NOE data are suggestive of a bulge near residues C1 and G2. It is possible that this may be due to the disulphide bond connected to C1. Residues V23 to R26 form the central strand of the triple stranded  $\beta$ -sheet. All four of these residues have slow exchange amide protons as expected for a central strand, and large coupling constants are observed for three out the four residues. The amide protons, which are predicted to form hydrogen bonds across the triple stranded  $\beta$ -sheet, all display slow exchange with  $^2\text{H}_2\text{O}$  solvent with the exception of K20. This is further discussed in the results of the structure calculations.

### **6.3.3 Three Dimensional Structure**

#### **6.3.3.1 Quality of Structures**

A set of 50 structures was generated using 374 NOE distance restraints, 17 backbone dihedral restraints, 9  $\chi_1$  angle restraints, 8 hydrogen bond definitions and stereospecific assignments where applicable. The table of NOE restraints is given in Appendix IV. The backbone dihedral restraints,  $\chi_1$  angle restraints and hydrogen bond definitions applied are listed in Appendix V. The twelve best structures were selected on the basis of lowest total energy, lowest NOE energy and minimum RMSD value. The mean pairwise RMSDs over the twelve best structures were 1.01 Å for the backbone atoms and 1.67 Å for all non-hydrogen atoms.

Geometric and energetic statistics that were calculated for the twelve best structures are shown in Table 6.2. In general, the structures show no significant deviation from ideal covalent geometry and fit the experimental restraints with minimal violations. There were no NOE violations greater than

<b>Mean Pairwise RMS deviations (Å)<sup>a</sup></b>	
Backbone	1.01 ± 0.29
Heavy atom	1.67 ± 0.25
<b>Mean RMS deviations from experimental restraints</b>	
Interproton distances (Å)	0.02 ± 0.001
Dihedral angles (deg.)	0.18 ± 0.04
<b>Mean RMS deviations from idealized covalent geometry<sup>b</sup></b>	
Bonds (Å)	0.008 ± 0.0002
Angles (deg.)	2.15 ± 0.03
Impropers (deg.)	0.16 ± 0.01
<b>Mean Energies (kcal mol<sup>-1</sup>)</b>	
$E_{\text{NOE}}^c$	6.63 ± 0.42
$E_{\text{cdih}}^c$	0.05 ± 0.02
$E_{\text{LJ}}^d$	-116.2 ± 5.8
$E_{\text{bond}}$	4.89 ± 0.28
$E_{\text{improper}}$	0.50 ± 0.08
$E_{\text{angle}}$	51.7 ± 1.39
$E_{\text{total}}$	-43.7 ± 2.60
<b>Ramachandran plot statistics<sup>e</sup></b>	
Residues in most favoured regions (%)	64
Residues in additional allowed regions (%)	33
Residues in generously allowed regions (%)	3
Residues in disallowed regions (%)	0
The values in the Table are the mean ± standard deviation.	
<sup>a</sup> RMS deviation measured for the whole molecule.	
<sup>b</sup> Idealised geometry as defined by CHARMM force field and as implemented within XPLOD.	
<sup>c</sup> Force constants for the calculation of square-well potentials for the NOE and dihedral angle restraints were 50 kcal mol <sup>-1</sup> Å <sup>-1</sup> and 200 kcal mol <sup>-1</sup> rad <sup>-2</sup> , respectively.	
<sup>d</sup> The Lennard-Jones van der Waals energy was calculated with the CHARMM empirical energy function.	
<sup>e</sup> Ramachandran statistics were calculated using PROCHECK (Laskowski <i>et al.</i> , 1996).	

**Table 6.2.** Geometric and energetic statistics for the twelve best final structures of circulin A.

0.2 Å and no dihedral violations greater than 2°. Angular order parameters were calculated for these structures to determine regions of order/disorder (Hyberts *et al.*, 1992). These are given in Table 6.3. The twelve structures were then examined in greater detail and compared with the 3D structure of kalata-B1.

#### 6.3.3.2 Three Dimensional Arrangement

The superimposition of the twelve best structures is displayed in Figure 6.7. It is apparent that the structure of circulin A is well defined over most regions. There is some disorder exhibited over residues 11 to 17 and around residue 28. This is further indicated geometrically by the angular order parameters, which approach unity for the majority of residues, except residues 13-17 and 27-29. However, superimposition over the cystine knot region, consisting of residues 2-12 and 18-27, is significantly improved with RMSD values of 0.32 Å for the backbone atoms and 1.45 Å for all heavy atoms of the knot. This finding is in line with the general observation that the cystine knot feature is conserved across the family of macrocyclic peptides.

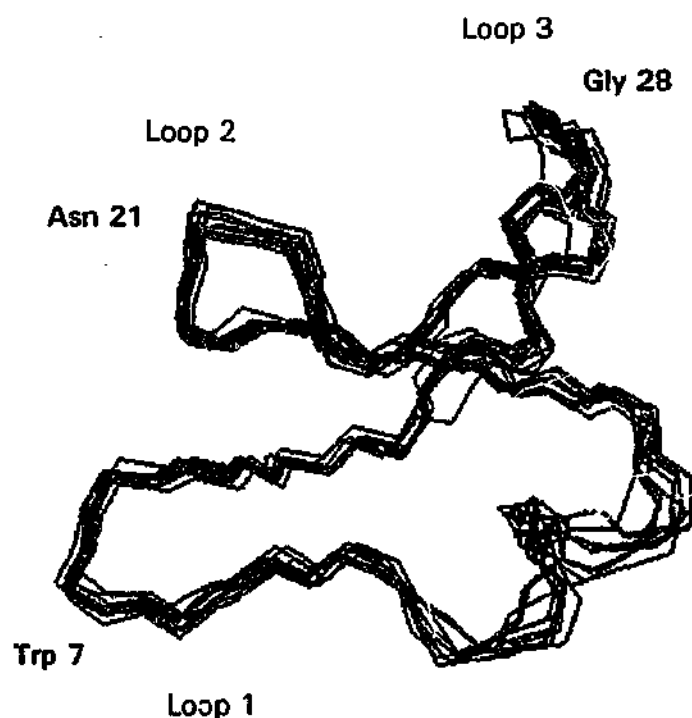


Figure 6.7. Superimposition of the twelve best NMR-derived structures over the backbone (N, C $^{\alpha}$ , C') cysteine residues.

Table 6.3. Angular order parameters for circulin A.

Residue	phi	psi	chi 1
1	1.00	0.99	0.98
2	0.98	0.94	0
3	0.91	0	0.93
4	0	1.00	1.00
5	1.00	0.99	1.00
6	0.98	1.00	1.00
7	0.99	0.99	0.98
8	1.00	0.98	0.53
9	1.00	0.99	1.00
10	1.00	0.97	0.76
11	0.98	0.99	0.99
12	1.00	0.95	0.62
13	0.70	0.93	0
14	0.71	0.89	0
15	0.69	0.75	0.77
16	0.82	0.85	0
17	0.99	0.92	0.83
18	0.88	0.980	0.56
19	0.98	0.98	0.99
20	0.98	0.99	0.71
21	1.00	0.99	0.41
22	0.99	0.99	0.87
23	0.99	1.00	0.99
24	1.00	0.98	0.95
25	0.99	1.00	1.00
26	1.00	1.00	0.90
27	1.00	1.00	0.63
28	1.00	1.00	0
29	1.00	1.00	1.00
30	1.00	1.00	0.99

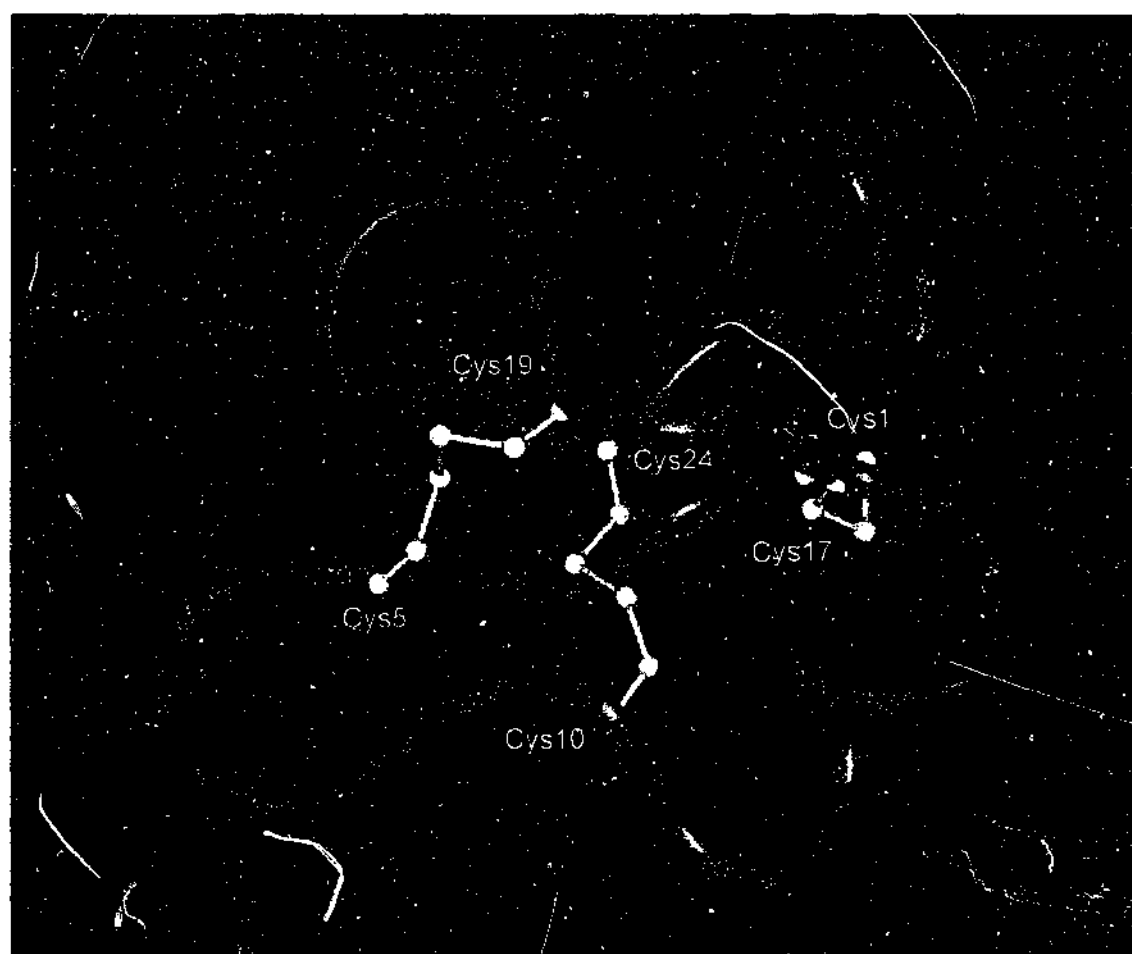
Overall, the 3D spatial arrangement of circulin A, forms a highly compact globular structure in which three loops in the peptide backbone can be identified, as shown in Figure 6.7. Two of the loops, namely 1 and 2 are folded close to one another and the turns at the ends of these loops are located one above the other. Loop 2 corresponds to the  $\beta$ -hairpin. The closely packed arrangement of circulin A is braced by the disulphide bonds across diagonally opposed strands of the loops.

The configuration of the disulphide bonds and the  $\beta$ -strand is compliant with that of the cystine knot motif (McDonald and Hendrickson, 1993; Saether *et al.*, 1995; Isaacs, 1995) which uniquely defines the cystine knot family of peptides. More specifically, the embedded eight amino acid loop formed by the backbone of segments Cys1 to Cys5 and Cys17 to Cys19, together with their disulphide bonds Cys1-Cys17 and Cys5-Cys19 is threaded by the third disulphide bond Cys10-Cys24. Together with the structure of kalata-B1, the embedded loop in circulin A is the smallest known loop through which a disulphide bond may pass. The structure of circulin A is illustrated in Figure 6.8 and can be further discussed in terms of formal classification of local secondary structural elements.

#### 6.3.3.3 Analysis of Structure

The twelve best structures of circulin A were examined using the program PROMOTIF (Hutchinson and Thornton, 1996) which classifies structural elements according to specific definitions. The results of the analysis confirmed that the 3D structure of circulin A consists of predominantly turns. This is not surprising given the cyclic backbone, high concentration of cysteine residues and small, compact, defined structure. All the turns identified are classified as  $\beta$ -turns, based on the distance between the  $C^\alpha$  of residue  $i$  and  $C^\alpha$  of residue  $i+3$  being less than 7 Å. In the majority of the twelve structures,  $\beta$ -turns were identified between residues 5-8, 6-9, 10-13, 19-22, 20-23, 25-28, 26-29. The turn involving the residues 6-9 is





**Figure 6.8.** Lowest energy structure of circulin A. The backbone is shown by a blue ribbon and the disulphide bonds are shown in yellow in ball and stick form.

classified as type VIII and the 26-29 turn which is stabilized by hydrogen bonds is classified as a type I'  $\beta$ -turn. The remaining turns do not readily fit standard turn types and are classified as type IV  $\beta$ -turns. Further information on structural features was also obtained from the analysis using PROMOTIF.

In all the calculated structures, a  $\beta$ -hairpin is present from residues 18-25 comprising of  $\beta$ -strands over residues 18-20 and 23-25. This is in agreement with the predicted secondary structure. In several structures, the amide of residue 20 is involved in a hydrogen bond across the  $\beta$ -hairpin. However, the amide of residue 20 does not exhibit slow exchange, suggesting that this is not likely to be a strong hydrogen bond. From the qualitative analysis of the NOE and slow exchange data, this  $\beta$ -hairpin forms part of a  $\beta$ -sheet, which is not formally identified by the program PROMTIF. This may be explained by the strand involving residues 30-5 containing a  $\beta$ -bulge and being at a slight angle to the other two strands of the  $\beta$ -sheet. This is not an uncommon occurrence for molecules of the size of circulin A and high cysteine content to have distorted elements of secondary structure. The disulphide bonds Cys1-Cys 17 and Cys5-Cys19 link the regions involved in the  $\beta$ -sheet and are likely to account for its distortion and also perhaps the strength of the hydrogen bond involving residue 20.

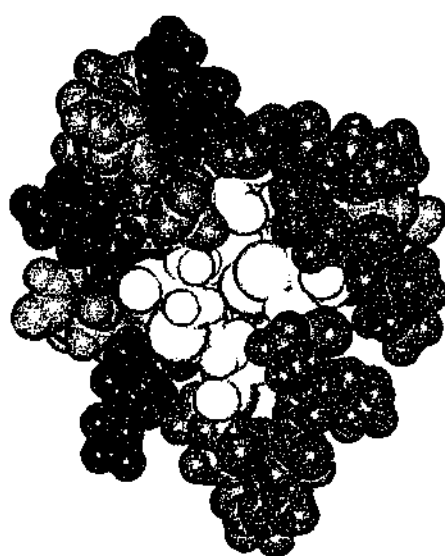
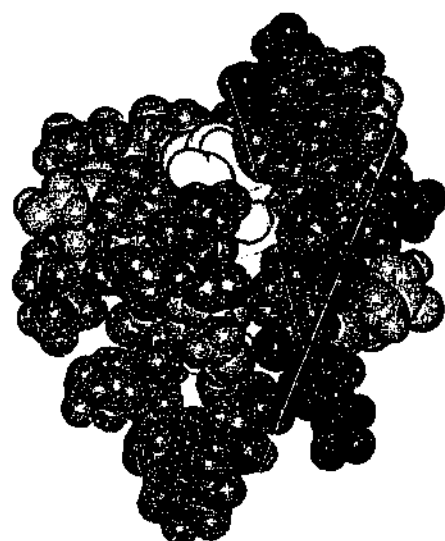
In addition to the  $\beta$ -hairpin, there is one more interesting feature of secondary structure identified by the program PROMOTIF. A small region of  $\alpha$ -helix over residues 11-16 is evident in eleven of the twelve structures. This  $\alpha$ -helix is classified as a type 3<sub>10</sub> helix in nine of the structures and  $\alpha$ -helical in two of the structures. In the structure where the  $\alpha$ -helix is not defined, a series of three type IV  $\beta$ -turns is classified. This latter observation is also found in other structures not included in the set of final twelve structures. In general, the region of the  $\alpha$ -helix is relatively disordered and this may reflect mobility of this helical segment. Finally, PROMOTIF was used to examine the configuration of the disulphide bonds.

Of the three disulphide bonds stabilizing circulin A, two have well defined conformations across the family of structures. The torsion angles of

the disulphide bond Cys1-Cys17 (*i.e.* I-IV) are representative of a left-handed spiral for the majority of the twelve structures. This configuration is the most common one for disulphide bonds based on an analysis of protein structures deposited in the Brookhaven Protein Data Bank (Thornton, 1981; Srinivasan *et al.*, 1990). Similarly well defined, the torsion angles of the disulphide bonds Cys5-Cys19 (*i.e.* II-V) are characteristic of a right-handed spiral for all the twelve structures. There is much greater diversity for the disulphide structures in right-handed conformations than those of left-handed conformations (Thornton, 1981). The conformation of the third disulphide bond, Cys10-Cys24 (*i.e.* III-VI) varies amongst the individual structures within the calculated family, having a left-handed conformation in about half the structures and a right-handed conformation in the other half. In summary, the cystine knotted framework in circulin A does not appear to require an unusual conformation for the disulphide groups. Lastly, the NMR-derived structure for circulin A was examined in terms of distribution of residue types in 3D space.

#### 6.3.3.4 Charge and Hydrophobic Distribution

The space-filling model for circulin A was generated using the program INSIGHT (Biosym Technologies, San Diego, CA). To study the distribution of residues in 3D space, residue types were colour-coded. Hydrophobic residues were coloured green, negatively charged residues red, positively charged residues dark blue, polar residues light blue and cysteine residues yellow. The colour-coded space-filling model is illustrated in Figure 6.9 along with 90° and 180° views rotated about the vertical axis. The space-filling model shows the impact of the braced loop on the 3D shape. Circulin A is significantly distorted from a spherical globular shape and may be described as an oblate spheroid. The three views of Figure 6.9 highlight the flattened shape which has a cross-sectional diameter of approximately 22 Å viewed from the front, and 16 Å viewed from the side. The distribution of residue types on the surface is interesting.



**Figure 6.9.** Space-filling models of circulin A showing the distribution of hydrophobic (green), negatively charged (red), positively charged (dark blue), polar (light blue) and cysteine (yellow) residues. The middle and lowest views are the 90° and 180° rotations respectively of the upper view about the vertical axis.

The different types of residues are clustered. One face is almost entirely composed of hydrophobic residues, while on the opposite face the cystine core is exposed to solvent. The few positive charges are located along one peripheral edge, rather than on the flatter faces. The functional significance of this distribution is currently not known. However, the fact that many hydrophobic residues are exposed on the surface of the molecule suggests that the formation of a hydrophobic core is a less important driving force in the 3D folding of circulin A than the structural stabilisation brought about *via* intertwining of the three disulphide bonds to form the cystine knot motif. This motif assists formation of the central scaffold of the molecule about which three loops are built.

One of the loops is rich in hydrophobic residues which essentially shield much of the disulphide core from solvent on this face of the molecule. However, the location of the residues on the lower loop is such that one face of the disulphide core is nearly completely solvent exposed. Finally, the views of Figure 6.9 show the protuberance of the single tryptophan residue from the main core of the protein. The solvent exposed position of this tryptophan is another interesting feature, which is also observed in kalata-B1, as discussed in the following section.

#### 6.4 Comparison of Circulin A and Kalata-B1

A comparison of structures of circulin A and kalata-B1 is given in Figure 6.10 in which the peptides are superimposed over the cysteine backbone residues. An RMSD value of 1.02 Å is obtained for this superimposition. A significantly better RMSD value of 0.59 Å is obtained when the molecules are superimposed over the structural motif of the inhibitor cystine knot (Pallaghy *et al.*, 1994). This motif is defined as  $XXC, XCX, XXCX$ , where X is any residue except cysteine and the C residues are CysII, CysV and CysVI, respectively.

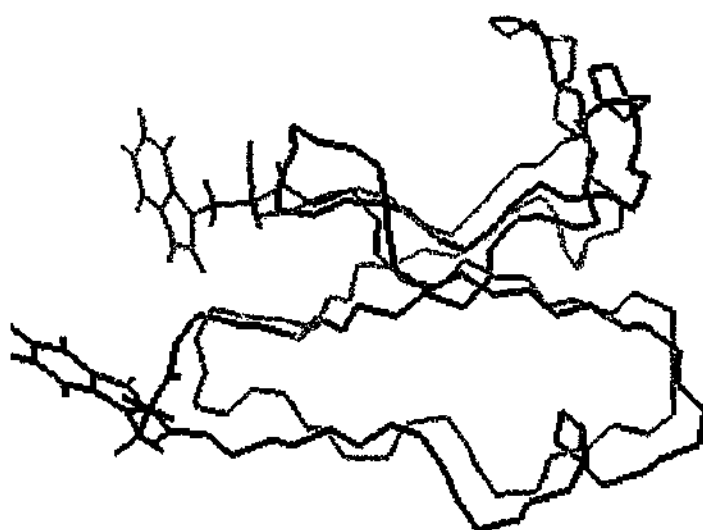


Figure 6.10. Superimposition of circulin A (black) and kalata-B1 (grey) over the cysteine backbone (N, C $\alpha$ , C') residues. The positions of the tryptophan side-chains are displayed for each molecule.

In addition to the similarities in disulphide topologies, kalata-B1 and circulin A have very similar global folds. Both peptides contain distorted triple-stranded  $\beta$ -sheets and numerous  $\beta$ -turns. The hydrogen bonds in circulin A predicted from the slow exchange data are equivalent to those seen in kalata-B1, with the exception of residue 20. Several turns link the regions involved in the distorted triple-stranded  $\beta$ -sheet.

In kalata-B1, the two loops equivalent to loops 1 and 2 in circulin A characterized by  $\beta$ -turns at their termini are parallel to one another. In circulin A, these loops are in a slightly different orientation than that observed in kalata-B1, as the turn at the end of loop 1 bends away from the turn in loop 2. The residues in the  $\beta$ -turns within these loops are not conserved between circulin A and kalata-B1. The central residues for loop 2 in kalata-B1 contain the sequence Trp-Pro, in which the proline appears to be in a *cis* conformation. Circulin A does not contain this proline residue. Interestingly, circulin A does contain a single tryptophan residue, but it occurs in a different turn, being in loop 1 rather than in loop 2 for kalata-B1. It is remarkable that despite being in a completely different position in the sequence, the tryptophan residues are in similar spatial locations in the two peptides. The

side-chains of the tryptophan residues are highly solvent exposed in both peptides, as illustrated in Figure 6.10.

The sequence alignment of kalata-B1 and circulin A (Figure 1.1.10 - chapter 1.1) reveals two extra residues in circulin A between CysIII and CysIV. In addition, an extra residue is present in kalata-B1 between CysVI and CysI. The two extra residues in circulin A between CysIII and CysIV do not protrude further from the cysteine core of the peptide, but are involved in a new element of secondary structure, namely the helical segment. It appears that this region of the structure is too short in kalata-B1 to form a helix, but in circulin A, it is sufficiently long to contain elements of helical structure. It might be anticipated that further extension to this loop might lead to a better-formed helix.

Overall, there is a high degree of similarity between the structures of circulin A and kalata-B1. Consequently, it could be expected that there is likely to be some functional similarity. Comparison of the structure of circulin A with other well-studied peptides could give some insight into their function in nature.

## 6.5 Discussion of Biological Activity

It may be speculated that these cyclic peptides may form part of plant defense mechanisms, since they share general structural similarities to a range of known plant defense molecules. For example, there are several members of the squash family of trypsin inhibitors that are of similar size (<35 amino acid residues) and contain three disulphide bonds. These include CMTI (Holak *et al.*, 1989a,b) and EETI (Favel *et al.*, 1989; Nielsen *et al.*, 1994a). While none of these molecules have a cyclic amide backbone, their overall folds are quite similar to that of circulin A and kalata-B1. Breaking the circulin A ring in loop 3 of the structure would lead to an open chain analogue with similar molecular topology.

Other recently discovered cystine-rich defense molecules in plants include a novel series of proteinase inhibitors produced in the stigmas of *Nicotiana glauca*. It has been proposed that these inhibitors have a function related to the protection of reproductive tissue against potential pathogens (Atkinson *et al.*, 1993). It is interesting that these protease inhibitors have a flattened spherical 3D shape (Nielsen *et al.* 1994b, 1995), not unlike that seen for circulin A, although these molecules are significantly larger (53 residues).

Moreover, the structure of circulin A appears to be very similar to that of the AVR9 elicitor peptide (Van den Ackerveken *et al.*, 1993) whose 3D structure was reported recently (Vervoort *et al.*, 1997). This peptide has been implicated in the gene-for-gene model (Flor, 1971) for host specificity in plant pathogen interactions. In this model, the products of avirulence genes of a pathogen induce a hypersensitive response in plants that carry the corresponding resistance genes. The hypersensitive response results in the rapid death of a few cells that surround the infection site, thereby preventing further growth of the pathogen. Interaction between the fungal pathogen *Cladosporium fulvum* and tomato is a model system, which has been used to study the molecular basis of gene-for-gene based resistance (De Wit, 1995). The AVR9 gene of *C. fulvum* encodes a 63 amino acid peptide, which after secretion is processed by fungal and plant proteases to the mature 28 amino acid AVR9 elicitor peptide. This peptide is the only factor responsible for the induction of active defense responses in tomato plants that carry the complementary *Cf-9* resistance gene. The AVR9 peptide consists of a triple-stranded anti-parallel  $\beta$ -sheet with three disulphide bridges grouped in a cystine knot motif, very similar to that described for circulin A.

In addition, there are some structural similarities between circulin A and the family of plant defensins, a class of cystine-rich anti-microbial peptides (Broekaert *et al.*, 1995). While these molecules are slightly larger than circulin A and kalata-B1, consisting of typically 45-54 amino acids residues and containing additional cysteine residues, they have some structural features in



common. In particular, their 3D structures are dominated by a triple-stranded anti-parallel  $\beta$ -sheet and a single  $\alpha$ -helix in parallel with the  $\beta$ -sheet. The helix in circulin A is much smaller than that in the defensins. This may be a reflection of the smaller size of circulin A relative to plant defensins. Broekaert *et al.* have noted that plant defensins belong to a superfamily of similarly folded anti-microbial peptides that has representatives in vertebrates, invertebrates and plants, suggesting that these diverse molecules pre-date the evolutionary divergence of animals and plants. It is thus not surprising that there are structural similarities between circulin A and some small disulphide-rich animal proteins. The similarity of kalata-B1 to the  $\omega$ -conotoxin GVIA, a potent calcium-channel blocker from the venom of *Conus geographus*, has already been noted (Chapter 1.2).

Also of interest is the anti-HIV activity of circulin peptides. Although the mechanism is hitherto unknown, the structure of circulin A described above lends itself to speculation that interaction with biological membranes may be involved. The separation of hydrophobic and positively charged residues suggests the possibility that circulin A could insert its hydrophobic region into a membrane, with the clustered positively charged groups interacting with the negative membrane surface. Membrane disruption could similarly account for the diverse biological activities observed for the other macrocyclic peptides.

It remains the subject of future investigations to ascertain the mode of anti-HIV activity, to discover why such a complex motif has evolved and the function such peptides serve in host plants. Continued studies on circulin A will provide many answers and the determination of its 3D structure will play a key role in fully understanding biological activity.

## 6.6 Summary

The three-dimensional solution structure of circulin A, a 30 residue polypeptide from the African plant *Chassalia parvifolia*, has been determined using two dimensional  $^1\text{H}$ -NMR spectroscopy. Circulin A was originally

identified based upon its inhibition of the cytopathic effects and the replication of the HIV virus (Gustafson *et al.*, 1994). Structural restraints consisting of 374 interproton distances inferred from nuclear Overhauser effects, 17 backbone dihedral and 9  $\chi_1$  angle restraints from spin-spin coupling constants and 8 hydrogen bond definitions were used as input for simulated annealing calculations and energy minimisation in the program X-PLOR. The final set of 12 structures have mean pairwise rms differences over the whole molecule of 1.01 Å for the backbone atoms, and 1.67 Å for all heavy atoms. For the well-defined region encompassing residues 2-12 and 18-27, the corresponding values were 0.32 Å and 1.45 Å, respectively.

Despite its small size, circulin A shows an exceptionally well defined 3D fold which consists of mainly  $\beta$ -turns. A  $\beta$ -hairpin is present and also a small region of  $\alpha$ -helix. A third  $\beta$ -strand links somewhat loosely to the  $\beta$ -hairpin and while the orientation and hydrogen bonding patterns are such that it can not rigorously be defined as a triple-stranded  $\beta$ -sheet, it does have the geometrical characteristics of a distorted triple-stranded anti-parallel  $\beta$ -sheet. The peptide backbone is extensively folded back on itself and is braced with disulphide bonds, two of which form an embedded loop completed by the backbone fragments connecting the cysteine residues. A third disulphide bond threads through the centre of this loop to form a cystine knot motif.

In essence, this motif consists of a central disulphide core which, due to the distribution of residue types, is solvent exposed on one face. In terms of spatial arrangement, circulin A is a flattened shape. The other face shields the central disulphide core by a concentration of hydrophobic residues. Positive charged residues are placed along one peripheral edge and the single tryptophan residue protrudes from the main core. The overall fold, shape and distribution of in particular the hydrophobic residues are very similar to kalata-B1. It can be expected that their biological action is also similar. The 3D structures of these molecules are crucial to their activities and the determination of their structure will be important in the future investigation to characterise activity.

**CHAPTER SEVEN**

**STRUCTURE DETERMINATION**

**OF CIRCULIN B**

## 7.1 Introduction

Circulin B peptide was kindly supplied by the research group associated with discovery, purification and characterisation of the circulins (Gustafson *et al.*, 1994). Circulin B is similar in sequence to circulin A and also has anti-HIV activity. However, there are a few differences and it was of interest to ascertain the effect of these differences on the structure. The sequence of circulin B, aligned according to that used for the first collective alignment of members of the inhibitor cystine knot family (Pallaghy *et al.*, 1994), is given below.

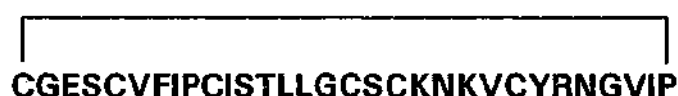


Figure 7.1. Sequence of circulin B (Gustafson *et al.*, 1994).

Circulin B has an additional residue to circulin A, thereby consisting of 31 residues. All the differences are in significant positions. The additional residue is located between C(6) and C(1) in the region corresponding to the disordered loop of kalata-B1. This loop has been postulated to be important in receptor recognition and binding (Saether *et al.*, 1995). Other differences are that W7 of circulin A, which is part of a  $\beta$ -turn, is replaced with F7 in circulin B, and A13 and A14 of the  $\alpha$ -helix in circulin A are replaced with bulkier residues T13 and L14. This chapter reports on the determination of the 3D structure of circulin B and compares it with circulin A. The numbering of residues used throughout this chapter will be in line with the sequence cited above in Figure 7.1 so that residue 1 is C1 of the cystine knot family. The experimental procedure differs from that used for circulin A and is described in the following section.

## 7.2 Materials and Methods

An overview of methods used in this thesis is given in Chapter 2. In the following section, specific details relevant to the study of circulin B are recorded.

### 7.2.1 NMR Experiments

Circulin B material was prepared for NMR data acquisition in 5 mm Wilmad NMR tubes as 1-2 mM concentration solutions which were pH adjusted to around pH 4. A 10%:90% by volume  $^2\text{H}_2\text{O}$ : $\text{H}_2\text{O}$  solution was used for the majority of the experiments including the collection of data for structure calculations. For deuterium exchange experiments, E-COSY experiments and additional information on alpha protons, a circulin B solution in 99.99%  $^2\text{H}_2\text{O}$  was used.

NMR data were kindly recorded by Dr. Norelle Daly (Centre for Drug Design and Development, University of Queensland, Queensland, Australia) on either a Bruker ARX 500 or a Bruker DRX750 spectrometer, each of which is equipped with a shielded gradient unit. Two dimensional data were acquired with a 5 mm inverse broadband probe in the phase sensitive mode using time proportional phase incrementation for quadrature detection in the  $t_1$  dimension (Marion and Wüthrich *et al.*, 1983) over around 6024 Hz. The data were acquired at several different temperatures namely 280, 282, 285, 290 and 300 K. The temperature was maintained using a temperature controlled stream of air, which was supplied by a Bruker BCU05 refrigeration unit and controlled by a B-VT 2000 control unit. Temperature and acquisition parameters were varied in an effort to maximise information and resolve spectral ambiguities. Details of pulse sequences used to extract structure information are described in the following section.

TOCSY spectra (Brauschweiler and Ernst, 1983) were recorded using an MLEV-17 spin lock sequence (Bax and Davis, 1985) with mixing periods of

either 80, 100, 109 or 165 ms. NOESY spectra (Jeener *et al.*, 1979) were recorded with mixing times of either 100, 250, 300 or 400 ms. Solvent suppression for NOESY and TOCSY spectra was achieved using a modified WATERGATE sequence (Piotto *et al.*, 1992) in which two gradient pulses of 2 ms duration and 6 G cm<sup>-1</sup> strength were applied either side of a binomial 3-9-19 pulse. The spectra were acquired with 4096 complex data points in F<sub>2</sub> and 512 increments in the F<sub>1</sub> dimension with 8 or 16 scans per increment for the TOCSY spectra and between 40 or 72 scans per increment for the NOESY spectra.

DQF-COSY (Rance *et al.*, 1983) and E-COSY (Griesinger *et al.*, 1987) spectra were also acquired with 4096 complex data points in F<sub>2</sub> and 512 increments in the F<sub>1</sub> dimension. The number scans per increment was between 40 to 60 for the DQF-COSY spectra and 84 for the E-COSY spectra. Solvent suppression for the COSY-type spectra was attained using selective low-power irradiation of the water resonance during a relaxation delay of 1.8 s.

Slowly exchanging NH protons were identified by acquiring a series of 1D and TOCSY spectra of the fully protonated peptide immediately following dissolution in <sup>2</sup>H<sub>2</sub>O. The <sup>3</sup>J<sub>NH-H $\alpha$</sub>  coupling constants were measured from DQF-COSY spectra. The <sup>3</sup>J<sub>H $\alpha$ -H $\beta$</sub>  coupling constants were measured from E-COSY spectra for non-valine residues. The DQF-COSY and E-COSY data were strip transformed to 16 K in both F<sub>1</sub> and F<sub>2</sub> dimensions for coupling constant measurement. A deconvolution program which fits Lorentzian line shapes to 1D rows was used to improve the accuracy of <sup>3</sup>J<sub>NH-H $\alpha$</sub>  coupling constants measured from DQF-COSY spectra, as this procedure minimises error due to finite line widths of anti-phase peak components.

All spectral data were processed on a Silicon Graphics Indy workstation using XWINMR (Bruker) software. The t<sub>1</sub> dimension of 2D data was zero-filled to 2048 real data points and 90° phase-shifted sine-bell window functions were applied prior to Fourier transformation. Polynomial baseline correction was used in regions either side of the water resonance to improve the spectral appearance of the 2D data. The 1D data were multiplied by an exponential

function with line broadening of 0.3 prior to Fourier transformation. Chemical shifts were referenced to DDS set at 0.00 ppm. The processed data were then used to determine distance restraints for structure calculations, which were conducted as follows.

### 7.2.2 Structure Calculations

Distance restraints were determined from a 400 ms NOESY spectrum recorded at 290K on the 750 MHz spectrometer. Distance data from spectra acquired with lower mixing times, at different temperatures and field strengths were used to confirm ambiguous information. Peak volumes were visually classified as strong, medium, weak or very weak and upper bound interproton distance restraints of 2.7, 3.5, 5.0 and 6.0 Å, respectively, were assigned (Williamson *et al.*, 1985; Clore *et al.*, 1986a). Pseudoatom corrections were applied to non-stereospecifically assigned methylene and methyl protons (Wüthrich *et al.*, 1983), so that for distances involving methylene protons, 1.0 Å was added to the upper limit, and for distances involving methyl protons, 1.5 Å was added to the upper limit.

Backbone dihedral restraints were inferred from  $^3J_{\text{NH-H}\alpha}$  coupling constants, with  $\phi$  restrained to  $-120 (\pm 30)^\circ$  for  $^3J_{\text{NH-H}\alpha}$  in the range of 8.0 to 9.5 Hz,  $-120 (\pm 15)^\circ$  for  $^3J_{\text{NH-H}\alpha}$  greater than 9.5 Hz and  $-65 (\pm 15)^\circ$  for  $^3J_{\text{NH-H}\alpha}$  less than 5.8 Hz (Pardi *et al.*, 1984; Wüthrich, 1986). Additional  $\phi$  restraints of  $100 (\pm 80)^\circ$  were applied to residues where the intraresidue  $\text{H}^\alpha\text{-H}_\text{N}$  NOE interaction was weaker than the sequential  $\text{H}^\alpha\text{-H}_\text{N}$  NOE interaction (Clubb *et al.*, 1994). Residues with intense intraresidue  $\text{H}^\alpha\text{-H}_\text{N}$  NOE interactions and a  $^3J_{\text{NH-H}\alpha}$  of  $\sim 7$  Hz were restrained to  $\phi$  of  $50 (\pm 40)^\circ$  (Ludvigsen and Poulsen, 1992; Omecinsky *et al.*, 1996). Finally,  $\chi_1$  dihedral angle restraints and stereospecific assignment of methylene protons were derived from  $^3J_{\text{H}\alpha\text{-H}\beta}$  coupling constants and intraresidue NOE intensities (Zuiderweg *et al.*, 1985; Wagner *et al.*, 1987). Ranges of  $\pm 30^\circ$  were used for all  $\chi_1$  dihedral angle restraints.

Hydrogen bonds were also included in the final 3D calculations. Hydrogen-deuterium exchange experiments permitted the identification of the amide proton hydrogen-bond donors. The hydrogen-bond acceptors for the amide protons were ascertained through spatial examination of the initial structures generated. Hydrogen bonds ( $i$ - $j$ ) were then expressed as distance restraints where distances of  $1.58 \leq d \leq 2.3$  Å for  $\text{HN}-\text{O}_j$  and  $1.58 \leq d \leq 3.2$  Å for  $\text{N}-\text{O}_j$  were defined.

Three dimensional structures were generated using simulated annealing and energy minimisation protocols in the program X-PLOR Online 3.851 (Brünger, 1992). The calculations were carried out on a Silicon Graphics (4D/30) computer. Initially, an *ab initio* simulated annealing protocol (Nilges *et al.*, 1988) was used to generate a set of 50 structures, starting from template structures with randomised  $\phi$  and  $\psi$  angles and extended side-chains. The backbone was cyclised throughout the calculations and all peptide bonds were defined as *trans*. At this stage, the disulphide bonds were included as pseudo-NOE restraints and the simulated annealing protocol consisted of 20 ps of high temperature molecular dynamics at 1000 K with a low weighting on the repel force constant and the NOE distance restraints.

This was followed by a further 10 ps of high temperature dynamics with an increased force constant on the NOE distance restraints. The disulphide bonds were formally bonded and the dihedral force constant increased prior to cooling the system to 0 K and increasing the repel force constant over 15 ps of dynamics. The NOE restraints were checked for violations and ambiguous cross peaks were resolved on the basis of interproton distances in the family of structures. A further 50 structures were calculated and these structures were refined using the conjugate gradient Powell algorithm with 1000 cycles of energy minimisation (Clore *et al.*, 1986b) and a refined force field based on the program CHARMM (Brooks *et al.*, 1983).

Structures were examined using PROMOTIF (Hutchinson and Thornton, 1996), checked using PROCHECK (Laskowski *et al.*, 1993) and displayed



using INSIGHT (Biosym Technologies, San Diego, CA). The resulting set of structures together with the analysis of the NMR data allowed for a detailed insight into the 3D topology of circulin B, which could then be compared to that of circulin A.

## 7.3 Results and Discussion

### 7.3.1 Proton Resonance Assignments

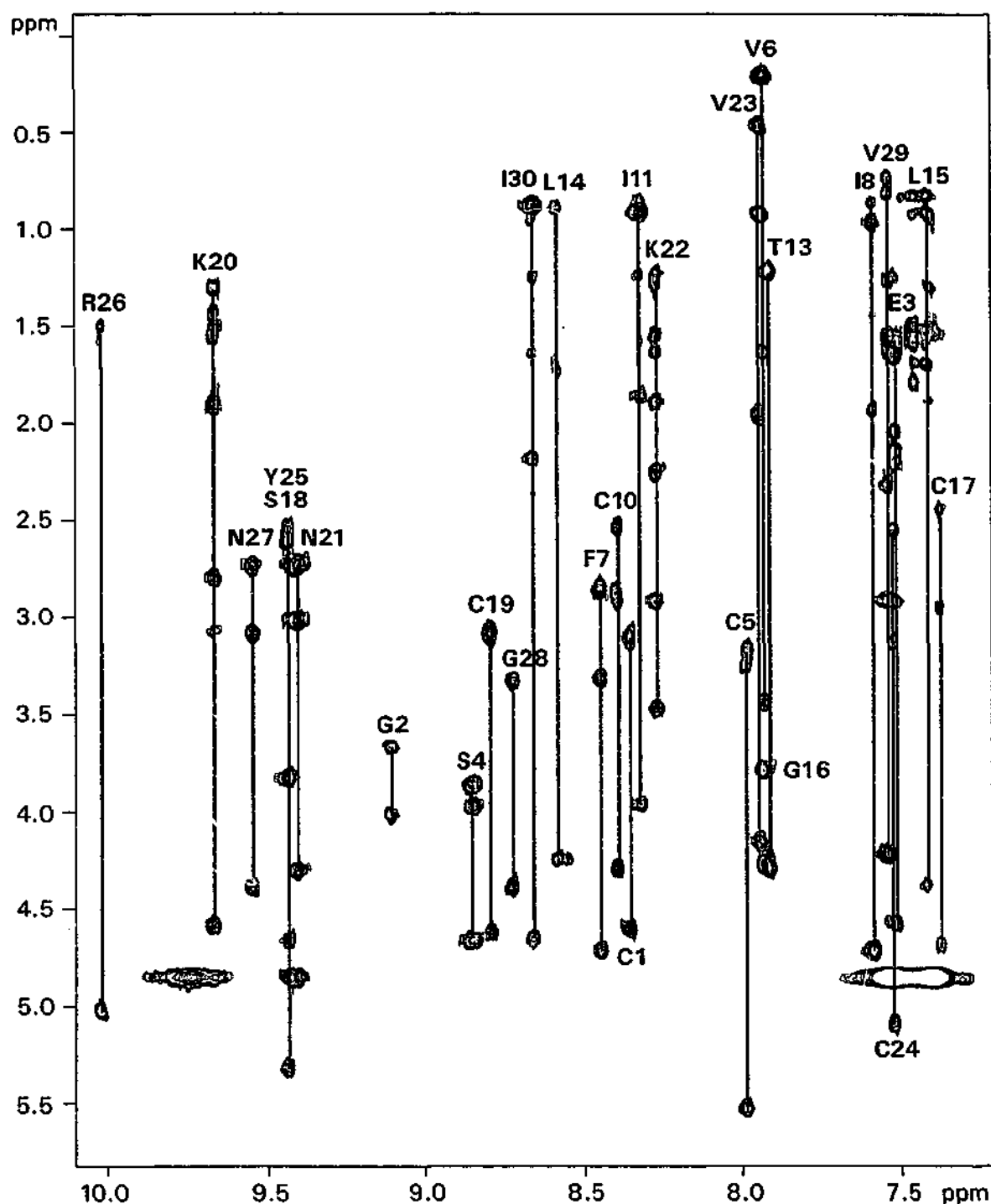
The proton assignment protocol used for circulin A was applied for circulin B (Wüthrich, 1986; King and Mackay, 1996). The resulting proton chemical shifts for the backbone and side-chain protons are given in Table 7.1. The fingerprint region of the TOCSY spectrum, with the spin systems labeled, is displayed in Figure 7.2, and the fingerprint region of the NOESY spectrum, showing the closed circuit of sequential NOEs in accordance with the cyclic nature of circulin B, is shown in Figure 7.3. The spectra resemble those of circulin A with well dispersed, easily recognisable amino acid spin systems indicative of a highly structured molecule. As in the case of circulin A, the presence of strong NOE cross peaks between  $\alpha\text{H}_{i-1}$  -  $\delta\text{H}_i$  protons confirms the *trans* configuration of the amide bonds involving both prolines and their preceding sequential residues. This latter finding is consistent with the bracelet-like subfamily of the cyclic cystine knot (CCK) family of peptides (Craik *et al.*, 1999) to which the circulins belong. The proton assignment could be then used to indicate secondary structure especially in relation to circulin A.

### 7.3.2 Secondary Structure

The NMR data of circulin B was analysed for secondary structure in the same manner as for circulin A. The procedure involved firstly, calculation of CSI values based on alpha proton chemical shifts (Wishart *et al.* 1992). These

**Table 7.1.** Proton chemical shifts (ppm) for Circulin B at 290 K. Shifts are measured relative to DDS set at 0 ppm.

	RESIDUE	HN	$\alpha$ H	$\beta$ H	OTHER	
1.	CYS	8.34	4.61	3.12		
2.	GLY	9.11	4.03, 3.68			
3.	GLU	7.52	4.58	2.05, 1.69	$\gamma$ CH <sub>2</sub> 2.21, 2.16	
4.	SER	8.85	4.67	3.98, 3.88		
5.	CYS	7.99	5.53	3.26, 3.18		
6.	VAL	7.93	3.45	1.64	$\gamma$ CH <sub>3</sub> 0.94, 0.22	
7.	PHE	8.45	4.72	3.32, 2.86	C2,6 7.23 C3,5 7.31	
8.	ILE	7.59	4.72	1.94	$\gamma$ CH <sub>2</sub> 1.42, 1.03 $\gamma$ CH <sub>3</sub> 0.98	$\delta$ CH <sub>3</sub> 0.87
9.	PRO		4.37	2.36, 1.78	$\gamma$ CH <sub>2</sub> 1.96, 1.89	$\delta$ CH <sub>2</sub> 3.73, 3.58
10.	CYS	8.39	4.30	2.93, 2.54		
11.	ILE	8.27	3.97	1.87	$\gamma$ CH <sub>2</sub> 1.59, 1.25 $\gamma$ CH <sub>3</sub> 0.92	$\delta$ CH <sub>3</sub> 0.86
12.	SER	9.81	4.59	4.14, 3.88		
13.	THR	7.92	4.29	4.26	$\gamma$ CH <sub>3</sub> 1.23	
14.	LEU	8.58	4.25	1.75, 1.57	$\gamma$ CH 1.67	$\delta$ CH <sub>3</sub> 0.92, 0.88
15.	LEU	7.42	4.39	1.71	$\gamma$ CH 1.61	$\delta$ CH <sub>3</sub> 0.93, 0.84
16.	GLY	7.93	3.80			
17.	CYS	7.38	4.70	2.96, 2.45		
18.	SER	9.44	4.67	3.83		
19.	CYS	8.79	4.63	3.10		
20.	LYS	9.67	4.59	1.91	$\gamma$ CH <sub>2</sub> 1.31 $\delta$ CH <sub>2</sub> 1.55, 1.45	$\epsilon$ CH <sub>2</sub> 3.08, 2.80 $\epsilon$ NH <sub>3</sub> <sup>+</sup> 7.37
21.	ASN	9.40	4.31	3.03, 2.75	$\gamma$ NH <sub>2</sub> 7.61, 6.91	
22.	LYS	8.27	3.48	2.27, 1.90	$\gamma$ CH <sub>2</sub> 1.27 $\delta$ CH <sub>2</sub> 1.64, 1.55	$\epsilon$ CH <sub>2</sub> 2.93 $\epsilon$ NH <sub>3</sub> <sup>+</sup> 7.51
23.	VAL	7.95	4.16	1.97	$\gamma$ CH <sub>3</sub> 0.94, 0.47	
24.	CYS	7.52	5.10	3.14, 2.56		
25.	TYR	9.44	5.32	2.62, 2.55	C2,6 H 6.71 C3,5 H 6.86	
26.	ARG	10.02	5.03	1.80	$\gamma$ CH <sub>2</sub> 1.59, 1.51	$\delta$ CH <sub>2</sub> 3.22, 3.05 NH 7.46
27.	ASN	9.54	4.39	3.10, 2.74	$\gamma$ NH <sub>2</sub> 7.34, 6.91	
28.	GLY	8.72	4.39, 3.34			
29.	VAL	7.54	4.22	2.33	$\gamma$ CH <sub>3</sub> 0.82, 0.74	
30.	ILE	8.66	4.66	2.19	$\gamma$ CH <sub>2</sub> 1.65, 1.26 $\gamma$ CH <sub>3</sub> 0.88	$\delta$ CH <sub>3</sub> 0.96
31.	PRO		4.30	2.23, 1.89	$\gamma$ CH <sub>2</sub> 2.23, 2.05	$\delta$ CH <sub>2</sub> 4.40, 3.87



**Figure 7.2.** Fingerprint region of a 500 MHz TOCSY spectrum for circulin B showing assignments of side chain spin systems linked with their corresponding amide protons at 290 K in 90% H<sub>2</sub>O / 10% <sup>2</sup>H<sub>2</sub>O.

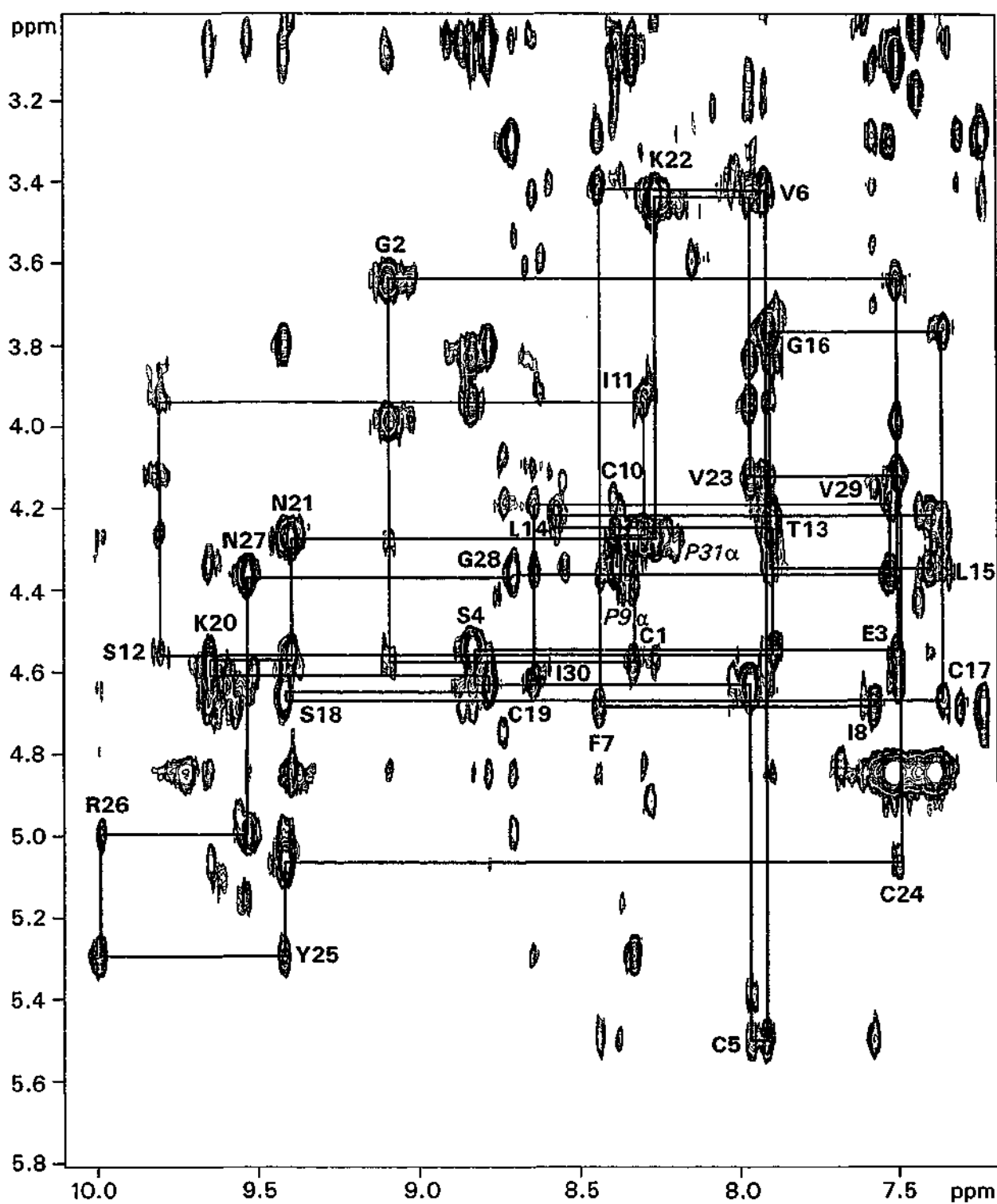
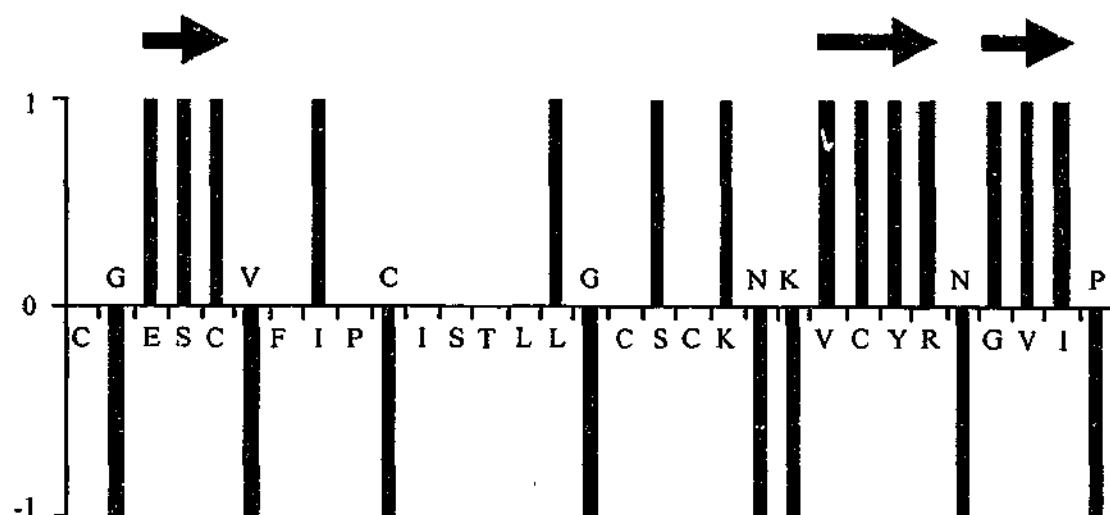


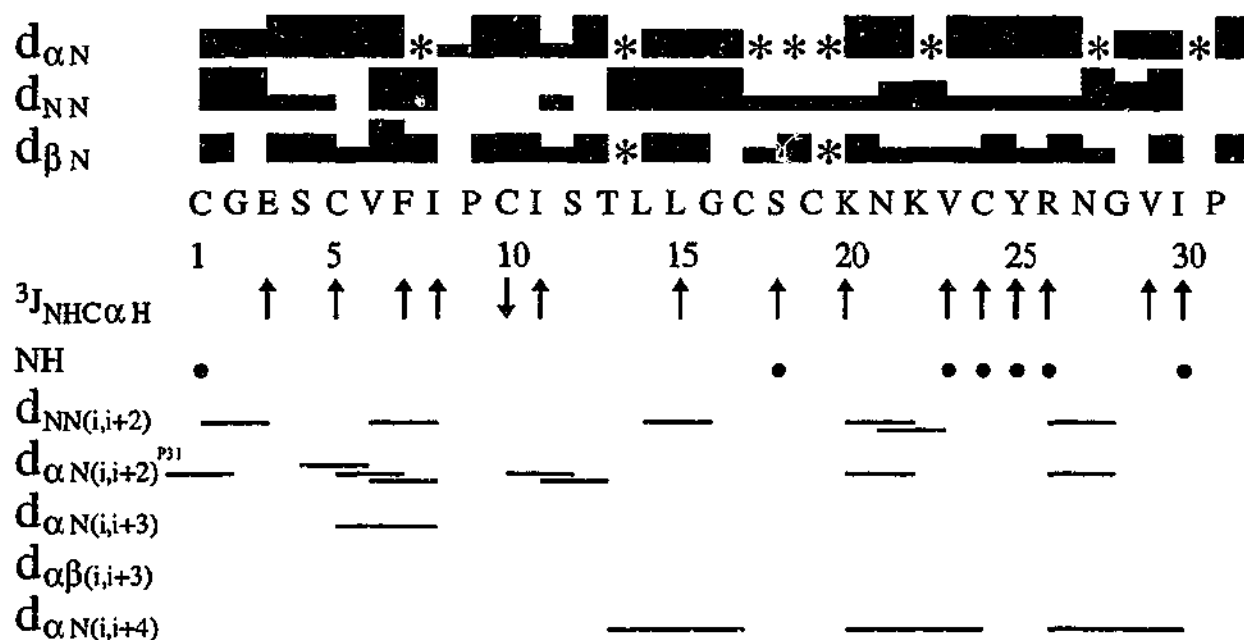
Figure 7.3. Fingerprint region of the 500 MHz NOESY spectrum of circulin B at 290 K in 90% H<sub>2</sub>O / 10% 2H<sub>2</sub>O. The cyclic sequential connectivity pattern is shown.

are graphically displayed in Figure 7.4. Secondly, short to medium range NOE interactions,  $^3J_{\text{NH-H}\alpha}$  coupling constants and slow exchanging NH protons were determined, and in combination are summarised in Figure 7.5. Finally, the non-sequential NOE interactions of the triple-stranded  $\beta$ -sheet were established. These are schematically shown in Figure 7.6, along with the slowly exchanging NH protons.

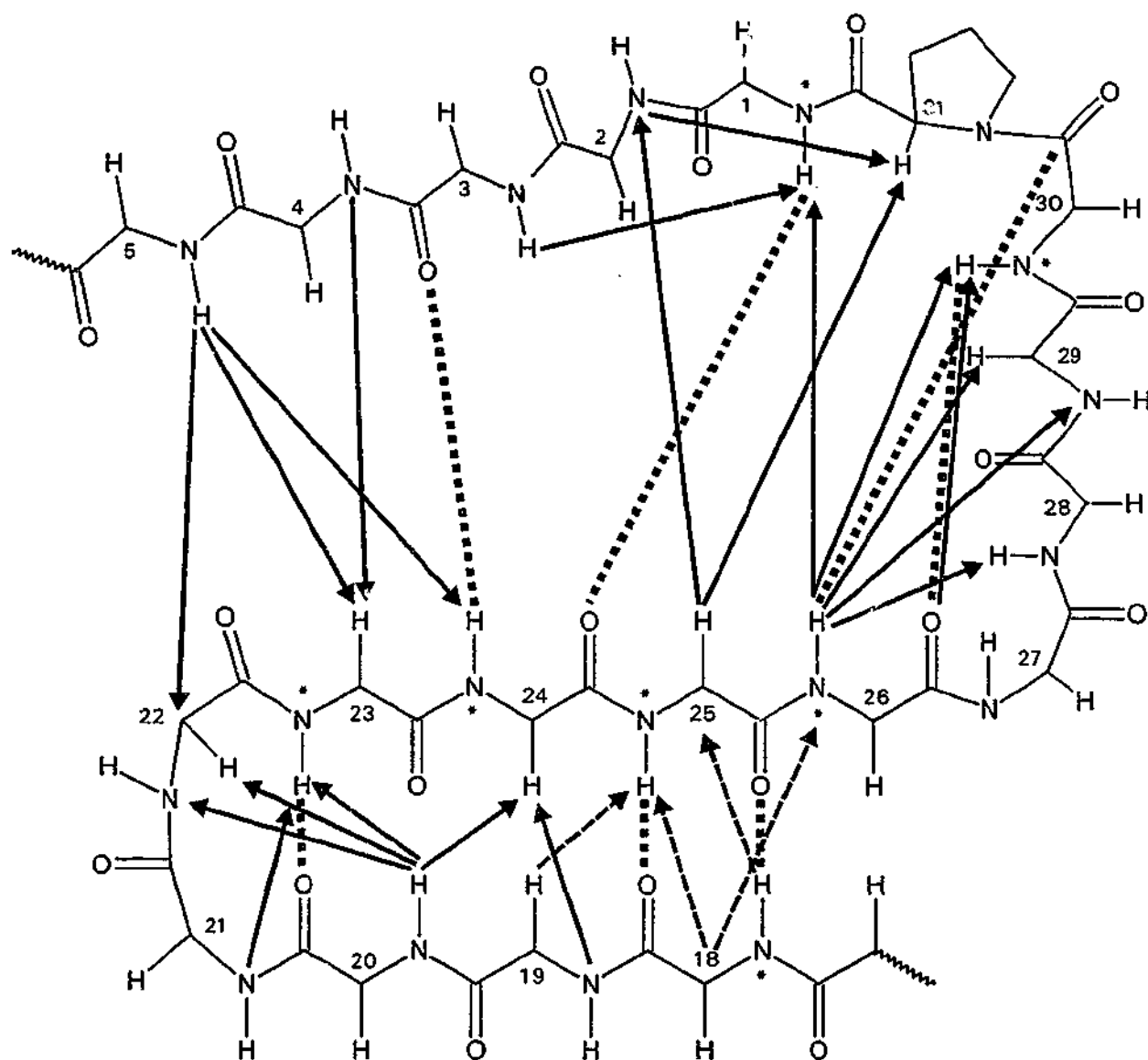
Commencing with examination of the CSI values, it is immediately apparent that like circulin A, circulin B is highly structured, with CSI values for the majority of residues differing significantly from their random chemical coil values (Figure 7.4).  $\beta$ -Strand character predicted in circulin A is found over residues V23 to R26 (C24 to R26 in circulin A), but is not predicted over residues G16 to K20. Additional  $\beta$ -strand character in circulin B is predicted over residues E3 to C5 and G28 to I30. While part of the CSI differences between the circulins may be due to the different temperature and pH conditions under which CSI values were determined, another influence may be due to slight but real difference in local structure. For instance, the  $\beta$ -strand over residues G28 to I30 may be the direct consequence of the additional residue in this loop of circulin B, which allows the loop to be long enough to form a  $\beta$ -strand in circulin B. The absence of the G16 to K20  $\beta$ -strand of circulin A may be the result of bulkier  $\alpha$ -helix residues in circulin B (T13, L14) compared to circulin A (A13, A14) which produce distortion of the neighbouring region in circulin B. The prediction of  $\beta$ -strand character around E3 to C5 is not surprising as this segment forms part of the triple-stranded  $\beta$ -sheet. The fact that it is not evident from the CSI values for circulin A may be related to the less bulky side-chain of the aromatic F7 residue in circulin B, which is replaced by the comparatively bulkier side-chain of the aromatic W7 residue in circulin A. The remainder of the CSI values are random variations of positive, negative and zero values not unlike circulin A, indicating the presence of turn features. Combining CSI values with other parameters measured from the NMR data gives further support to these and other observations (Figure 7.5).



**Figure 7.4.** CSI values derived from alpha proton chemical shifts of circulin B at 290 K. The prediction of  $\beta$ -strand character is indicated by solid arrows.



**Figure 7.5.** Summary of the sequential and medium-range NOE connectivities,  $^3J_{\text{NH}\alpha\text{H}}$  coupling constants and slowly exchanging NH protons observed for circulin B. Filled bars indicate sequential connectivities observed in a 250 ms NOESY spectrum at 290 K, pH 4. Shaded bars correspond to sequential  $\text{H}^\alpha\text{-H}^{\beta+1}$  and  $\text{NH-H}^{\beta+1}$  connectivities for proline residues. The height of the bar indicates the strength of the NOE interaction. Overlapping NOE interactions are indicated by an asterisk; ( $\uparrow$ ) indicate  $^3J_{\text{NH}\alpha\text{H}}$  coupling constants  $\geq 8.5$  Hz. Slowly exchanging backbone NH protons observed in a TOCSY spectrum recorded in  $^2\text{H}_2\text{O}$  are indicated by filled circles.



**Figure 7.6.** Schematic diagram of the  $\beta$ -sheet region of circulin B showing the interstrand NOE interactions (arrows), potential hydrogen bonds (broken lines) and slowly exchanging amide NH bonds (\*). For clarity, sequential NOE interactions are not shown. The broken arrows correspond to NOE interactions which are ambiguous due to overlap.

As expected, large coupling constants are mostly associated with  $\beta$ -strand character identified from CSI values (Wüthrich, 1986). Furthermore, the presence of several  $d_{NN}(i,i+2)$  and  $d_{\alpha N}(i,i+2)$  NOE cross peaks are consistent with the occurrence of turns. Many of these are similar to circulin A, although more such NOE cross peaks are generally observed for circulin B in the same areas. This may be due to the use of higher field NMR data in the analysis of circulin B. However, there is evidence for an additional turn in circulin B around residues P31 to C2 and C1 to E3. It is possible that this may be due to a sharper turn in this region which must occur between the two  $\beta$ -strands on either side. As in the case of circulin A, an  $\alpha$ -helix is suggested by  $d_{\alpha N}(i,i+4)$  NOE interaction between T13 and C17, but is not supported by the requisite series of negative CSI in the region.

Also similar to circulin A, slow exchange NH protons are observed involving the same residues in circulin B (Figures 7.5 and 7.6). These include residues C1, S18, V23, C24, Y25, R26, I30. However, there is one exception and this is C5 where a slowly exchanging NH proton is not evident in circulin B. It may be speculated that perhaps the presence of a more hydrophobic aromatic F7 residue may cause a weakening of any hydrogen bond interaction involving C5. Despite the absence of the hydrogen bond, a long range NOE interaction between the NH proton of C5 and the alpha proton of C22 does indicate their close proximity, thereby confirming the presence of the triple-stranded  $\beta$ -sheet region of the cystine knot as shown in Figure 7.6.

Moreover, Figure 7.6 clearly shows the network of potential hydrogen bonds and NOE interactions between the non-sequential backbone atoms of the triple-stranded  $\beta$ -sheet region of circulin B. There are more inter-strand NOE interactions than obtained for circulin A. Once again, this may be due to the acquisition of the NMR data at a higher magnetic field for circulin B. The diagram shows the inter-strand relationship with termination points indicated by CSI values as turn centres at N21, K22, at N27 and at P31. A bulge/turn around C1 and G2 is also illustrated, as indicated by NOE interactions between these residues and residues Y25, R26 and what appears as a



termination point at G2 according to CSI values and the NOE interactions in this region. This latter observation is not obvious in circulin A.

In summary, the combined information from the measurable parameters in NMR spectra indicate a high degree of similarity in terms of the global fold for circulin A and B, and highlight regions of subtle structural changes, which may be related to residual differences and consequently may be expected to affect activity. Examination of the 3D calculated structure for circulin B gives further detail of structural changes.

### 7.3.3 Three Dimensional Structure

#### 7.3.3.1 Quality of Structures

A set of 50 structures was generated from dynamical simulated annealing calculations based on NOE distance and angle restraints determined from the NMR data. A total of 600 distance restraints was used. These are listed in Appendix VI. In addition, 17 backbone dihedral restraints, 11  $\chi_1$  angle restraints and 14 distance restraints inferred from 7 slowly exchanging NH protons and likely carbonyl acceptors were also used. These restraints are given in Appendix VII. From the 50 structures, the twelve structures with the lowest NOE energies were selected for detailed examination. The energetic and geometric statistics of these twelve structures are shown in Table 7.2.

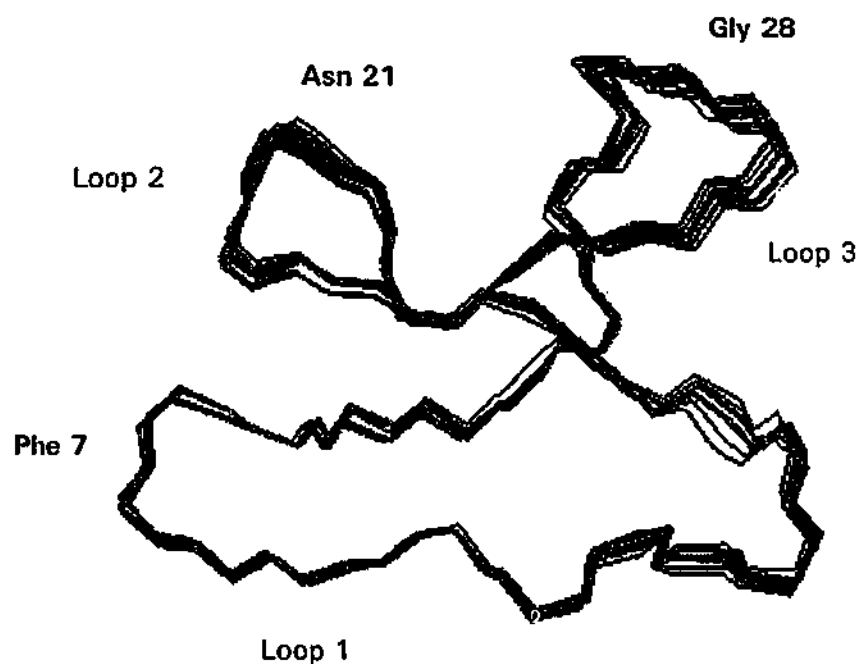
In regard to the quality of the final structures, the statistical data indicate that ideal covalent geometry has been largely maintained with no significant deviations. The structures fit the experimental parameters with minimal violations. There were no NOE violations greater than 0.2 Å and no dihedral violations greater than 2°. A mean pairwise RMSD of 0.30 Å was achieved over the backbone atoms and 0.83 Å for all non-hydrogen atoms. The twelve structures, displayed superimposed in Figure 7.7, are of sufficiently high quality for analysis of the 3D spatial arrangement.

<i>Mean Pairwise RMS deviations (Å)<sup>a</sup></i>	
Backbone	0.30 ± 0.09
Heavy atom	0.84 ± 0.18
<i>Mean RMS deviations from experimental restraints</i>	
Interproton distances (Å)	0.025 ± 0.0003
Dihedral angles (deg.)	0.34 ± 0.05
<i>Mean RMS deviations from idealized covalent geometry<sup>b</sup></i>	
Bonds (Å)	0.009 ± 0.0002
Angles (deg.)	2.21 ± 0.02
Impropers (deg.)	0.29 ± 0.02
<i>Mean Energies (kcal mol<sup>-1</sup>)</i>	
$E_{\text{NOE}}^c$	11.53 ± 0.26
$E_{\text{cdih}}^c$	0.25 ± 0.06
$E_{\text{L-J}}^d$	-107.3 ± 1.7
$E_{\text{bond}}$	6.12 ± 0.23
$E_{\text{improper}}$	1.53 ± 0.19
$E_{\text{angle}}$	56.6 ± 0.98
$E_{\text{total}}$	-24.6 ± 1.69
<i>Ramachandran plot statistics<sup>e</sup></i>	
Residues in most favoured regions (%)	64
Residues in additional allowed regions (%)	30
Residues in generously allowed regions (%)	6
Residues in disallowed regions (%)	0
The values in the Table are the mean ± standard deviation.	
<sup>a</sup> RMS deviation measured for the whole molecule.	
<sup>b</sup> Idealised geometry as defined by CHARMM force field and as implemented within XPLOR.	
<sup>c</sup> Force constants for the calculation of square-well potentials for the NOE and dihedral angle restraints were 50 kcal mol <sup>-1</sup> Å <sup>-1</sup> and 200 kcal mol <sup>-1</sup> rad <sup>-2</sup> , respectively.	
<sup>d</sup> The Lennard-Jones van der Waals energy was calculated with the CHARMM empirical energy function.	
<sup>e</sup> Ramachandran statistics were calculated using PROCHECK (Laskowski <i>et al.</i> , 1996).	

**Table 7.2.** Geometric and energetic statistics for the twelve lowest NOE energy structures for circulin B.

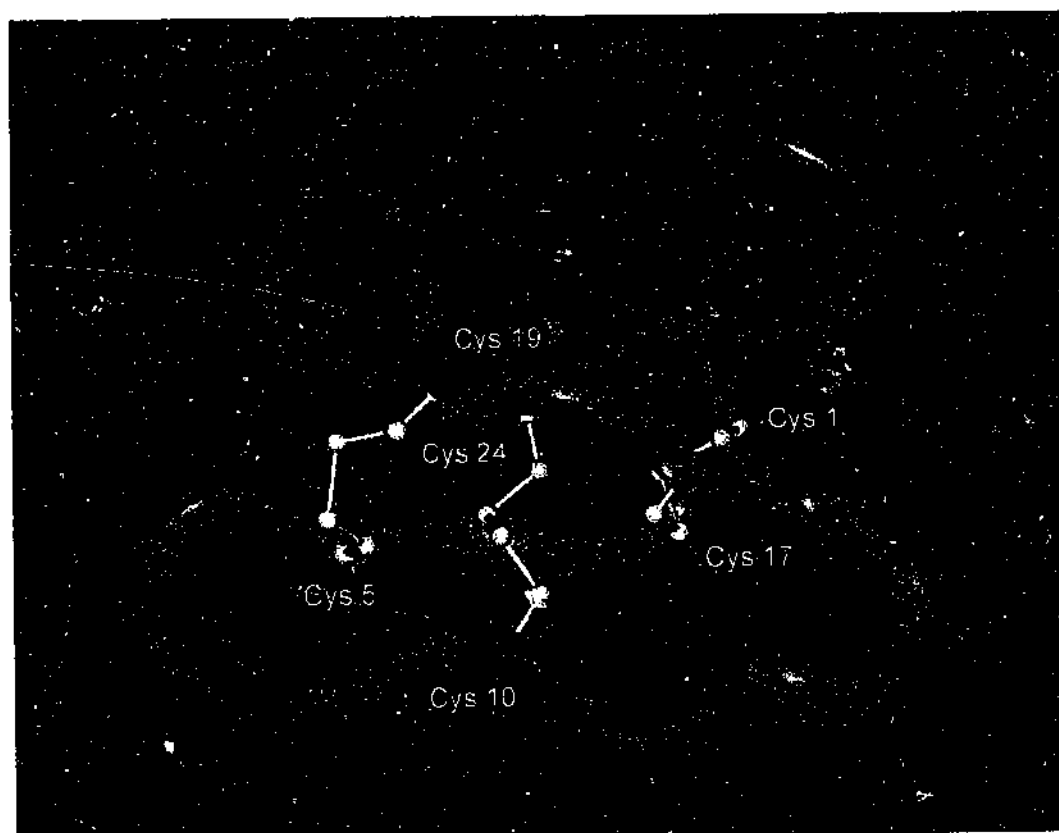
### 7.3.3.2 Three Dimensional Arrangement

The lowest energy structure is exhibited in Figure 7.8. It is immediately apparent that the overall fold is very much like that of circulin A. The



**Figure 7.7.** Superimposition of the twelve lowest NOE energy NMR-derived structures over the backbone (N, C $\alpha$ , C') cysteine residues.

structure is compact and consists of several turns. Application of the structural classification program PROMOTIF (Hutchinson and Thornton, 1996) to the twelve structures resulted in the identification of secondary structure that is mostly similar to circulin A, but with some differences. The  $\beta$ -turns occur between residues 5-8, 19-22, 20-23, 25-28, 26-29, 27-30. The  $\beta$ -turn between residues 27-30 is present in circulin B only and is consistent with previous findings of perhaps a sharper turn in this region due to the additional residue of circulin B being part of a unique  $\beta$ -strand. The  $\beta$ -turns between residues 6-9 and 10-13 in circulin A are not clearly evident in circulin B. These regions are sites of residual substitution in circulin B, being the different aromatic residue at F7, and the bulkier  $\alpha$ -helix residues at T13 and L14. All the  $\beta$ -turns in circulin B are in the miscellaneous, type IV category of  $\beta$ -turns,



**Figure 7.8.** Lowest energy structure of circulin B. The backbone is shown by a pink ribbon and the disulphide bonds are shown in yellow in ball and stick form.

except the  $\beta$ -turn between residues 20-23, which is type I'. In circulin A, this  $\beta$ -turn is non-standard, type IV and the  $\beta$ -turn involving residues 26-29 is type I'. The differences in the angles of these particular turns between the circulins may once again be related to the structural and residual changes in the region. Other less subtle differences, as well as similarities, between the circulins are also observed.

The  $\beta$ -hairpin in circulin B is defined over residues 18-25 and is comprised of  $\beta$ -strands including residues 18-25 and 23-25. The  $\beta$ -hairpin is identically classified in circulin A. A  $3_{10}$   $\alpha$ -helix is classified over residues 11-15 in all the structures. In circulin A, this  $\alpha$ -helix is defined as a  $3_{10}$  helix in most structures,  $\alpha$ -helical in a minority of structures and is not always present. The difference in  $\alpha$ -helices between the circulins may be due to increased NOE information for circulin B, or it may be, in part, a consequence of the  $\alpha$ -helix being more stable in circulin B, as a result of substitution of the two central residues of the  $\alpha$ -helix between the circulins. In addition, the  $\alpha$ -helix of circulin B is slightly longer, with averages in length of 8.68 Å and pitch 6.48 Å compared to circulin A with averages in length of 8.41 Å and pitch 5.81 Å. This is not surprising considering that the two different  $\alpha$ -helix residues are bulkier in circulin B. The potential for increased stability of the  $\alpha$ -helix in circulin B is supported by more order in the  $\alpha$ -helix region. Angular order parameters were calculated for the twelve structures and these are listed in Table 7.3. Some degree of disorder is inferred around residues T13 and G16 for circulin B, while in circulin A disorder is evident over the entire of the  $\alpha$ -helix residues 11-17.

Other than the  $\alpha$ -helix region, the regions of disorder are much the same in both the circulins. For circulin B, disorder in other regions is noted at residue 28 and from 31-2. The remainder of the molecule is largely ordered, with order angular parameters approaching unity. For circulin A, disorder in other regions is found at residues 29 and 2-4, and the rest of the peptide is ordered. There is a slight shift in the region of disorder between the circulins around residue 1. This may be explained by the residue differences in the

Table 7.3. Angular order parameters for circulin B.

Residue	phi	psi	chi 1
1	0	0.99	0.87
2	0.99	1.00	0
3	1.00	1.00	1.00
4	0.99	1.00	1.00
5	1.00	1.00	0.89
6	1.00	1.00	1.00
7	1.00	1.00	1.00
8	1.00	1.00	1.00
9	1.00	1.00	1.00
10	1.00	1.00	1.00
11	1.00	1.00	1.00
12	1.00	1.00	0.99
13	1.00	1.00	0
14	1.00	1.00	1.00
15	1.00	0.98	0.97
16	0.97	0.99	0
17	1.00	0.99	0.91
18	1.00	1.00	0.87
19	1.00	1.00	0.99
20	0.99	1.00	1.00
21	1.00	1.00	0.40
22	0.99	0.99	1.00
23	0.99	0.99	1.00
24	1.00	1.00	1.00
25	1.00	1.00	1.00
26	1.00	1.00	1.00
27	1.00	1.00	1.00
28	1.00	1.00	0
29	1.00	0.99	0.84
30	0.99	1.00	1.00
31	1.00	0	1.00

region, noted already in the results section. Finally, charge and hydrophobic distribution is considered below.

The space-filling model for circulin B was produced as for circulin A. The three views including 90° and 180° degree rotations about the vertical axis are display in Figure 7.9. The distribution of residue function is indicated by colour-coding. Hydrophobic residues are coloured green, negatively charged residues in red, positively charged residues in dark blue, polar residues in light blue and cysteines in yellow. From a superficial, overall view, circulin B is remarkably similar to circulin A. The shape is identical, being flat with similar cross-sectional diameters of approximately 22 Å viewed from the front and 16 Å viewed from the side. The distribution of the residue type is also similar. One of the flat faces is rich in hydrophobic residues, while the opposite face exposes the cystine core to solvent. The positively charged residues are located along one peripheral edge and the F7 residue protrudes from the peptide core.

These observations, along with the previous findings from the analysis of the 3D calculated structure are in line with structural implications from NMR measurable parameters. The circulins A and B have a very similar global fold, albeit, subtle differences in structure, evidently a consequence of a single residue addition and three residual substitutions. A more comprehensive and thorough comparison of the topologies of the circulins, especially in relation to activity data is given below.

## **7.4 Comparison of Circulin Topology**

### **7.4.1 Conserved Features**

The lowest energy structures of circulin A and B were superimposed over the disulphide backbone atoms. An RMSD value of 0.84 Å is achieved, which is significantly better than that for the superimposition of circulin A and kalata-B1 (Chapter 6). The value is improved, giving an RMSD of 0.79 Å,



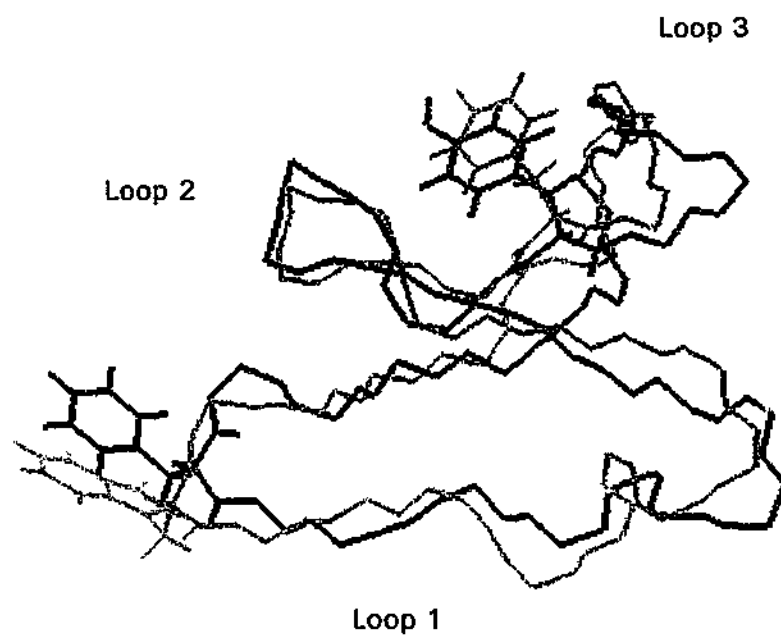
**Figure 7.9.** Space-filling models of circulin B showing the distribution of hydrophobic (green), negatively charged (red), positively charged (dark blue), polar (light blue) and cysteine (yellow) residues. The middle and lowest views are the 90° and 180° rotations respectively of the upper view about the vertical axis.



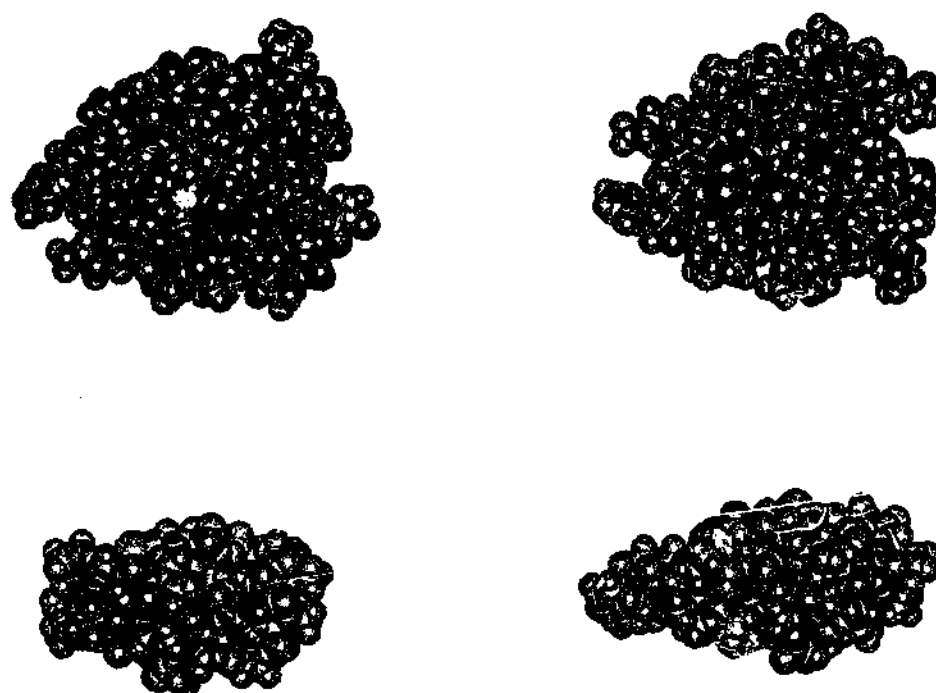
when measured over the backbone atoms of the cystine knot residues, XXC(2), XC(5)X, XXC(6)X, where X is any residue except cysteine and C are the Cys residues (Pallaghy *et al.*, 1994). The superimposition over backbone knot residues is displayed in Figure 7.10. The diagram illustrates several features in common.

Overall, the general fold is maintained. Major turns occur in the same regions. A single twist of a  $3_{10}$   $\alpha$ -helix is found in both circulins that is not present in kalata-B1, and furthermore the  $\alpha$ -helix is in the same location. Particularly notable is the position of residue 7, which in circulin A and B is aromatic. This residue is phenylalanine in circulin B and tryptophan in circulin A. As in circulin A, the residue is suspended from the turn at around residues 5-9 of loop1 and protrudes into the solvent environment. The residue is hydrophobic in character and it is highly unusual to have hydrophobic residues extending into the external environment. This does not occur in kalata-B1. It suggests that the aromatic residue 7 might play a role in activity, such as in peptide recognition by a membrane, and could therefore be a conserved feature in circulin peptides. The second aromatic residue in the circulins is the same for both circulins, being tyrosine 25. It is located between loops 2 and 3, buried within the compact peptide core. The spatial locations of the aromatic residues are shown in Figure 7.10.

Another striking aspect of the circulins is their anisotropy. The peptides are flat structures, consisting of two larger faces and a thinner edge. Figure 7.11 displays the molecular anisotropy exhibited by the space-filling diagram of circulin B. Relating to this peculiar structure, is the clustering of hydrophobic and charged residues, resulting in one face forming a hydrophobic surface and one edge containing the three cationic residues. This feature of clustering is characteristic of antimicrobial cationic peptides (Hancock *et al.*, 1995). Finally, a high percentage of around 90% of the amino acids of the circulins are identical, including the cationic residues. Despite the high degree of chemical and physical similarity of the fold in circulin A and B, there are major differences in activity.



**Figure 7.10.** Superimposition of circulin A (grey) and circulin B (black) over the cystine knot backbone (N, C $^{\alpha}$ , C') residues. The aromatic side-chains are displayed for each molecule.



**Figure 7.11.** Space-filling diagram of circulin B of the two larger faces and the thinner edges illustrating the molecular anisotropy of the circulins.

#### 7.4.2 Topological Differences in Relation to Activity

Recently, data on extensive activity testing of four CCK peptides, kalata-B1, circulin A, circulin B and cyclopsychotride A, on a range on microbes, fungi and cells was published (Tam *et al.*, 1999b). The results for circulin A and B are given in Table 7.4. Except for the effect on circulin B on a couple of the microbes, all activity was affected by the salt concentration, and the peptides were essentially inactive in a high salt concentration. This finding implicates an important electrostatic component in the factors determining activity. The activities for circulin A and B are remarkably different, with circulin B having a broader spectrum of activity, being twice as hemolytic and also being cytotoxic.

In an effort to explain the activity differences between circulin A and B in relation to structure, Tam *et al.* applied molecular modeling to the peptides to generate their structures based on the NMR-determined structure of kalata-B1. From the comparison of the resulting structures for the circulins, Tam *et al.* argue that as a consequence of insertion of the additional residue in circulin B, there are changes in the conformations of the cationic residues. This results in increased exposure of the cationic residues and a continuous alignment of the cationic residues along one edge of the circulin B molecule. Furthermore, the arginine residue in circulin A is folded back to its backbone and is shielded from the tryptophan residue absent in circulin B. The differences in comparison to the structures of cyclopsychotride A and kalata-B1 were considered responsible for the activity of circulin B against Gram-negative bacteria. To ascertain consistent differences in structure that could elucidate the differences in activities and verify assertions by Tam *et al.*, the average structures of circulin A and B were calculated from the family of twelve finally selected structures. The average structures and their surfaces were compared in detail.

With regard to the orientations of the cationic residues, Tam *et al.* appear to have a valid point. The side-chains of the three cationic residues

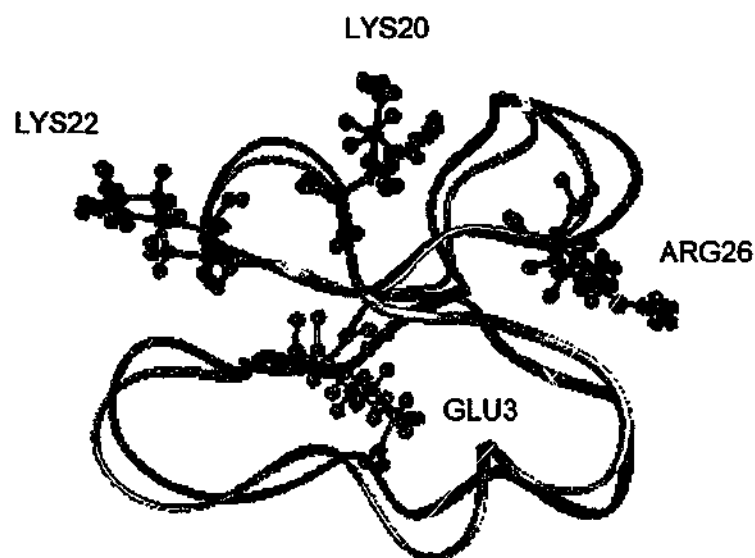
Activity	Organism	Circulin A MIC, $\mu$ M	Circulin B MIC, $\mu$ M
Bacterial, Gram negative	<i>Escherichia coli</i>	>500	0.41
	<i>Pseudomonas aeruginosa</i>	>500	25.5 <sup>a</sup>
	<i>Proteus vulgaris</i>	54.6	6.8
	<i>Klebsiella oxytoca</i>	>500	8.2 <sup>a</sup>
Bacterial, Gram positive	<i>Staphylococcus aureus</i>	0.19	13.5
	<i>Micrococcus luteus</i>	>500	>500
Fungal	<i>Candida albicans</i>	>500	>500
	<i>Candida kefyr</i>	18.6	29.0
	<i>Candida tropicalis</i>	19.4	>500
Hemolytic	Human erythrocytes	1020 <sup>b</sup>	550 <sup>b</sup>
Cytotoxic	Mouse fibroblasts	-	820 <sup>c</sup>

a Activity is largely unaffected by the salt concentration.

b concentration at which there is 50% red blood cell hemolysis.

c concentration at which there is 50% cell growth inhibition.

**Table 7.4.** Antimicrobial, hemolytic and cytotoxic activity of circulin A and B in low salt assays (Tam et al., 1999b).



**Figure 7.12.** Superimposition of circulin A (gray ribbon) and circulin B (black ribbon) over the cystine knot backbone (N, C $^{\alpha}$ , C') residues. The side-chains of the ionic residues are displayed for each molecule in ball and stick representation with light blue for the cationic residues of circulin A, dark blue for the cationic residues of circulin B, pink for the anionic residue of circulin A and red for the anionic residue of circulin B.

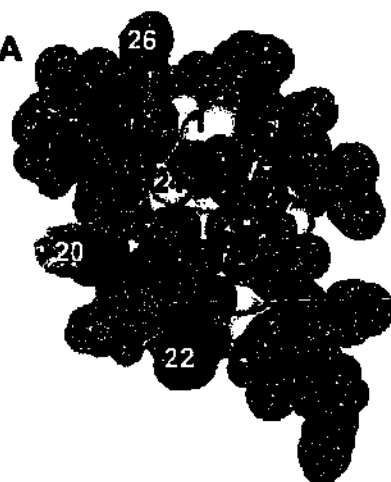
occupy different spatial orientations. The positions of the side-chains of K22 are dramatically altered between the circulins. The positions of side-chains of R26 are less different, although they are still markedly different, while it is only the ends of the side-chains of K20 that vary quite obviously. The spatial positions of the anionic residue of E3 are only slightly different. These observations are shown in Figure 7.12 in which the backbones of the circulins, represented by ribbons, are superimposed over corresponding backbone cystine residues and the ionic residues are displayed in ball and stick form.

Protrusion of the ionic residues in different dimensions is clearly evident in the colour-coded space filling diagrams of the circulins depicted in Figure 7.13. Based on the superimposition over corresponding backbone cystine residues, the identical surfaces of the two faces and the edge containing the cationic residues of the circulins are shown side by side. Colour-coding used previously in this chapter was applied; in other words, the hydrophobic residues are coloured green, negatively charged residues in red, positively charged residues in dark blue, polar residues in light blue and cysteines in yellow. Comparison of the surface differences is important, as the surfaces are the initial area of exposure and possibly interaction between the peptide and a membrane. Indeed, examination of the surfaces reveals poignant differences, which undoubtedly relate to the activity differences between the circulins.

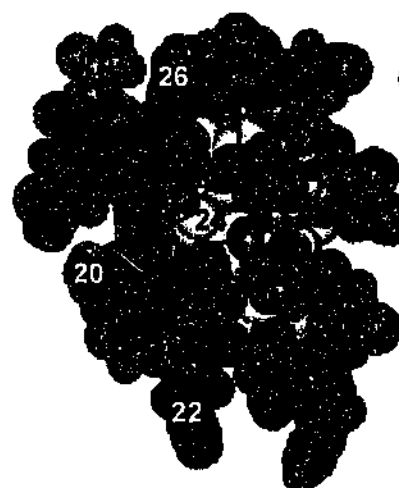
The first general observation is that the total visual surface areas occupied by the hydrophobic and cystine core faces appear to be slightly larger for circulin B. However, the total visual surface area occupied by the cationic edge appears slightly larger for circulin A. With consideration to the hydrophobic face (Figure 7.13), it can be seen that F7 is oriented towards K22 in circulin B, but the corresponding residues in circulin A, being W7 and K22, are oriented in orthogonal dimensions. This is quite likely to affect hydrogen bonding between C5 and K22, which is absent in circulin B but

### Hydrophobic Face

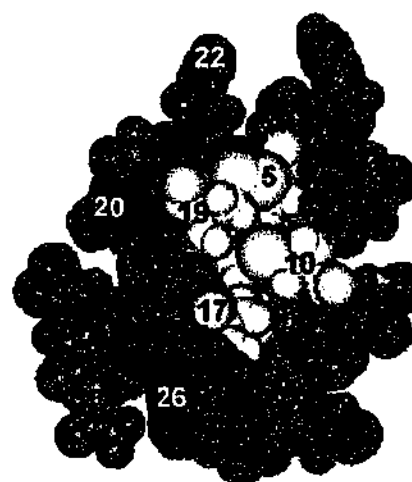
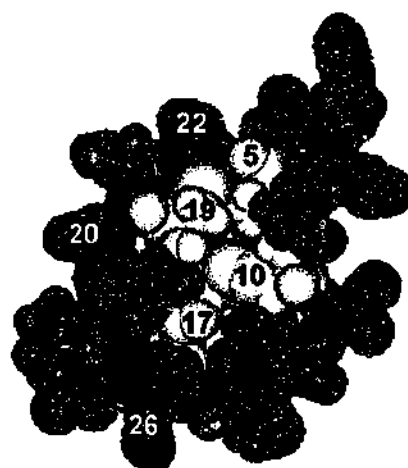
Circulin A



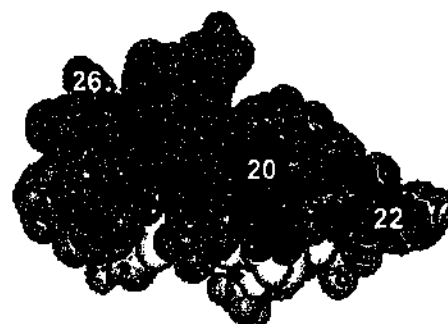
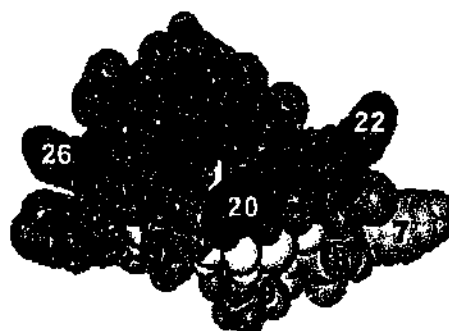
Circulin B



### Cystine Core Face



### Cationic Edge



**Figure 7.13.** Space-filling diagrams of the identical hydrophobic faces, cystine core faces and cationic edges for circulin A and circulin B. Hydrophobic residues are coloured green, negatively charged residues in red, positively charged residues in dark blue, polar residues in light blue and cysteine residues in yellow. Residues numbers are marked.

present in circulin A. Other differences that can be associated with residue substitution and insertion are also evident.

The top left corners of the hydrophobic surfaces of the peptides are also notably different (Figure 7.13). In circulin B, the corner is dominated by the polar N27 residue and in circulin A, by the hydrophobic I29 residue. Most probably, this observation is a consequence of the additional V29 residue in circulin B. On the opposite right top corner, the polar S12 residue is exposed in circulin A, but is more hidden in circulin B. In between the top corners, along the top edge, is a portion of the  $\alpha$ -helix. In circulin B, it is obviously more hydrophobic, with the two leucine residues exposed. In circulin A, only the single leucine residue protrudes. The net effect of the relocation of the  $\alpha$ -helix is that the cationic arginine residue is closer to the  $\alpha$ -helix in circulin B than in circulin A, where the arginine is slightly more exposed to the solvent environment. These differences are reinforced in the view of the cystine core face, although the cystine knot core itself is largely the same between the peptides.

Similarly, the edge where the cationic residues are resident, several of the previous observations are also further actualised, in particular with the corresponding cationic residues lying in different positions. However, the observation by Tam *et al.*, that the cationic residues in circulin B are more exposed and aligned continuously along one edge in circulin B is not upheld. In fact, it is in circulin A that polar and cationic residues form a continuous line across the lower section of this edge. The cationic K22 and R26 residues at each end of the line, respectively, are turned upward rather like a smile. The line is capped by a cluster of hydrophobic residues, involving residues V23, Y25, G28, I29 and P30. Below the line, a small portion of the cystine core emerges and also more hydrophobic residues. In circulin B, the line is more straight and is interrupted by the hydrophobic cap, which extends across the line and is no longer clustered as in circulin A. The position of K22 almost entirely obscures the suspended aromatic residue F7 from view, unlike in circulin A, where the W7 residue is virtually completely in full view due to

K22 turning upward. Circulin B also has an additional polar residue, T13, partially exposed on lowest edge of this face originating from the  $\alpha$ -helix which in circulin A is the hydrophobic G16 residue that is shifted to the left.

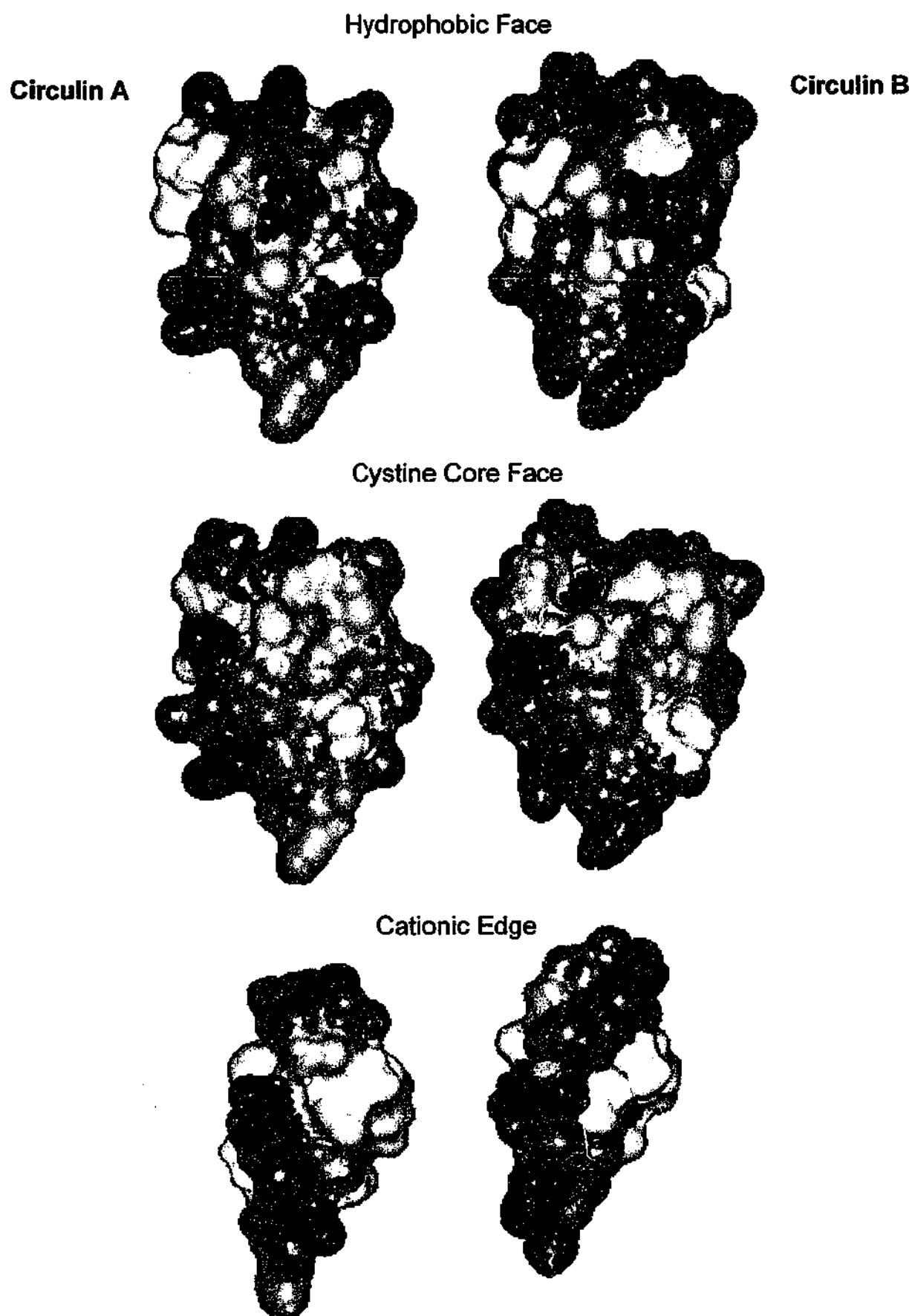
Tam *et al.* also suggest that the arginine residue is folded back towards the backbone and is shielded from the tryptophan residue. This does not appear to be the case, as the arginine residue is definitely protruding from the peptide core in circulin A and is at the opposite end of the cationic edge, so that shielding of the arginine residue by tryptophan is not relevant. Lastly, it was considered of interest to ascertain whether these structural changes affect the degree of hydrophobicity of any of these surfaces. The Connolly surfaces were generated for the surfaces analysed above. These surfaces are displayed for the same orientations of the circulins in Figure 7.14 according to the property of hydrophobicity. In the diagram, the flow of colour from blue through white to pink to red represents increasing hydrophobicity.

Interestingly, the surfaces are generally more hydrophobic for circulin B.

It may be speculated that these differences in the hydrophobic, polar and charge distribution serve to alter the chemical recognition and docking process necessary for interaction with membrane surfaces where the positions of the reactive centres vary for specific microbes. In other words, the lock and key mechanism of interaction is finely tuned and highly specific. This suggests that the CCK fold, probably being part of the plant defense system, has evolved to deal with the threat of hostile biological organisms in a unique manner targeted at each specific microbe. This highlights the prime importance of the CCK fold in a general sense. The same basic fold with minor modification may be used to combat a diverse range of invading microbial species.

Overall, it seems that it is the combination of three residue substitutions and one residue insertion in circulin B that alters the local backbone fold slightly, many of the residue side-chain orientations and hence the surface characteristics. It is not merely the presence of an additional residue, as





**Figure 7.14.** Connolly surface diagrams of the identical hydrophobic faces, cystine core faces and cationic edges for circulin A and circulin B. The surfaces are displayed according to the property of hydrophobicity, with the flow of colour from blue through white to pink to red representing increasing hydrophobic character.

suggested by Tam *et al.* The study and comparison of the circulins support subtle structural changes noted throughout the structural determination and analysis. While the global fold is essentially conserved between the circulins, minute structural differences manifested as surface differences appear to significantly alter specific biological activities. This finding illustrates the potential for circulins in drug design in that they form a highly stable, robust template in which minimal residue alternation can have marked effects on specificity of microbial, fungal and cellular activity. An identification and understanding of the key residues would potentially allow the development of antimicrobial medications where side effects may be controlled and minimised.

### 7.5 Summary

The three dimensional solution structure of circulin B, a polypeptide with 31 residues extracted from the African plant *Chassalia parvifolia*, has been determined using two dimensional  $^1\text{H}$ -NMR spectroscopy. Circulin B, together with circulin A, was originally identified based upon its inhibition of the cytopathic effects and the replication of the HIV virus (Gustafson *et al.*, 1994). Structural restraints consisting of 600 interproton distances inferred from nuclear Overhauser effects, 17 backbone dihedral and 11  $\chi_1$  angle restraints from spin-spin coupling constants and 14 distance restraints ascertained from likely hydrogen bonds were used as input for simulated annealing calculations and energy minimisation in the program X-PLOR. The final set of 12 structures have mean pairwise rms differences over the whole molecule of 0.30 Å for the backbone atoms, and 0.83 Å for all heavy atoms.

The overall fold of circulin B is very similar to that of circulin A and other CCK peptides. Circulin B consists of several  $\beta$ -turns, which form a distorted triple-stranded anti-parallel  $\beta$ -sheet. There are three disulphide bonds, which produce the knotted cystine core where two of the disulphide bonds form a ring and the third disulphide bond is threaded through the centre of the ring. In addition, several features between circulin A and B are conserved.

This includes a  $3_{10}$   $\alpha$ -helix in which the same sequential residues are involved, a non identical aromatic residue which protrudes into the solvent environment, a second identical aromatic residue located within the inner core of the peptides, molecular anisotropy whereby the peptide molecules are a flattened shape in which the two faces are significantly larger than the edge and lastly, hydrophobic and charge residual distribution characteristic of cationic peptides where the hydrophobic residues are concentrated on one face and the cationic residues are contained along a peripheral edge.

Despite the similarity of fold and sequence, a separate study found substantial differences in microbial, fungal and cellular activity between circulin A and B (Tam *et al.*, 1999b). A detailed comparison of the three dimensional structures and their surfaces was carried out. Subtle structural and surface differences were noted and appear to be related to the three residual substitutions and the additional residue in circulin B. The backbone topology is slightly altered foremost in the regions of residual differences. As a result, there are several significant topological consequences.

The  $3_{10}$   $\alpha$ -helix occupies a different position and may be more stable in circulin B. The loop containing the extra residue in circulin B is associated with additional  $\beta$ -strand character both within and adjacent to the loop. The spatial orientations of the important cationic residues are different, especially those of K22 and R26. Exposure of the protruding non identical aromatic residue is different, being hidden for circulin B in the view of the cationic edge and completely exposed for circulin A. This reorientation affects the hydrogen bonding of the peptide core in the region. The distribution of hydrophobic and cationic residues is different, resulting in slight differences of surface area, circulin B being larger on the faces and circulin A on the cationic edge. In addition, circulin B in general displays greater hydrophobic character. Furthermore, along the cationic edge there is a continuous line of cationic and polar residues capped by a hydrophobic cluster in circulin A. In circulin B, the line is more straight and is interrupted by the hydrophobic cluster. However, the cystine core itself is largely maintained between circulin A and B.

The comparison of the 3D structures and surfaces of the circulin peptides in relation to activity indicates the strategic importance of the CCK fold. The fold is evidently robust in that the inner cystine knot core is maintained. The same, basic, thematic fold could be used in directed biological defense where by minor residual modification results in significant structural and surface changes that enable more specific pathogens to be destroyed. In this sense, the essential CCK framework may be used to resist a range of microorganisms. The CCK peptides therefore offer scope for the development of target specific antimicrobial drugs.

## **CHAPTER EIGHT**

## **CONCLUSIONS**

Extensive structural studies on a novel class of macrocyclic plant peptides, known as the cyclic cystine knot family, have been conducted for this thesis. The results have broadened the perception of the 3D nature of members of this family and shed light on structure/activity relationships, demonstrating their potential for future drug design.

Attention was initially directed towards the first discovered member of the family, kalata-B1, for which the 3D solution structure, sequential residual composition and pharmacological properties were already known. A set of six fragments was designed to study the effect of the absence of each section on structure and activity. The members of the set were referred to as acyclic permutants, as each comprised an acyclic peptide in which a selected region of the backbone was removed. The study of acyclic permutants is a novel approach for the study of folding and structure/activity relationships.

Potential folding/unfolding pathways of cyclic peptides and their acyclic permutants were explored using the 'kinemage' computer graphics tool. The set of acyclic permutants was used to create animations, in which the polypeptide backbones were unthreaded by a series of kinemages to produce the 2D planar arrangements. The 2D planar arrangements were compared to that of PCI, a non-cyclic member of the inhibitor cystine knot family of which cyclic cystine knots are a sub-classification. The 2D planar arrangement of the acyclic permutant in which the residues of the disordered region of kalata-B1 are absent matches that of PCI. This comparison provided a topological basis for suggesting that the location of the starting residue in plant synthesis may be within the disordered region.

The unthreading of the backbones of the acyclic permutants illustrated the simplicity of unthreading in the absence of the embedded ring of inhibitor cystine knot peptides. If unthreading may be equated to unfolding, an explanation may be available in these animations, for the findings of experimental folding studies, where if there is a break in the embedded ring of the cystine knot, the native fold of kalata-B1 is not achieved. The animations show that the presence of the ring restricts the unthreading process. In the

absence of the ring, there are more pathways to unthreading, and these simply involve a collapsing of the backbone 3D structures.

An animation where a series of kinemages show the formation of the 2D reduced of kalata-B1 was also created. The resultant 2D reduced graph proves that kalata-B1 and cyclic cystine knot peptides are not topologically simple. They are in fact nonplanar, chiral and have D chirality like other polypeptides with four disulphide bonds.

The structure of kalata-B1 in aqueous TFE and in D<sub>2</sub>O was studied using 2D homonuclear and heteronuclear NMR spectroscopy. The <sup>13</sup>C chemical shifts for kalata-B1 were derived from HMQC and HMQC-TOCSY data. The <sup>13</sup>C chemical shifts were not particularly useful for predicting the location of secondary structure or structural changes due to a TFE solvent environment. However, <sup>13</sup>C chemical shift data provided unequivocal evidence for the conformations of the three proline residues, including the less common *cis* arrangement for one of these. The 3D structure of kalata-B1 in 20% TFE was determined from interproton distance data from nuclear Overhauser spectra. The structure of kalata-B1 in the TFE environment is largely maintained relative to that in water, indicating the stable, resilient nature of the cyclic cystine knot structure. The structural changes in TFE are minor. The disulphide core appears to be slightly tighter and there is a net increase in the hydrophobic surface area, resulting in some loss of classical secondary structure definition.

Structural studies on kalata-B1 in TFE were followed by similar studies on one of the key acyclic permutants in which residues of the disordered region are omitted. This particular acyclic permutant was of interest, as it unthreads to form the same 2D planar arrangement as that of non-cyclic cystine knot peptides and it had been speculated that the missing section is likely to be involved in receptor binding. The linear reduced acyclic permutant was oxidised in 0.1 M NH<sub>4</sub>NCO<sub>3</sub> at pH of 8 and a mixture of species was obtained. The predominant species was isolated and collected using RP-HPLC in a water/acetonitrile/trifluoroacetic acid buffer system. The predominant

component was fully oxidised and was the most hydrophilic fraction of the mixture. This fraction consisted of two distinct species, which are the product of *cis/trans* isomerisation of the proline residue identified as the *cis* proline in kalata-B1.

The proton chemical shifts were assigned for both isomers. From parameters measured from the NMR data, including amide and alpha proton chemical shifts, CSI values and mid to long range NOE connectivities, the structures of the isomers were inferred. The isomers are similar in structure. However, they are significantly different from the parent kalata-B1. The isomers are comprised of a centrally located  $\beta$ -strand, with several turns on either side. The only common region between the isomers and kalata-B1 is a section beside the  $\beta$ -strand, on the *N*-terminus side of the isomer backbone and the connecting region of kalata-B1. In this region, the NOE connectivities are similar between the isomers and kalata-B1. From this similarity and NOE connectivities between disulphide residues, a disulphide bonding pattern was deduced, in which the native disulphide bond Cys(2)-Cys(5) is retained, and two non-native disulphide bonds, Cys(1)-Cys(3) and Cys(4)-Cys(6), are formed. This particular disulphide pattern is consistent with all of the structural observations in the isomers.

Finally, 2D NMR spectroscopy was applied to two new macrocyclic cystine knot peptides, circulin A and circulin B, which have anti-HIV and anti-microbial activities. The circulin peptides belong to the bracelet like sub-classification of cyclic cystine knot peptides. Using NMR derived interproton distances and angle restraints in simulated annealing calculations, the 3D solution structures were determined. The structures of circulin A and B are very similar to that of kalata-B1, which is the prototypic member of the Moebius strip sub-classification of cyclic cystine knot peptides. Like kalata-B1, the circulins consist mainly of  $\beta$ -turns and  $\beta$ -strands, forming a distorted triple-stranded  $\beta$ -sheet, braced with three disulphide bonds linked in the pattern characteristic of the inhibitor cystine knot motif. Furthermore, the spatial location of the protruding aromatic residue is retained, as well as, the



distribution of hydrophobic residues which are concentrated on one face, shielding the cystine rich core on this face.

However, the circulins possess an additional short section of  $3_{10}$   $\alpha$ -helix in the region corresponding to the disordered region of kalata-B1. This has been attributed to the increased number of residues in the region for the circulins compared to kalata-B1. The circulins also have three cationic residues, as opposed to one in kalata-B1. The cationic residues are located along one peripheral edge of the circulin molecules. Overall, the distribution of cationic and hydrophobic residues is reminiscent of known cationic antibactericidal peptides.

A detailed comparison of the structures of circulins was carried out. It was noted that certain features are conserved. These include the  $3_{10}$   $\alpha$ -helix over the same sequential residual positions, a non-conserved aromatic residue which extends conspicuously into the solvent environment, a second, conserved residue located within the inner core of the molecular structures, and molecular anisotropy. The general shape is flattened so that there are two larger faces and a thinner edge. On one of the larger faces, the cystine core is protected by a concentration of hydrophobic residues, while on the other face the cystine core is exposed.

Despite the similarities between circulin A and circulin B, there are poignant superficial differences, which are related to the three different residues and an additional residue in circulin B. There are several implications of these differences. The side-chains of the cationic residues, which are likely to play an important role in recognition, adopt different spatial orientations, especially K22 and R26. In the view of the cationic edge, the protruding non-identical aromatic is completely exposed for circulin A and largely hidden in circulin B. The locations of polar and hydrophobic residues on the hydrophobic face, and along the cationic edge are notably altered. In general, circulin B occupies a larger surface area on the larger faces, while circulin A occupies a larger surface area on the cationic edge. Circulin B also has slightly greater

hydrophobic character. Nonetheless, the face exposing the cystine core is very similar, illustrating conservation of the cystine core structure.

These observed structural and surface differences could explain the considerable activity differences towards a range of common pathogens and fungi. Although, the circulins are very similar in overall structure, residual content and distribution, they exhibit specific and complementary activities. The study of the structures of circulin A and circulin B highlights the potential of cystine knot peptides whereby the stable cystine knot acts as a robust template for activity from which subtle modification of certain residues can dramatically alter activity through structural and surface changes. Through an understanding of which residues elicit certain responses, cyclic cystine knot peptides could one day be used for clinical applications. The work for this thesis has contributed to and increased knowledge of the cyclic cystine knot family.

**PUBLICATIONS AND  
COMMUNICATIONS**

## **PUBLICATIONS ASSOCIATED WITH THIS THESIS**

Daly, N. D., Koltay, A., Gustafson, K. R., Boyd, M. R., Casas-Finet, J. R., Craik, D. J. (1999) Solution structure by NMR of circulin A: a macrocyclic knotted peptide having anti-HIV activity. *J. Mol. Biol.* **285**, 333-345.

## **CONFERENCE PRESENTATIONS**

Structural Refinement of a Uterotonic Peptide ISMAR 95 (International NMR Conference), Sydney, Australia, 1995.

## **APPENDICES**

## APPENDIX I

**Table A1.1.** Proton chemical shifts (ppm) for kalata-B1 in 6% TFE at 303 K. Shifts are measured at 500 MHz and referenced to water calibrated against 2,2-dimethyl-silapentane-5-sulfonate (DDS).

	RESIDUE	NH	$\alpha$ H	$\beta$ H	OTHER	
1.	SER	8.97	4.77	3.91, 3.86		
2.	TRP	7.69	4.05	3.22		
3.	PRO		3.43	1.68, -0.34	$\gamma$ CH <sub>2</sub> 1.33, 1.23	$\delta$ CH <sub>2</sub> 3.21
4.	VAL	8.36	4.19	1.88		$\gamma$ CH <sub>3</sub> 0.85, 0.81
5.	CYS	7.63	5.11	3.27, 2.78		
6.	THR	9.94	5.09	3.73		$\gamma$ CH <sub>3</sub> 0.89
7.	ARG	8.70	4.77	1.69	$\gamma$ CH <sub>2</sub> 1.47	$\delta$ CH <sub>2</sub> 3.16, NH 6.93
8.	ASN	9.45	4.36	3.10, 2.85		NH <sub>2</sub> 6.85, 7.46
9.	GLY	8.70	4.24, 3.56			
10.	LEU	7.77	5.07	1.72, 1.40	$\gamma$ CH 1.99	$\delta$ CH <sub>2</sub> 0.98, 0.91
11.	PRO		5.08	2.47, 1.67	$\gamma$ CH <sub>2</sub> 2.15, 2.07	$\delta$ CH <sub>2</sub> 3.79, 3.69
12.	VAL	8.12	4.63	2.59		$\gamma$ CH <sub>3</sub> 0.89, 0.84
13.	CYS	7.81	4.47	3.38, 3.03		
14.	GLY	8.43	3.85, 3.73			
15.	GLU	7.13	4.77	1.90	$\gamma$ CH <sub>2</sub> 2.63, 2.47	
16.	THR	8.40	4.58	4.48		$\gamma$ CH <sub>3</sub> 1.12
17.	CYS	8.53	5.08	3.26, 2.76		
18.	VAL	8.52	3.77	2.06		$\gamma$ CH <sub>3</sub> 1.07, 0.98
19.	GLY	8.46	4.37, 3.81			
20.	GLY	8.21	4.49, 4.07			
21.	THR	7.79	4.70	4.07		$\gamma$ CH <sub>3</sub> 1.14
22.	CYS	8.60	4.69	3.03, 2.77		
23.	ASN	10.23	4.72	2.79		NH <sub>2</sub> 6.85, 7.61
24.	THR	8.81	4.42	4.23		$\gamma$ CH <sub>3</sub> 1.34
25.	PRO		4.26	2.32, 1.91	$\gamma$ CH <sub>2</sub> 2.15, 2.02	$\delta$ CH <sub>2</sub> 4.17, 3.71
26.	GLY	8.72	4.16, 3.68			
27.	CYS	7.68	5.35	3.82, 2.66		
28.	THR	9.50	4.55	4.04		$\gamma$ CH <sub>3</sub> 1.13
29.	CYS	8.86	4.64	3.10, 2.80		

**Table A1.2.** Proton chemical shifts (ppm) for kalata-B1 in 20% TFE at at 280 K. Shifts are measured at 500 MHz and referenced to water calibrated against DDS.

	RESIDUE	NH	$\alpha$ H	$\beta$ H	OTHER	
1.	SER	9.18	4.98	4.17, 4.07		
2.	TRP	-	4.23	3.48, 3.40		
3.	PRO		3.64	1.90, - 0.17	$\gamma$ CH <sub>2</sub> 1.52, 1.44	$\delta$ CH <sub>2</sub> 3.49, 3.40
4.	VAL	8.72	3.94	2.27		$\gamma$ CH <sub>3</sub> 1.29, 1.21
5.	CYS	7.82	5.33	3.51, 3.02		
6.	THR	10.17	5.31	3.96		$\gamma$ CH <sub>3</sub> -
7.	ARG	8.93	5.99	1.92	$\gamma$ CH <sub>2</sub> 1.69	$\delta$ CH <sub>2</sub> 3.39, 3.33 NH 7.10
8.	ASN	9.58	4.55	3.34, 3.06		
9.	GLY	8.87	4.46, 3.77			
10.	LEU	8.00	5.29	2.26, 1.56	$\gamma$ CH 1.96	$\delta$ CH <sub>2</sub> 1.19, 1.12
11.	PRO		5.29	2.69, 1.84	$\gamma$ CH <sub>2</sub> 2.34, 2.21	$\delta$ CH <sub>2</sub> 3.98, 3.81
12.	VAL	8.31	4.82	2.80		$\gamma$ CH <sub>3</sub> 1.09, 1.04
13.	CYS	7.91	4.69	3.59, 3.25		
14.	GLY	8.55	4.02, 3.90			
15.	GLU	7.33	4.98	2.12	$\gamma$ CH <sub>2</sub> 2.90, 2.70	
16.	THR	8.55	4.81	4.61		$\gamma$ CH <sub>3</sub> 1.34
17.	CYS	8.82	5.33	3.52, 2.92		
18.	VAL	8.61	4.39	2.09		$\gamma$ CH <sub>3</sub> 1.06, 1.01
19.	GLY	-	-			
20.	GLY	8.43	4.72, 4.30			
21.	THR	7.97	4.93	4.27		$\gamma$ CH <sub>3</sub> 1.35
22.	CYS	8.80	4.90	3.25, 3.00		
23.	ASN	10.49	4.95	3.00		
24.	THR	8.99	4.63	4.46		$\gamma$ CH <sub>3</sub> 1.56
25.	PRO		4.47	2.50, 2.10	$\gamma$ CH <sub>2</sub> 2.33, 2.21	$\delta$ CH <sub>2</sub> 4.39, 3.91
26.	GLY	8.90	4.35, 3.88			
27.	CYS	7.88	5.58	4.05, 2.89		
28.	THR	9.73	4.77	4.27		$\gamma$ CH <sub>3</sub> 1.36
29.	CYS	9.04	4.89	3.33, 3.01		

**Table A1.3.** Proton chemical shifts (ppm) for kalata-B1 in 20% TFE at at 300 K. Shifts are measured at 750 MHz and referenced to water calibrated against DDS.

	RESIDUE	NH	$\alpha$ H	$\beta$ H	OTHER	
1.	SER	9.03	4.83	4.02, 3.91		
2.	TRP	7.54	4.10	3.35, 3.26		
3.	PRO		4.10	3.34, - 0.3	$\gamma$ CH <sub>2</sub> 3.24, 1.75	$\delta$ CH <sub>2</sub> 3.79, 3.49
4.	VAL	8.45	4.24	1.94		$\gamma$ CH <sub>3</sub> 0.91, 0.87
5.	CYS	7.67	5.18	3.36, 2.87		
6.	THR	10.02	5.18	3.81		$\gamma$ CH <sub>3</sub> 0.97
7.	ARG	8.78	4.85	1.77	$\gamma$ CH <sub>2</sub> 1.56	$\delta$ CH <sub>2</sub> 3.25, 3.15 NH 6.95
8.	ASN	9.42	4.41	3.19, 2.90		NH <sub>2</sub> 6.83, 7.38
9.	GLY	8.70	4.31, 3.61			
10.	LEU	7.86	5.14	2.10, 1.40	$\gamma$ CH 1.81	$\delta$ CH <sub>2</sub> 1.04, 0.97
11.	PRO		5.14	2.56, 1.70	$\gamma$ CH <sub>2</sub> 2.20, 2.12	$\delta$ CH <sub>2</sub> 3.86, 3.68
12.	VAL	8.18	4.66	2.65		$\gamma$ CH <sub>3</sub> 0.94, 0.90
13.	CYS	7.75	4.53	3.44, 3.10		
14.	GLY	8.40	3.92, 3.74			
15.	GLU	7.18	4.83	1.95	$\gamma$ CH <sub>2</sub> 2.73, 2.52	
16.	THR	8.40	4.66	4.55		$\gamma$ CH <sub>3</sub> 1.18
17.	CYS	8.67	5.19	3.36, 2.78		
18.	VAL	8.55	3.79	2.12		$\gamma$ CH <sub>3</sub> 1.15, 1.05
19.	GLY	8.40	4.46, 3.85			
20.	GLY	8.28	4.53, 4.14			
21.	THR	7.81	4.77	4.13		$\gamma$ CH <sub>3</sub> 1.21
22.	CYS	8.61	4.78	3.10, 2.85		
23.	ASN	10.53	4.73	2.85		NH <sub>2</sub> 6.86, 7.64
24.	THR	8.81	4.47	4.30		$\gamma$ CH <sub>3</sub> 1.40
25.	PRO		4.32	2.36, 1.96	$\gamma$ CH <sub>2</sub> 2.20, 2.07	$\delta$ CH <sub>2</sub> 4.24, 3.76
26.	GLY	8.72	4.19, 3.73			
27.	CYS	7.73	5.42	3.90, 2.75		
28.	THR	9.58	4.61	4.12		$\gamma$ CH <sub>3</sub> 1.21
29.	CYS	8.88	4.74	3.17, 2.86		



**Table A1.4.** Proton chemical shifts (ppm) for kalata-B1 in 20% TFE at 333 K. Shifts are measured at 500 MHz and referenced to water calibrated against DDS.

	RESIDUE	NH	$\alpha$ H	$\beta$ H	OTHER	
1.	SER	9.05	4.95	4.07		
2.	TRP	-	4.25	3.49, 3.39		
3.	PRO		3.66	1.86, 0.037	$\gamma$ CH <sub>2</sub> 1.45, 1.49	$\delta$ CH <sub>2</sub> 3.50, 3.38
4.	VAL	8.28	4.02	2.24		$\gamma$ CH <sub>3</sub> 1.24, 1.15
5.	CYS	7.69	5.34	3.46, 3.01		
6.	THR	9.97	5.26	3.94		$\gamma$ CH <sub>3</sub> 1.11
7.	ARG	8.82	5.00	1.91	$\gamma$ CH <sub>2</sub> 1.72	$\delta$ CH <sub>2</sub> 3.33, NH 6.99
8.	ASN	9.32	4.50	3.33, 3.01		
9.	GLY	8.51	4.45, 3.74			
10.	LEU	7.92	5.23	2.22, 1.56	$\gamma$ CH 1.94	$\delta$ CH <sub>2</sub> 1.11
11.	PRO		5.24	2.69, 1.84	$\gamma$ CH <sub>2</sub> 2.29	$\delta$ CH <sub>2</sub> 4.04, 3.79
12.	VAL	8.19	4.76	2.73		$\gamma$ CH <sub>3</sub> 1.05
13.	CYS	7.65	4.65	3.55, 3.24		
14.	GLY	8.37	4.08, 3.88			
15.	GLU	7.25	4.97	2.23, 2.10	$\gamma$ CH <sub>2</sub> 2.74, 2.64	
16.	THR	8.28	4.81	4.65		$\gamma$ CH <sub>3</sub> 1.31
17.	CYS	8.59	5.26	3.51, 3.05		
18.	VAL	8.50	4.45	2.09		$\gamma$ CH <sub>3</sub> 1.02
19.	GLY	8.42	4.51, 4.00			
20.	GLY	8.19	4.64, 4.21			
21.	THR	7.75	4.88	4.28		$\gamma$ CH <sub>3</sub> 1.34
22.	CYS	8.43	4.91	3.26, 3.03		
23.	ASN	10.04	4.92	2.93		
24.	THR	8.66	4.65	4.40		$\gamma$ CH <sub>3</sub> 1.51
25.	PRO		4.44	2.48, 2.08	$\gamma$ CH <sub>2</sub> 2.31, 2.19	$\delta$ CH <sub>2</sub> 4.31, 3.89
26.	GLY	8.61	4.33, 3.85			
27.	CYS	7.79	5.49	3.99, 2.91		
28.	THR	9.62	4.76	4.26		$\gamma$ CH <sub>3</sub> 1.35
29.	CYS	8.85	4.86	3.34, 3.04		

## APPENDIX II

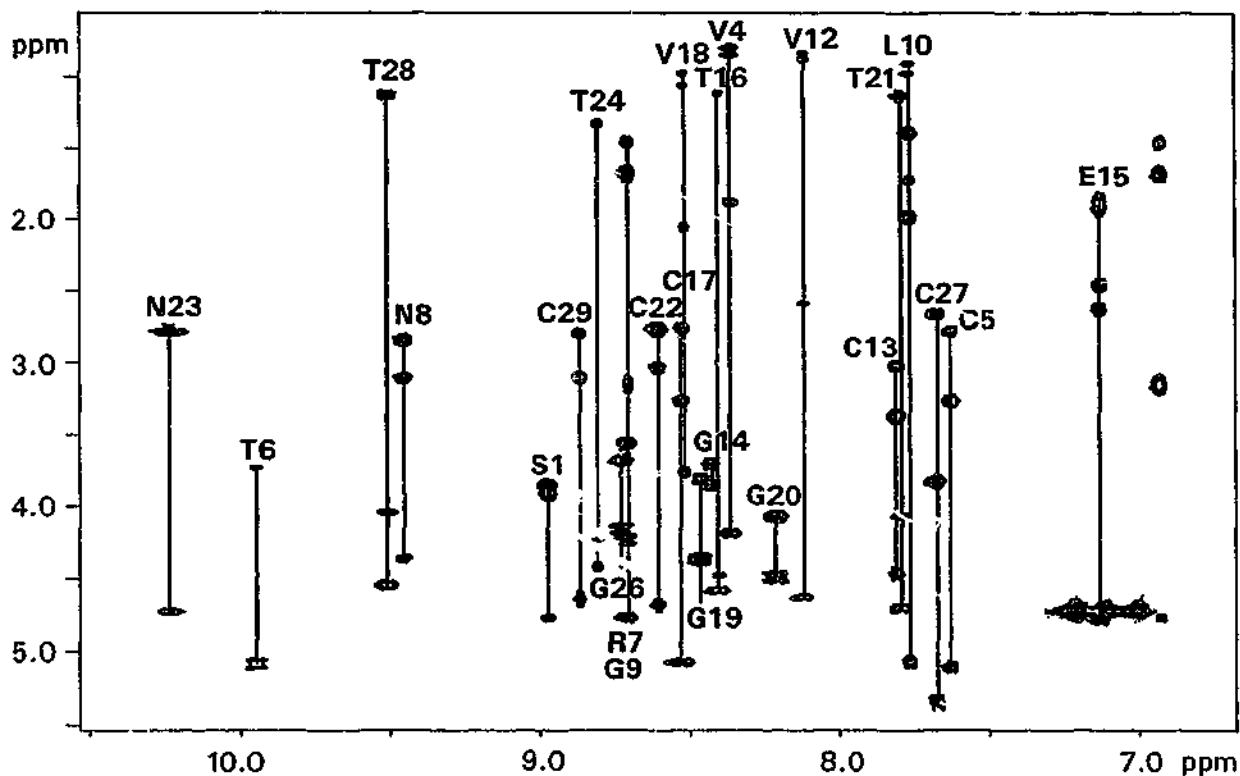


Figure AII.1. TOCSY fingerprint region showing assignments of side chain spin systems linked with their corresponding amide protons for kalata-B1 in 6% TFE at 303 K.

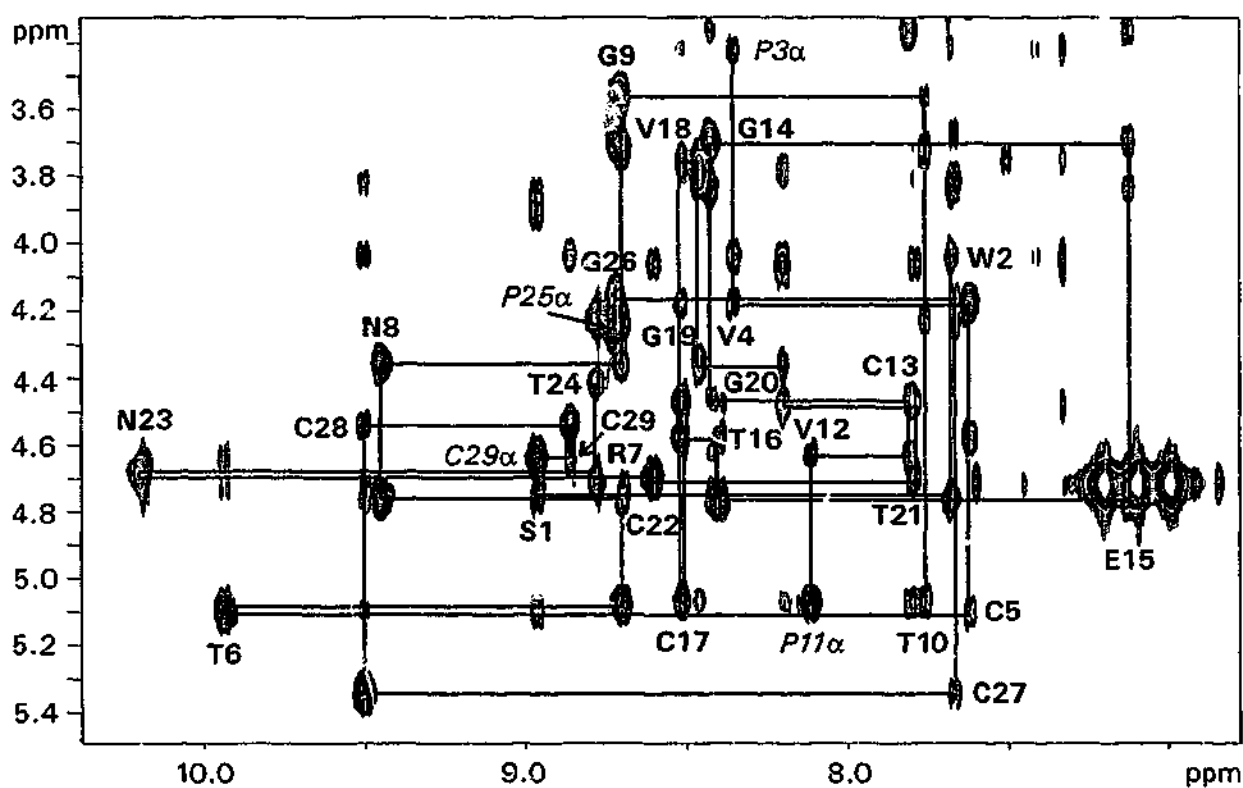


Figure AII.2. NOESY fingerprint region showing sequential connectivity for kalata-B1 in 6% TFE at 303 K.

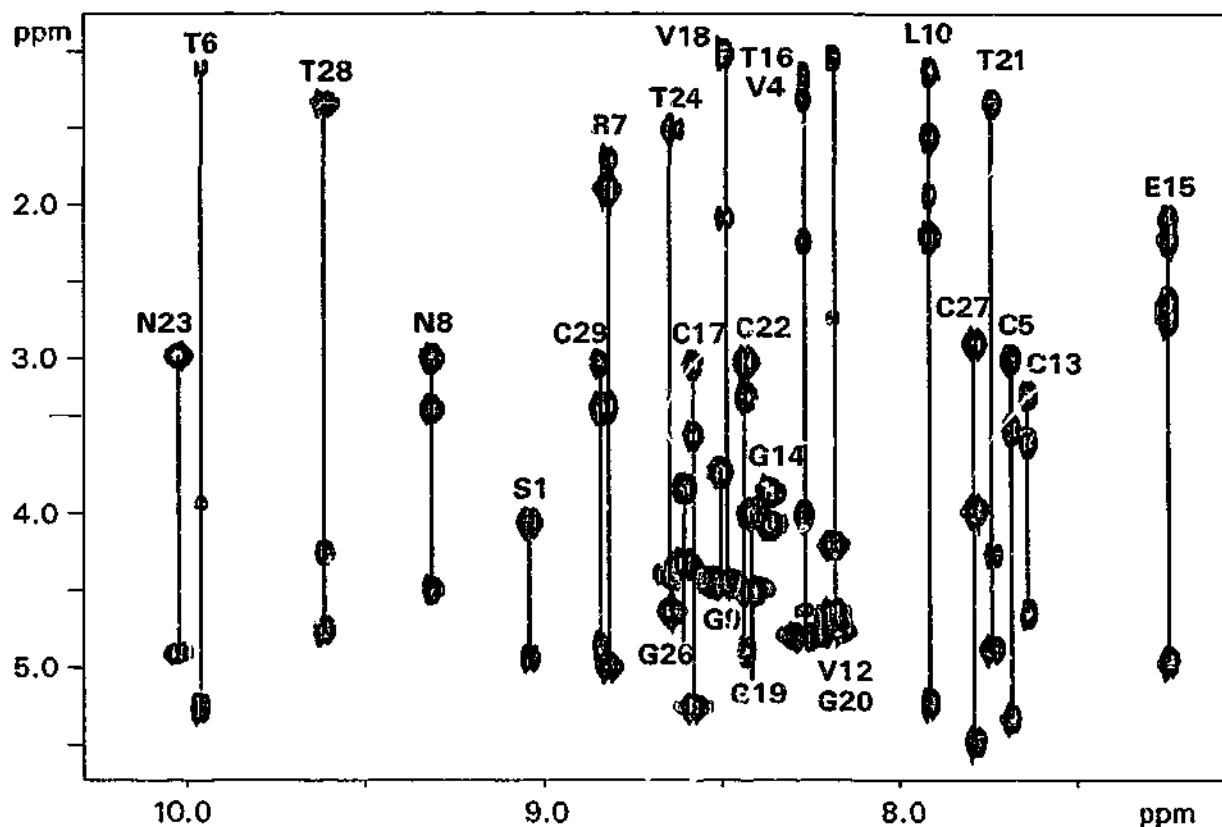


Figure AII.3. TOCSY fingerprint region showing assignments of side chain spin systems linked with their corresponding amide protons for kalata-B1 in 20% TFE at 333 K.

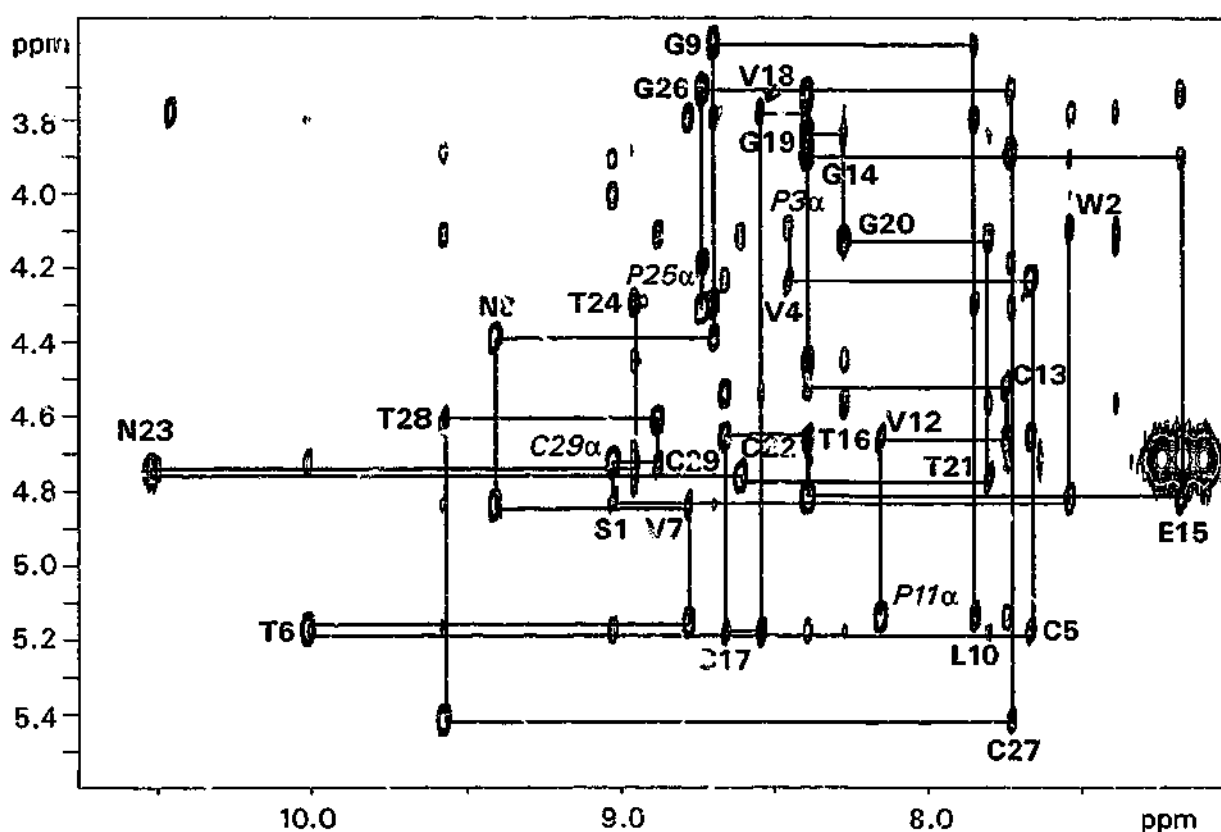


Figure AII.4. NOESY fingerprint region showing sequential connectivity for kalata-B1 in 20% TFE at 300 K.

### APPENDIX III

Table of distance restraints used in the simulated annealing protocol for the determination of the structure of kalata-B1 in 20% TFE (Chapter 4). The lower bound of the distance restraint given in the second distance column is not specified, and is, for the purposes of the calculations determined from the Van der Waal's forces. The upper bound distance values, as determined from the relative intensities of the NOE cross peaks and pseudoatom adjusted where relevant, are given in the third distance column.

	d <sub>-</sub>	d <sub>+</sub>
assign (resid 1 and name HN) (resid 1 and name HA)	0 0	3.5
assign (resid 1 and name HN) (resid 1 and name HB1)	0 0	3.5
assign (resid 1 and name HN) (resid 1 and name HB2)	0 0	3.5
assign (resid 1 and name HA) (resid 1 and name HB*)	0 0	5
assign (resid 1 and name HB1) (resid 1 and name HB2)	0 0	3.5
assign (resid 1 and name HN) (resid 4 and name HN)	0 0	5
assign (resid 1 and name HN) (resid 4 and name HG*)	0 0	9
assign (resid 1 and name HB*) (resid 4 and name HN)	0 0	7
assign (resid 1 and name HB*) (resid 4 and name HG1*)	0 0	8.5
assign (resid 1 and name HB*) (resid 4 and name HG2*)	0 0	8.5
assign (resid 1 and name HN) (resid 5 and name HA)	0 0	3.5
assign (resid 1 and name HN) (resid 6 and name HN)	0 0	3.5
assign (resid 1 and name HN) (resid 6 and name HG2*)	0 0	3.5
assign (resid 1 and name HN) (resid 29 and name HA)	0 0	2.7
assign (resid 1 and name HN) (resid 29 and name HB1)	0 0	5
assign (resid 1 and name HN) (resid 29 and name HB2)	0 0	5
assign (resid 2 and name HE1) (resid 2 and name HB*)	0 0	7
assign (resid 2 and name HE1) (resid 3 and name HD1)	0 0	6
assign (resid 2 and name HE1) (resid 3 and name HD2)	0 0	6
assign (resid 2 and name HE1) (resid 18 and name HA)	0 0	5
assign (resid 2 and name HE1) (resid 18 and name HB)	0 0	5
assign (resid 2 and name HE1) (resid 18 and name HG1*)	0 0	6.5
assign (resid 2 and name HE1) (resid 18 and name HG2*)	0 0	6.5
assign (resid 2 and name HE1) (resid 20 and name HN)	0 0	5
assign (resid 3 and name HA) (resid 3 and name HB1)	0 0	5
assign (resid 3 and name HA) (resid 3 and name HB2)	0 0	5
assign (resid 3 and name HB1) (resid 3 and name HB2)	0 0	3.5
assign (resid 3 and name HB*) (resid 3 and name HG1)	0 0	6
assign (resid 3 and name HB*) (resid 3 and name HG2)	0 0	6
assign (resid 3 and name HB*) (resid 3 and name HD1)	0 0	6
assign (resid 3 and name HB*) (resid 3 and name HD2)	0 0	6
assign (resid 3 and name HG*) (resid 3 and name HD1)	0 0	4.5
assign (resid 3 and name HG*) (resid 3 and name HD2)	0 0	4.5
assign (resid 3 and name HD1) (resid 3 and name HD2)	0 0	5
assign (resid 3 and name HG1) (resid 4 and name HN)	0 0	6
assign (resid 3 and name HG2) (resid 4 and name HN)	0 0	6
assign (resid 3 and name HD*) (resid 4 and name HN)	0 0	6
assign (resid 3 and name HB*) (resid 5 and name HB*)	0 0	7
assign (resid 3 and name HG*) (resid 16 and name HB)	0 0	6
assign (resid 3 and name HG*) (resid 17 and name HN)	0 0	6
assign (resid 3 and name HD1) (resid 17 and name HN)	0 0	6
assign (resid 3 and name HD2) (resid 17 and name HN)	0 0	6
assign (resid 4 and name HN) (resid 4 and name HA)	0 0	3.5

assign (resid 4 and name HN) (resid 4 and name HB)	0 0 2.7
assign (resid 4 and name HN) (resid 4 and name HG*)	0 0 5.7
assign (resid 4 and name HB) (resid 4 and name HG1*)	0 0 5
assign (resid 4 and name HB) (resid 4 and name HG2*)	0 0 5
assign (resid 4 and name HN) (resid 5 and name HN)	0 0 5
assign (resid 4 and name HA) (resid 5 and name HN)	0 0 2.7
assign (resid 4 and name HB) (resid 5 and name HN)	0 0 5
assign (resid 4 and name HG*) (resid 11 and name HG*)	0 0 9
assign (resid 4 and name HA) (resid 17 and name HN)	0 0 3.5
assign (resid 5 and name HN) (resid 5 and name HA)	0 0 3.5
assign (resid 5 and name HN) (resid 5 and name HB1)	0 0 3.5
assign (resid 5 and name HN) (resid 5 and name HB2)	0 0 3.5
assign (resid 5 and name HA) (resid 5 and name HB*)	0 0 3.7
assign (resid 5 and name HB1) (resid 5 and name HB2)	0 0 2.7
assign (resid 5 and name HN) (resid 6 and name HN)	0 0 5
assign (resid 5 and name HB1) (resid 6 and name HN)	0 0 5
assign (resid 5 and name HB2) (resid 6 and name HN)	0 0 5
assign (resid 5 and name HN) (resid 15 and name HN)	0 0 6
assign (resid 5 and name HB1) (resid 15 and name HN)	0 0 6.0
assign (resid 5 and name HB2) (resid 15 and name HN)	0 0 6.0
assign (resid 5 and name HN) (resid 16 and name HA)	0 0 2.7
assign (resid 5 and name HN) (resid 16 and name HG2*)	0 0 6.5
assign (resid 5 and name HN) (resid 17 and name HN)	0 0 5
assign (resid 5 and name HA) (resid 28 and name HN)	0 0 5
assign (resid 5 and name HB*) (resid 28 and name HN)	0 0 6
assign (resid 6 and name HN) (resid 6 and name HB)	0 0 5
assign (resid 6 and name HA) (resid 6 and name HG*)	0 0 5
assign (resid 6 and name HN) (resid 7 and name HN)	0 0 5
assign (resid 6 and name HN) (resid 7 and name HA)	0 0 6
assign (resid 6 and name HB) (resid 7 and name HN)	0 0 2.7
assign (resid 6 and name HG2*) (resid 7 and name HN)	0 0 5
assign (resid 6 and name HB) (resid 9 and name HN)	0 0 3.5
assign (resid 6 and name HG2*) (resid 9 and name HN)	0 0 6.5
assign (resid 6 and name HB) (resid 10 and name HN)	0 0 2.7
assign (resid 6 and name HN) (resid 27 and name HA)	0 0 5
assign (resid 6 and name HN) (resid 28 and name HN)	0 0 2.7
assign (resid 6 and name HN) (resid 29 and name HA)	0 0 3.5
assign (resid 7 and name HN) (resid 7 and name HA)	0 0 3.5
assign (resid 7 and name HN) (resid 7 and name HB*)	0 0 3.7
assign (resid 7 and name HN) (resid 7 and name HG*)	0 0 7
assign (resid 7 and name HA) (resid 7 and name HB*)	0 0 3.7
assign (resid 7 and name HA) (resid 7 and name HG*)	0 0 4.5
assign (resid 7 and name HB*) (resid 7 and name HG*)	0 0 4.7
assign (resid 7 and name HB*) (resid 7 and name HD1)	0 0 3.7
assign (resid 7 and name HB*) (resid 7 and name HD2)	0 0 3.7
assign (resid 7 and name HB*) (resid 7 and name HE)	0 0 3.7
assign (resid 7 and name HG*) (resid 7 and name HD1)	0 0 4.5
assign (resid 7 and name HG*) (resid 7 and name HD2)	0 0 4.5
assign (resid 7 and name HG*) (resid 7 and name HE)	0 0 6
assign (resid 7 and name HD1) (resid 7 and name HE)	0 0 3.5
assign (resid 7 and name HD2) (resid 7 and name HE)	0 0 3.5
assign (resid 7 and name HN) (resid 8 and name HN)	0 0 5
assign (resid 7 and name HN) (resid 8 and name HA)	0 0 2.7
assign (resid 7 and name HB*) (resid 8 and name HN)	0 0 4.5
assign (resid 7 and name HG*) (resid 8 and name HN)	0 0 4.5
assign (resid 7 and name HA) (resid 9 and name HN)	0 0 5

assign (resid 7 and name HN) (resid 10 and name HN)	0 0 2.7
assign (resid 7 and name HN) (resid 10 and name HB*)	0 0 6
assign (resid 7 and name HA) (resid 10 and name HN)	0 0 5
assign (resid 7 and name HN) (resid 12 and name HN)	0 0 3.5
assign (resid 7 and name HG*) (resid 12 and name HD1)	0 0 8
assign (resid 7 and name HG*) (resid 12 and name HD2)	0 0 8
assign (resid 7 and name HG*) (resid 12 and name HE)	0 0 8
assign (resid 7 and name HA) (resid 28 and name HN)	0 0 5
assign (resid 7 and name HB*) (resid 27 and name HA)	0 0 6.0
assign (resid 8 and name HN) (resid 8 and name HA)	0 0 2.7
assign (resid 8 and name HN) (resid 8 and name HB1)	0 0 5
assign (resid 8 and name HN) (resid 8 and name HB2)	0 0 5
assign (resid 8 and name HA) (resid 8 and name HB1)	0 0 3.5
assign (resid 8 and name HA) (resid 8 and name HB2)	0 0 3.5
assign (resid 8 and name HB1) (resid 8 and name HB2)	0 0 2.7
assign (resid 8 and name HB*) (resid 8 and name HD21)	0 0 6.5
assign (resid 8 and name HB*) (resid 8 and name HD22)	0 0 6.5
assign (resid 8 and name HD21) resid 8 and name HD22)	0 0 2.7
assign (resid 8 and name HN) (resid 9 and name HN)	0 0 2.7
assign (resid 8 and name HA) (resid 9 and name HN)	0 0 3.5
assign (resid 8 and name HB*) (resid 9 and name HN)	0 0 7
assign (resid 8 and name HN) (resid 10 and name HN)	0 0 5
assign (resid 8 and name HB1) (resid 10 and name HN)	0 0 6
assign (resid 8 and name HB2) (resid 10 and name HN)	0 0 6
assign (resid 9 and name HN) (resid 9 and name HA1)	0 0 3.5
assign (resid 9 and name HN) (resid 9 and name HA2)	0 0 3.5
assign (resid 9 and name HA1) (resid 9 and name HA2)	0 0 2.7
assign (resid 9 and name HN) (resid 10 and name HN)	0 0 2.7
assign (resid 9 and name HN) (resid 10 and name HG)	0 0 6
assign (resid 9 and name HA1) (resid 10 and name HN)	0 0 3.5
assign (resid 9 and name HA2) (resid 10 and name HN)	0 0 3.5
assign (resid 10 and name HN) (resid 10 and name HA)	0 0 3.5
assign (resid 10 and name HN) (resid 10 and name HG)	0 0 2.7
assign (resid 10 and name HN) (resid 10 and name HD1*)	0 0 6.5
assign (resid 10 and name HN) (resid 10 and name HD2*)	0 0 6.5
assign (resid 10 and name HA) (resid 10 and name HB1)	0 0 3.5
assign (resid 10 and name HA) (resid 10 and name HB2)	0 0 3.5
assign (resid 10 and name HA) (resid 10 and name HD*)	0 0 8
assign (resid 10 and name HB1) (resid 10 and name HB2)	0 0 2.7
assign (resid 10 and name HB1) (resid 10 and name HG)	0 0 3.5
assign (resid 10 and name HB1) (resid 10 and name HD1*)	0 0 5
assign (resid 10 and name HB1) (resid 10 and name HD2*)	0 0 5
assign (resid 10 and name HB2) (resid 10 and name HD1*)	0 0 5
assign (resid 10 and name HB2) (resid 10 and name HD2*)	0 0 5
assign (resid 10 and name HB1) (resid 10 and name HG)	0 0 3.5
assign (resid 10 and name HB2) (resid 10 and name HG)	0 0 3.5
assign (resid 10 and name HG) (resid 10 and name HD1*)	0 0 5
assign (resid 10 and name HG) (resid 10 and name HD2*)	0 0 5
assign (resid 10 and name HB*) (resid 11 and name HD1)	0 0 4.5
assign (resid 10 and name HB*) (resid 11 and name HD2)	0 0 4.5
assign (resid 10 and name HG*) (resid 11 and name HD*)	0 0 9
assign (resid 10 and name HD1*) (resid 11 and name HD*)	0 0 6
assign (resid 10 and name HD2*) (resid 11 and name HD*)	0 0 6
assign (resid 10 and name HB1) (resid 12 and name HN)	0 0 5
assign (resid 10 and name HB2) (resid 12 and name HN)	0 0 5
assign (resid 10 and name HB1) (resid 12 and name HG*)	0 0 6.5

assign (resid 10 and name HB2) (resid 12 and name HG*)	0 0 6.5
assign (resid 11 and name HA) (resid 11 and name HB1)	0 0 2.7
assign (resid 11 and name HA) (resid 11 and name HB2)	0 0 2.7
assign (resid 11 and name HA) (resid 11 and name HG1)	0 0 5
assign (resid 11 and name HA) (resid 11 and name HG2)	0 0 5
assign (resid 11 and name HB1) (resid 11 and name HB2)	0 0 2.7
assign (resid 11 and name HB1) (resid 11 and name HG1)	0 0 3.5
assign (resid 11 and name HB1) (resid 11 and name HG2)	0 0 3.5
assign (resid 11 and name HB2) (resid 11 and name HG1)	0 0 3.5
assign (resid 11 and name HB2) (resid 11 and name HG2)	0 0 3.5
assign (resid 11 and name HB1) (resid 11 and name HD1)	0 0 5
assign (resid 11 and name HB1) (resid 11 and name HD2)	0 0 5
assign (resid 11 and name HB2) (resid 11 and name HD1)	0 0 5
assign (resid 11 and name HB2) (resid 11 and name HD2)	0 0 5
assign (resid 11 and name HG1) (resid 11 and name HG2)	0 0 2.7
assign (resid 11 and name HG1) (resid 11 and name HD1)	0 0 3.5
assign (resid 11 and name HG1) (resid 11 and name HD2)	0 0 3.5
assign (resid 11 and name HG2) (resid 11 and name HD1)	0 0 3.5
assign (resid 11 and name HG2) (resid 11 and name HD2)	0 0 3.5
assign (resid 11 and name HD1) (resid 11 and name HD2)	0 0 2.7
assign (resid 11 and name HB*) (resid 4 and name HG1*)	0 0 7.5
assign (resid 11 and name HB*) (resid 4 and name HG2*)	0 0 7.5
assign (resid 11 and name HA) (resid 12 and name HN)	0 0 2.7
assign (resid 11 and name HB1) (resid 12 and name HN)	0 0 6
assign (resid 11 and name HB2) (resid 12 and name HN)	0 0 6
assign (resid 11 and name HD*) (resid 12 and name HG*)	0 0 9
assign (resid 11 and name HA) (resid 13 and name HN)	0 0 3.5
assign (resid 12 and name HN) (resid 12 and name HA)	0 0 3.5
assign (resid 12 and name HN) (resid 12 and name HB)	0 0 3.5
assign (resid 12 and name HN) (resid 12 and name HG*)	0 0 5.7
assign (resid 12 and name HN) (resid 13 and name HN)	0 0 2.7
assign (resid 12 and name HA) (resid 12 and name HB)	0 0 3.5
assign (resid 12 and name HA) (resid 12 and name HG1*)	0 0 6.5
assign (resid 12 and name HA) (resid 12 and name HG2*)	0 0 6.5
assign (resid 12 and name HB) (resid 12 and name HG1*)	0 0 4.2
assign (resid 12 and name HB) (resid 12 and name HG2*)	0 0 4.2
assign (resid 12 and name HA) (resid 13 and name HN)	0 0 3.5
assign (resid 12 and name HG*) (resid 13 and name HN)	0 0 6.5
assign (resid 12 and name HG*) (resid 13 and name HA)	0 0 8
assign (resid 13 and name HN) (resid 13 and name HA)	0 0 3.5
assign (resid 13 and name HN) (resid 13 and name HB1)	0 0 3.5
assign (resid 13 and name HN) (resid 13 and name HB2)	0 0 3.5
assign (resid 13 and name HA) (resid 13 and name HB1)	0 0 3.5
assign (resid 13 and name HA) (resid 13 and name HB2)	0 0 3.5
assign (resid 13 and name HB1) (resid 13 and name HB2)	0 0 2.7
assign (resid 13 and name HN) (resid 14 and name HN)	0 0 2.7
assign (resid 13 and name HB*) (resid 14 and name HN)	0 0 6
assign (resid 13 and name HN) (resid 15 and name HN)	0 0 5
assign (resid 13 and name HB1) (resid 15 and name HN)	0 0 5
assign (resid 13 and name HB2) (resid 15 and name HN)	0 0 5
assign (resid 13 and name HB*) (resid 24 and name HG2*)	0 0 7.5
assign (resid 13 and name HB*) (resid 27 and name HB*)	0 0 7
assign (resid 14 and name HN) (resid 14 and name HA1)	0 0 3.5
assign (resid 14 and name HN) (resid 14 and name HA2)	0 0 3.5
assign (resid 14 and name HA1) (resid 14 and name HA2)	0 0 2.7
assign (resid 14 and name HA1) (resid 15 and name HN)	0 0 3.5

assign (resid 14 and name HA2) (resid 15 and name HN)	0 0 3.5
assign (resid 15 and name HN) (resid 15 and name HA)	0 0 3.5
assign (resid 15 and name HN) (resid 15 and name HB1)	0 0 3.5
assign (resid 15 and name HN) (resid 15 and name HB2)	0 0 3.5
assign (resid 15 and name HN) (resid 15 and name HG1)	0 0 5
assign (resid 15 and name HN) (resid 15 and name HG2)	0 0 5
assign (resid 15 and name HA) (resid 15 and name HB1)	0 0 3.5
assign (resid 15 and name HA) (resid 15 and name HB2)	0 0 3.5
assign (resid 15 and name HA) (resid 15 and name HG1)	0 0 3.5
assign (resid 15 and name HA) (resid 15 and name HG2)	0 0 3.5
assign (resid 15 and name HB1) (resid 15 and name HG1)	0 0 2.7
assign (resid 15 and name HB1) (resid 15 and name HG2)	0 0 2.7
assign (resid 15 and name HB2) (resid 15 and name HG1)	0 0 2.7
assign (resid 15 and name HB2) (resid 15 and name HG2)	0 0 2.7
assign (resid 15 and name HG1) (resid 15 and name HG2)	0 0 2.7
assign (resid 15 and name HA) (resid 16 and name HN)	0 0 2.7
assign (resid 15 and name HG1) (resid 16 and name HN)	0 0 3.5
assign (resid 15 and name HG2) (resid 16 and name HN)	0 0 3.5
assign (resid 16 and name HN) (resid 16 and name HA)	0 0 3.5
assign (resid 16 and name HN) (resid 16 and name HB)	0 0 5
assign (resid 16 and name HN) (resid 16 and name HG2*)	0 0 4.2
assign (resid 16 and name HA) (resid 16 and name HG2*)	0 0 5
assign (resid 16 and name HB) (resid 16 and name HG2*)	0 0 5
assign (resid 16 and name HA) (resid 17 and name HN)	0 0 2.7
assign (resid 16 and name HB) (resid 17 and name HN)	0 0 2.7
assign (resid 16 and name HG2*) (resid 17 and name HN)	0 0 5
assign (resid 16 and name HB) (resid 18 and name HN)	0 0 3.5
assign (resid 16 and name HB) (resid 18 and name HG1*)	0 0 7.5
assign (resid 16 and name HB) (resid 18 and name HG2*)	0 0 7.5
assign (resid 17 and name HN) (resid 17 and name HA)	0 0 3.5
assign (resid 17 and name HN) (resid 17 and name HB1)	0 0 5
assign (resid 17 and name HN) (resid 17 and name HB2)	0 0 5
assign (resid 17 and name HA) (resid 17 and name HB*)	0 0 3.7
assign (resid 17 and name HB1) (resid 17 and name HB2)	0 0 2.7
assign (resid 17 and name HN) (resid 18 and name HN)	0 0 5
assign (resid 17 and name HN) (resid 18 and name HG*)	0 0 8
assign (resid 17 and name HA) (resid 18 and name HN)	0 0 3.5
assign (resid 17 and name HB1) (resid 18 and name HN)	0 0 5
assign (resid 17 and name HB2) (resid 18 and name HN)	0 0 5
assign (resid 17 and name HA) (resid 19 and name HN)	0 0 3.5
assign (resid 17 and name HA) (resid 20 and name HN)	0 0 5
assign (resid 17 and name HB1) (resid 20 and name HN)	0 0 5
assign (resid 17 and name HB2) (resid 20 and name HN)	0 0 5
assign (resid 17 and name HA) (resid 21 and name HN)	0 0 5
assign (resid 17 and name HB1) (resid 21 and name HN)	0 0 5
assign (resid 17 and name HB1) (resid 21 and name HN)	0 0 5
assign (resid 18 and name HN) (resid 18 and name HA)	0 0 2.7
assign (resid 18 and name HN) (resid 18 and name HB)	0 0 2.7
assign (resid 18 and name HN) (resid 18 and name HG1*)	0 0 4.2
assign (resid 18 and name HN) (resid 18 and name HG2*)	0 0 4.2
assign (resid 18 and name HA) (resid 18 and name HB)	0 0 3.5
assign (resid 18 and name HA) (resid 18 and name HG1*)	0 0 5
assign (resid 18 and name HA) (resid 18 and name HG2*)	0 0 5
assign (resid 18 and name HB) (resid 18 and name HG*)	0 0 6.5
assign (resid 18 and name HN) (resid 19 and name HN)	0 0 3.5
assign (resid 18 and name HB) (resid 19 and name HN)	0 0 3.5



assign (resid 18 and name HG*) (resid 19 and name HN)	0 0 8
assign (resid 19 and name HN) (resid 19 and name HA1)	0 0 3.5
assign (resid 19 and name HN) (resid 19 and name HA2)	0 0 3.5
assign (resid 19 and name HA1) (resid 19 and name HA2)	0 0 2.7
assign (resid 19 and name HN) (resid 20 and name HN)	0 0 2.7
assign (resid 19 and name HA1) (resid 20 and name HN)	0 0 3.5
assign (resid 19 and name HA2) (resid 20 and name HN)	0 0 3.5
assign (resid 19 and name HN) (resid 21 and name HN)	0 0 3.5
assign (resid 19 and name HA*) (resid 21 and name HN)	0 0 6
assign (resid 20 and name HN) (resid 20 and name HA1)	0 0 3.5
assign (resid 20 and name HN) (resid 20 and name HA2)	0 0 3.5
assign (resid 20 and name HA1) (resid 20 and name HA2)	0 0 2.7
assign (resid 20 and name HN) (resid 21 and name HN)	0 0 2.7
assign (resid 20 and name HA*) (resid 21 and name HN)	0 0 4.5
assign (resid 21 and name HN) (resid 21 and name HA)	0 0 3.5
assign (resid 21 and name HN) (resid 21 and name HB)	0 0 3.5
assign (resid 21 and name HN) (resid 21 and name HG2*)	0 0 5
assign (resid 21 and name HA) (resid 21 and name HG2*)	0 0 5
assign (resid 21 and name HN) (resid 22 and name HN)	0 0 5
assign (resid 21 and name HB) (resid 22 and name HN)	0 0 3.5
assign (resid 21 and name HG2*) (resid 22 and name HN)	0 0 5
assign (resid 22 and name HN) (resid 22 and name HB1)	0 0 3.5
assign (resid 22 and name HN) (resid 22 and name HB2)	0 0 3.5
assign (resid 22 and name HA) (resid 22 and name HB*)	0 0 4.5
assign (resid 22 and name HB1) (resid 22 and name HB2)	0 0 2.7
assign (resid 22 and name HB*) (resid 23 and name HN)	0 0 4.5
assign (resid 22 and name HB1) (resid 24 and name HN)	0 0 3.5
assign (resid 22 and name HB2) (resid 24 and name HN)	0 0 3.5
assign (resid 22 and name HB*) (resid 28 and name HA)	0 0 6
assign (resid 23 and name HD21) (resid 23 and name HD22)	0 0 2.7
assign (resid 23 and name HB*) (resid 23 and name HD21)	0 0 4.5
assign (resid 23 and name HB*) (resid 23 and name HD22)	0 0 4.5
assign (resid 23 and name HN) (resid 24 and name HN)	0 0 2.7
assign (resid 24 and name HN) (resid 24 and name HA)	0 0 3.5
assign (resid 24 and name HN) (resid 24 and name HB)	0 0 2.7
assign (resid 24 and name HA) (resid 24 and name HB)	0 0 3.5
assign (resid 24 and name HA) (resid 24 and name HG2*)	0 0 5
assign (resid 24 and name HB) (resid 24 and name HG2*)	0 0 5
assign (resid 24 and name HA) (resid 25 and name HG1)	0 0 5
assign (resid 24 and name HA) (resid 25 and name HG2)	0 0 5
assign (resid 24 and name HG2*) (resid 25 and name HD1)	0 0 5
assign (resid 24 and name HG2*) (resid 25 and name HD2)	0 0 5
assign (resid 24 and name HB) (resid 27 and name HB1)	0 0 3.5
assign (resid 24 and name HB) (resid 27 and name HB2)	0 0 3.5
assign (resid 24 and name HG2*) (resid 27 and name HN)	0 0 6.5
assign (resid 25 and name HA) (resid 25 and name HB1)	0 0 2.7
assign (resid 25 and name HA) (resid 25 and name HB2)	0 0 2.7
assign (resid 25 and name HA) (resid 25 and name HG1)	0 0 5
assign (resid 25 and name HA) (resid 25 and name HG2)	0 0 5
assign (resid 25 and name HB1) (resid 25 and name HB2)	0 0 2.7
assign (resid 25 and name HB1) (resid 25 and name HG1)	0 0 3.5
assign (resid 25 and name HB1) (resid 25 and name HG2)	0 0 3.5
assign (resid 25 and name HB2) (resid 25 and name HG1)	0 0 3.5
assign (resid 25 and name HB2) (resid 25 and name HG2)	0 0 3.5
assign (resid 25 and name HB1) (resid 25 and name HD1)	0 0 5
assign (resid 25 and name HB1) (resid 25 and name HD2)	0 0 5

assign (resid 25 and name HB2) (resid 25 and name HD1)	0 0 5
assign (resid 25 and name HB2) (resid 25 and name HD2)	0 0 5
assign (resid 25 and name HG1) (resid 25 and name HD1)	0 0 3.5
assign (resid 25 and name HG1) (resid 25 and name HD2)	0 0 3.5
assign (resid 25 and name HG2) (resid 25 and name HD1)	0 0 3.5
assign (resid 25 and name HG2) (resid 25 and name HD2)	0 0 3.5
assign (resid 25 and name HD1) (resid 25 and name HD2)	0 0 2.7
assign (resid 25 and name HA) (resid 26 and name HN)	0 0 2.7
assign (resid 25 and name HA) (resid 26 and name HA*)	0 0 6
assign (resid 25 and name HB*) (resid 26 and name HN)	0 0 4.5
assign (resid 25 and name HA) (resid 27 and name HN)	0 0 3.5
assign (resid 26 and name HN) (resid 26 and name HA1)	0 0 3.5
assign (resid 26 and name HN) (resid 26 and name HA2)	0 0 3.5
assign (resid 26 and name HA1) (resid 26 and name HA2)	0 0 2.7
assign (resid 26 and name HN) (resid 27 and name HN)	0 0 2.7
assign (resid 26 and name HA1) (resid 27 and name HN)	0 0 3.5
assign (resid 26 and name HA2) (resid 27 and name HN)	0 0 3.5
assign (resid 27 and name HN) (resid 27 and name HA*)	0 0 3.5
assign (resid 27 and name HN) (resid 27 and name HB1)	0 0 3.5
assign (resid 27 and name HN) (resid 27 and name HB2)	0 0 3.5
assign (resid 27 and name HA) (resid 27 and name HB1)	0 0 3.5
assign (resid 27 and name HA) (resid 27 and name HB2)	0 0 3.5
assign (resid 27 and name HB1) (resid 27 and name HB2)	0 0 2.7
assign (resid 27 and name HA) (resid 28 and name HN)	0 0 2.7
assign (resid 27 and name HB2) (resid 28 and name HN)	0 0 3.5
assign (resid 27 and name HB1) (resid 28 and name HN)	0 0 3.5
assign (resid 28 and name HN) (resid 28 and name HA)	0 0 3.5
assign (resid 28 and name HN) (resid 28 and name HB)	0 0 3.5
assign (resid 28 and name HN) (resid 28 and name HG2*)	0 0 5
assign (resid 28 and name HA) (resid 28 and name HG2*)	0 0 5
assign (resid 28 and name HN) (resid 29 and name HN)	0 0 5
assign (resid 28 and name HN) (resid 29 and name HB)	0 0 3.5
assign (resid 28 and name HN) (resid 29 and name HA)	0 0 2.7
assign (resid 29 and name HN) (resid 29 and name HA)	0 0 3.5
assign (resid 29 and name HN) (resid 29 and name HB1)	0 0 3.5
assign (resid 29 and name HN) (resid 29 and name HB2)	0 0 3.5
assign (resid 29 and name HA) (resid 29 and name HB*)	0 0 4.5
assign (resid 29 and name HB1) (resid 29 and name HB2)	0 0 2.7

## APPENDIX IV

Table of distance restraints used in the simulated annealing protocol for the determination of the structure of circulin A in the aqueous environment (Chapter 6). The numbering of the residues is according to the sequence as presented in the first publication on circulins (Gustafson *et al.*, 1994). The lower bound of the distance restraint given in the second distance column is not specified, and is, for the purposes of the calculations determined from the Van der Waal's forces. The upper bound distance values, as determined from the relative intensities of the NOE cross peaks and pseudoatom adjusted where relevant, are given in the third distance column.

	d	$\delta$ - $\delta$ +
assign (resid 1 and name HN) (resid 1 and name HA)	0 0	3.5
assign (resid 1 and name HN) (resid 1 and name HB1)	0 0	3.5
assign (resid 1 and name HN) (resid 1 and name HB2)	0 0	3.5
assign (resid 1 and name HA) (resid 1 and name HB1)	0 0	3.5
assign (resid 1 and name HA) (resid 1 and name HB2)	0 0	3.5
assign (resid 1 and name HB1) (resid 1 and name HB2)	0 0	2.7
assign (resid 1 and name HA) (resid 2 and name HN)	0 0	2.7
assign (resid 1 and name HB1) (resid 2 and name HN)	0 0	3.5
assign (resid 1 and name HB2) (resid 2 and name HN)	0 0	3.5
assign (resid 1 and name HB1) (resid 3 and name HN)	0 0	5.0
assign (resid 1 and name HB2) (resid 3 and name HN)	0 0	5.0
assign (resid 1 and name HA) (resid 20 and name HA)	0 0	3.5
assign (resid 1 and name HA) (resid 20 and name HG1*)	0 0	7.5
assign (resid 1 and name HA) (resid 20 and name HG2*)	0 0	7.5
assign (resid 1 and name HA) (resid 21 and name HN)	0 0	3.5
assign (resid 1 and name HN) (resid 30 and name HA)	0 0	2.7
assign (resid 1 and name HN) (resid 30 and name HB1)	0 0	3.5
assign (resid 1 and name HN) (resid 30 and name HB2)	0 0	3.5
assign (resid 1 and name HN) (resid 30 and name HG1)	0 0	5.0
assign (resid 1 and name HN) (resid 30 and name HG2)	0 0	5.0
assign (resid 1 and name HN) (resid 30 and name HN)	0 0	5.0
assign (resid 2 and name HN) (resid 2 and name HA)	0 0	2.7
assign (resid 2 and name HN) (resid 2 and name HB1)	0 0	3.5
assign (resid 2 and name HN) (resid 2 and name HB2)	0 0	3.5
assign (resid 2 and name HA) (resid 2 and name HB1)	0 0	2.7
assign (resid 2 and name HA) (resid 2 and name HB2)	0 0	2.7
assign (resid 2 and name HB1) (resid 2 and name HB2)	0 0	2.7
assign (resid 2 and name HA) (resid 3 and name HN)	0 0	2.7
assign (resid 2 and name HA) (resid 3 and name HG*)	0 0	7.5
assign (resid 2 and name HA) (resid 4 and name HN)	0 0	5.0
assign (resid 2 and name HA) (resid 5 and name HN)	0 0	3.5
assign (resid 2 and name HB1) (resid 6 and name HA)	0 0	5.0
assign (resid 2 and name HB2) (resid 6 and name HA)	0 0	5.0
assign (resid 2 and name HB1) (resid 7 and name HN)	0 0	5.0
assign (resid 2 and name HB2) (resid 7 and name HN)	0 0	5.0
assign (resid 2 and name HN) (resid 20 and name HA)	0 0	3.5
assign (resid 3 and name HN) (resid 3 and name HA)	0 0	3.5
assign (resid 3 and name HN) (resid 3 and name HB)	0 0	2.7
assign (resid 3 and name HN) (resid 3 and name HG1*)	0 0	6.5
assign (resid 3 and name HN) (resid 3 and name HG2*)	0 0	6.5
assign (resid 3 and name HA) (resid 3 and name HB)	0 0	3.5

assign (resid 3 and name HA)(resid 3 and name HG1*)	0 0 5.0
assign (resid 3 and name HA)(resid 3 and name HG2*)	0 0 5.0
assign (resid 3 and name HB)(resid 3 and name HG*)	0 0 7.5
assign (resid 3 and name HG1*)(resid 3 and name HG2*)	0 0 6.5
assign (resid 3 and name HN)(resid 4 and name HN)	0 0 3.5
assign (resid 3 and name HA)(resid 4 and name HN)	0 0 3.5
assign (resid 3 and name HB)(resid 4 and name HN)	0 0 2.7
assign (resid 3 and name HG1*)(resid 4 and name HN)	0 0 6.5
assign (resid 3 and name HG2*)(resid 4 and name HN)	0 0 6.5
assign (resid 3 and name HG*)(resid 4 and name HA)	0 0 7.5
assign (resid 3 and name HN)(resid 5 and name HN)	0 0 5.0
assign (resid 3 and name HA)(resid 19 and name HB1)	0 0 5.0
assign (resid 3 and name HA)(resid 19 and name HB2)	0 0 5.0
assign (resid 4 and name HN)(resid 4 and name HA)	0 0 2.7
assign (resid 4 and name HN)(resid 4 and name HB1)	0 0 3.5
assign (resid 4 and name HN)(resid 4 and name HB2)	0 0 3.5
assign (resid 4 and name HA)(resid 4 and name HB1)	0 0 3.5
assign (resid 4 and name HA)(resid 4 and name HB2)	0 0 3.5
assign (resid 4 and name HB1)(resid 4 and name HB2)	0 0 2.7
assign (resid 4 and name HN)(resid 5 and name HN)	0 0 2.7
assign (resid 4 and name HA)(resid 5 and name HN)	0 0 3.5
assign (resid 4 and name HB1)(resid 5 and name HN)	0 0 3.5
assign (resid 4 and name HB2)(resid 5 and name HN)	0 0 3.5
assign (resid 4 and name HB1)(resid 5 and name HG2*)	0 0 6.5
assign (resid 4 and name HB2)(resid 5 and name HG2*)	0 0 6.5
assign (resid 4 and name HB*)(resid 5 and name HD1*)	0 0 7.5
assign (resid 5 and name HN)(resid 5 and name HA)	0 0 2.7
assign (resid 5 and name HN)(resid 5 and name HB)	0 0 3.5
assign (resid 5 and name HN)(resid 5 and name HG11)	0 0 6.5
assign (resid 5 and name HN)(resid 5 and name HG12)	0 0 6.5
assign (resid 5 and name HN)(resid 5 and name HG2*)	0 0 7.5
assign (resid 5 and name HN)(resid 5 and name HD1*)	0 0 6.5
assign (resid 5 and name HA)(resid 5 and name HB)	0 0 2.7
assign (resid 5 and name HA)(resid 5 and name HG2*)	0 0 6.5
assign (resid 5 and name HA)(resid 5 and name HD1*)	0 0 6.5
assign (resid 5 and name HA)(resid 6 and name HD1)	0 0 2.7
assign (resid 5 and name HA)(resid 6 and name HD2)	0 0 2.7
assign (resid 5 and name HG2*)(resid 6 and name HD1)	0 0 7.5
assign (resid 5 and name HG2*)(resid 6 and name HD2)	0 0 7.5
assign (resid 6 and name HA)(resid 6 and name HB1)	0 0 3.5
assign (resid 6 and name HA)(resid 6 and name HB2)	0 0 3.5
assign (resid 6 and name HA)(resid 6 and name HG*)	0 0 4.5
assign (resid 6 and name HA)(resid 6 and name HD1)	0 0 5.0
assign (resid 6 and name HA)(resid 6 and name HD2)	0 0 5.0
assign (resid 6 and name HB1)(resid 6 and name HB2)	0 0 2.7
assign (resid 6 and name HB1)(resid 6 and name HG*)	0 0 6.5
assign (resid 6 and name HB2)(resid 6 and name HG*)	0 0 6.5
assign (resid 6 and name HB1)(resid 6 and name HD1)	0 0 5.0
assign (resid 6 and name HB1)(resid 6 and name HD2)	0 0 5.0
assign (resid 6 and name HB2)(resid 6 and name HD1)	0 0 5.0
assign (resid 6 and name HB2)(resid 6 and name HD2)	0 0 5.0
assign (resid 6 and name HG*)(resid 6 and name HD1)	0 0 4.5
assign (resid 6 and name HG*)(resid 6 and name HD2)	0 0 4.5
assign (resid 6 and name HD1)(resid 6 and name HD2)	0 0 2.7
assign (resid 6 and name HA)(resid 7 and name HN)	0 0 2.7
assign (resid 6 and name HB1)(resid 7 and name HN)	0 0 3.5

assign (resid 6 and name HB2)(resid 7 and name HN)	0 0 3.5
assign (resid 6 and name HG*)(resid 7 and name HN)	0 0 6.0
assign (resid 7 and name HN)(resid 7 and name HA)	0 0 3.5
assign (resid 7 and name HN)(resid 7 and name HB1)	0 0 5.0
assign (resid 7 and name HN)(resid 7 and name HB2)	0 0 5.0
assign (resid 7 and name HA)(resid 7 and name HB1)	0 0 2.7
assign (resid 7 and name HA)(resid 7 and name HB2)	0 0 2.7
assign (resid 7 and name HB1)(resid 7 and name HB2)	0 0 2.7
assign (resid 7 and name HN)(resid 8 and name HN)	0 0 5.0
assign (resid 7 and name HA)(resid 8 and name HN)	0 0 2.7
assign (resid 7 and name HB1)(resid 8 and name HN)	0 0 5.0
assign (resid 7 and name HB2)(resid 8 and name HN)	0 0 5.0
assign (resid 7 and name HA)(resid 9 and name HN)	0 0 5.0
assign (resid 7 and name HB1)(resid 9 and name HN)	0 0 5.0
assign (resid 7 and name HB2)(resid 9 and name HN)	0 0 5.0
assign (resid 7 and name HN)(resid 10 and name HB*)	0 0 6.0
assign (resid 7 and name HA)(resid 10 and name HN)	0 0 6.0
assign (resid 7 and name HB1)(resid 10 and name HN)	0 0 5.0
assign (resid 7 and name HB2)(resid 10 and name HN)	0 0 5.0
assign (resid 7 and name HN)(resid 16 and name HB*)	0 0 6.0
assign (resid 8 and name HN)(resid 8 and name HA)	0 0 3.5
assign (resid 8 and name HN)(resid 8 and name HB)	0 0 3.5
assign (resid 8 and name HN)(resid 8 and name HG1*)	0 0 5.0
assign (resid 8 and name HA)(resid 8 and name HB)	0 0 3.5
assign (resid 8 and name HA)(resid 8 and name HG1*)	0 0 5.0
assign (resid 8 and name HN)(resid 9 and name HN)	0 0 3.5
assign (resid 8 and name HA)(resid 9 and name HN)	0 0 3.5
assign (resid 8 and name HB)(resid 9 and name HN)	0 0 2.7
assign (resid 8 and name HG1*)(resid 9 and name HN)	0 0 6.5
assign (resid 8 and name HD*)(resid 9 and name HN)	0 0 5.5
assign (resid 9 and name HN)(resid 9 and name HA)	0 0 2.7
assign (resid 9 and name HN)(resid 9 and name HB1)	0 0 3.5
assign (resid 9 and name HN)(resid 9 and name HB2)	0 0 3.5
assign (resid 9 and name HA)(resid 9 and name HB1)	0 0 3.5
assign (resid 9 and name HA)(resid 9 and name HB2)	0 0 3.5
assign (resid 9 and name HN)(resid 10 and name HN)	0 0 3.5
assign (resid 9 and name HA)(resid 10 and name HN)	0 0 3.5
assign (resid 9 and name HB1)(resid 10 and name HN)	0 0 3.5
assign (resid 9 and name HB2)(resid 10 and name HN)	0 0 3.5
assign (resid 9 and name HA)(resid 12 and name HN)	0 0 5.0
assign (resid 9 and name HB1)(resid 12 and name HN)	0 0 6.0
assign (resid 9 and name HB2)(resid 12 and name HN)	0 0 6.0
assign (resid 10 and name HN)(resid 10 and name HA)	0 0 2.7
assign (resid 10 and name HN)(resid 10 and name HB*)	0 0 4.2
assign (resid 10 and name HA)(resid 10 and name HB*)	0 0 4.2
assign (resid 10 and name HN)(resid 11 and name HN)	0 0 3.5
assign (resid 10 and name HA)(resid 11 and name HN)	0 0 3.5
assign (resid 10 and name HB*)(resid 11 and name HN)	0 0 5.0
assign (resid 10 and name HA)(resid 12 and name HN)	0 0 5.0
assign (resid 10 and name HA)(resid 14 and name HN)	0 0 3.5
assign (resid 10 and name HA)(resid 14 and name HB*)	0 0 7.0
assign (resid 11 and name HN)(resid 11 and name HA)	0 0 3.5
assign (resid 11 and name HN)(resid 11 and name HB*)	0 0 4.2
assign (resid 11 and name HN)(resid 12 and name HN)	0 0 3.5
assign (resid 12 and name HN)(resid 12 and name HA)	0 0 2.7
assign (resid 12 and name HN)(resid 12 and name HB1)	0 0 3.5

assign (resid 12 and name HN)(resid 12 and name HB2)	0 0 3.5
assign (resid 12 and name HN)(resid 12 and name HG)	0 0 3.5
assign (resid 12 and name HN)(resid 12 and name HD*)	0 0 6.5
assign (resid 12 and name HA)(resid 12 and name HB1)	0 0 3.5
assign (resid 12 and name HA)(resid 12 and name HB2)	0 0 3.5
assign (resid 12 and name HA)(resid 12 and name HG)	0 0 3.5
assign (resid 12 and name HA)(resid 12 and name HD1*)	0 0 5.0
assign (resid 12 and name HA)(resid 12 and name HD2*)	0 0 5.0
assign (resid 12 and name HA)(resid 13 and name HN)	0 0 3.5
assign (resid 12 and name HB1)(resid 13 and name HN)	0 0 5.0
assign (resid 12 and name HB2)(resid 13 and name HN)	0 0 5.0
assign (resid 12 and name HD*)(resid 13 and name HN)	0 0 8.0
assign (resid 13 and name HN)(resid 13 and name HA1)	0 0 2.7
assign (resid 13 and name HN)(resid 13 and name HA2)	0 0 2.7
assign (resid 13 and name HA1)(resid 13 and name HA2)	0 0 2.7
assign (resid 13 and name HN)(resid 14 and name HN)	0 0 3.5
assign (resid 13 and name HA1)(resid 14 and name HN)	0 0 3.5
assign (resid 13 and name HA2)(resid 14 and name HN)	0 0 3.5
assign (resid 14 and name HN)(resid 14 and name HA)	0 0 3.5
assign (resid 14 and name HN)(resid 14 and name HB1)	0 0 3.5
assign (resid 14 and name HN)(resid 14 and name HB2)	0 0 3.5
assign (resid 14 and name HA)(resid 14 and name HB1)	0 0 3.5
assign (resid 14 and name HA)(resid 14 and name HB2)	0 0 3.5
assign (resid 14 and name HB1)(resid 14 and name HB2)	0 0 2.7
assign (resid 14 and name HN)(resid 15 and name HN)	0 0 5.0
assign (resid 14 and name HA)(resid 15 and name HN)	0 0 2.7
assign (resid 14 and name HB1)(resid 15 and name HN)	0 0 5.0
assign (resid 14 and name HB2)(resid 15 and name HN)	0 0 5.0
assign (resid 14 and name HB1)(resid 22 and name HN)	0 0 5.0
assign (resid 14 and name HB2)(resid 22 and name HN)	0 0 5.0
assign (resid 15 and name HN)(resid 15 and name HA)	0 0 3.5
assign (resid 15 and name HN)(resid 15 and name HB*)	0 0 3.7
assign (resid 15 and name HA)(resid 15 and name HB*)	0 0 4.5
assign (resid 15 and name HA)(resid 16 and name HN)	0 0 2.7
assign (resid 15 and name HB*)(resid 16 and name HN)	0 0 3.7
assign (resid 15 and name HN)(resid 22 and name HN)	0 0 5.0
assign (resid 15 and name HN)(resid 21 and name HA)	0 0 5.0
assign (resid 16 and name HN)(resid 16 and name HA)	0 0 3.5
assign (resid 16 and name HN)(resid 16 and name HB*)	0 0 3.7
assign (resid 16 and name HA)(resid 16 and name HB*)	0 0 3.7
assign (resid 16 and name HA)(resid 17 and name HN)	0 0 2.7
assign (resid 16 and name HB*)(resid 17 and name HN)	0 0 6.0
assign (resid 16 and name HA)(resid 22 and name HN)	0 0 5.0
assign (resid 16 and name HB*)(resid 22 and name HN)	0 0 6.0
assign (resid 16 and name HA)(resid 20 and name HN)	0 0 5.0
assign (resid 17 and name HN)(resid 17 and name HA)	0 0 3.5
assign (resid 17 and name HN)(resid 17 and name HB*)	0 0 3.7
assign (resid 17 and name HN)(resid 17 and name HE*)	0 0 7.0
assign (resid 17 and name HA)(resid 17 and name HB*)	0 0 3.7
assign (resid 17 and name HA)(resid 17 and name HG*)	0 0 6.0
assign (resid 17 and name HA)(resid 17 and name HD*)	0 0 6.0
assign (resid 17 and name HN)(resid 18 and name HN)	0 0 5.0
assign (resid 17 and name HA)(resid 18 and name HN)	0 0 2.7
assign (resid 17 and name HB*)(resid 18 and name HN)	0 0 6.0
assign (resid 17 and name HG*)(resid 18 and name HN)	0 0 6.0
assign (resid 17 and name HD*)(resid 18 and name HN)	0 0 6.0

assign (resid 17 and name HN)(resid 19 and name HN)	0 0 5.0
assign (resid 17 and name HA)(resid 19 and name HN)	0 0 5.0
assign (resid 17 and name HN)(resid 20 and name HN)	0 0 3.5
assign (resid 17 and name HN)(resid 21 and name HA)	0 0 3.5
assign (resid 17 and name HN)(resid 22 and name HN)	0 0 5.0
assign (resid 17 and name HN)(resid 22 and name HD*)	0 0 7.0
assign (resid 18 and name HN)(resid 18 and name HA)	0 0 2.7
assign (resid 18 and name HN)(resid 18 and name HB1)	0 0 5.0
assign (resid 18 and name HN)(resid 18 and name HB2)	0 0 5.0
assign (resid 18 and name HA)(resid 18 and name HB1)	0 0 3.5
assign (resid 18 and name HA)(resid 18 and name HB2)	0 0 3.5
assign (resid 18 and name HB1)(resid 18 and name HB2)	0 0 2.7
assign (resid 18 and name HN)(resid 19 and name HN)	0 0 3.5
assign (resid 18 and name HA)(resid 19 and name HN)	0 0 2.7
assign (resid 18 and name HB1)(resid 19 and name HN)	0 0 5.0
assign (resid 18 and name HB2)(resid 19 and name HN)	0 0 5.0
assign (resid 18 and name HB*)(resid 19 and name HD*)	0 0 7.0
assign (resid 19 and name HN)(resid 19 and name HA)	0 0 2.7
assign (resid 19 and name HN)(resid 19 and name HB1)	0 0 5.0
assign (resid 19 and name HN)(resid 19 and name HB2)	0 0 5.0
assign (resid 19 and name HN)(resid 19 and name HG*)	0 0 3.7
assign (resid 19 and name HA)(resid 19 and name HB1)	0 0 2.7
assign (resid 19 and name HA)(resid 19 and name HB2)	0 0 2.7
assign (resid 19 and name HA)(resid 19 and name HG*)	0 0 3.7
assign (resid 19 and name HB1)(resid 19 and name HB2)	0 0 2.7
assign (resid 19 and name HB1)(resid 19 and name HG*)	0 0 4.5
assign (resid 19 and name HB2)(resid 19 and name HG*)	0 0 4.5
assign (resid 19 and name HB*)(resid 19 and name HD1)	0 0 3.7
assign (resid 19 and name HB*)(resid 19 and name HD2)	0 0 3.7
assign (resid 19 and name HG*)(resid 19 and name HD1)	0 0 3.7
assign (resid 19 and name HG*)(resid 19 and name HD2)	0 0 3.7
assign (resid 19 and name HG*)(resid 19 and name HE*)	0 0 5.5
assign (resid 19 and name HD1)(resid 19 and name HE*)	0 0 4.5
assign (resid 19 and name HD2)(resid 19 and name HE*)	0 0 4.5
assign (resid 19 and name HN)(resid 20 and name HN)	0 0 3.5
assign (resid 19 and name HN)(resid 20 and name HA)	0 0 3.5
assign (resid 19 and name HB1)(resid 20 and name HN)	0 0 5.0
assign (resid 19 and name HB2)(resid 20 and name HN)	0 0 5.0
assign (resid 20 and name HN)(resid 20 and name HA)	0 0 3.5
assign (resid 20 and name HN)(resid 20 and name HB)	0 0 2.7
assign (resid 20 and name HN)(resid 20 and name HG1*)	0 0 6.5
assign (resid 20 and name HN)(resid 20 and name HG2*)	0 0 6.5
assign (resid 20 and name HA)(resid 20 and name HB)	0 0 3.5
assign (resid 20 and name HA)(resid 20 and name HG1*)	0 0 5.0
assign (resid 20 and name HA)(resid 20 and name HG2*)	0 0 5.0
assign (resid 20 and name HB)(resid 20 and name HG1*)	0 0 4.2
assign (resid 20 and name HB)(resid 20 and name HG2*)	0 0 4.2
assign (resid 20 and name HG1*)(resid 20 and name HG2*)	0 0 5.7
assign (resid 20 and name HA)(resid 21 and name HN)	0 0 2.7
assign (resid 20 and name HG1*)(resid 21 and name HN)	0 0 6.5
assign (resid 20 and name HG2*)(resid 21 and name HN)	0 0 6.5
assign (resid 20 and name HN)(resid 22 and name HD*)	0 0 7.0
assign (resid 20 and name HN)(resid 22 and name HE*)	0 0 7.0
assign (resid 20 and name HG1*)(resid 22 and name HD*)	0 0 8.5
assign (resid 20 and name HG2*)(resid 22 and name HD*)	0 0 8.5
assign (resid 20 and name HG1*)(resid 22 and name HE*)	0 0 7.0

assign (resid 20 and name HG2*)(resid 22 and name HE*)	0 0 7.0
assign (resid 20 and name HA)(resid 30 and name HN)	0 0 5.0
assign (resid 21 and name HN)(resid 21 and name HA)	0 0 3.5
assign (resid 21 and name HN)(resid 21 and name HB1)	0 0 3.5
assign (resid 21 and name HN)(resid 21 and name HB2)	0 0 3.5
assign (resid 21 and name HA)(resid 21 and name HB1)	0 0 3.5
assign (resid 21 and name HA)(resid 21 and name HB2)	0 0 3.5
assign (resid 21 and name HB1)(resid 21 and name HB2)	0 0 2.7
assign (resid 21 and name HA)(resid 22 and name HN)	0 0 2.7
assign (resid 21 and name HB1)(resid 22 and name HN)	0 0 5.0
assign (resid 21 and name HB2)(resid 22 and name HN)	0 0 5.0
assign (resid 21 and name HN)(resid 30 and name HB1)	0 0 5.0
assign (resid 21 and name HN)(resid 30 and name HB2)	0 0 5.0
assign (resid 22 and name HN)(resid 22 and name HA)	0 0 3.5
assign (resid 22 and name HN)(resid 22 and name HB1)	0 0 3.5
assign (resid 22 and name HN)(resid 22 and name HB2)	0 0 3.5
assign (resid 22 and name HA)(resid 22 and name HB*)	0 0 3.7
assign (resid 22 and name HB1)(resid 22 and name HB2)	0 0 2.7
assign (resid 22 and name HD*)(resid 22 and name HA)	0 0 4.7
assign (resid 22 and name HD*)(resid 22 and name HB1)	0 0 4.7
assign (resid 22 and name HD*)(resid 22 and name HB2)	0 0 4.7
assign (resid 22 and name HE*)(resid 22 and name HA)	0 0 7.0
assign (resid 22 and name HE*)(resid 22 and name HB1)	0 0 7.0
assign (resid 22 and name HE*)(resid 22 and name HB2)	0 0 7.0
assign (resid 22 and name HA)(resid 23 and name HN)	0 0 2.7
assign (resid 22 and name HB1)(resid 23 and name HN)	0 0 3.5
assign (resid 22 and name HB2)(resid 23 and name HN)	0 0 3.5
assign (resid 22 and name HB1)(resid 25 and name HN)	0 0 3.5
assign (resid 22 and name HB2)(resid 25 and name HN)	0 0 3.5
assign (resid 22 and name HA)(resid 26 and name HN)	0 0 5.0
assign (resid 22 and name HB1)(resid 26 and name HN)	0 0 5.0
assign (resid 22 and name HB2)(resid 26 and name HN)	0 0 5.0
assign (resid 22 and name HD*)(resid 26 and name HA)	0 0 7.0
assign (resid 22 and name HE*)(resid 26 and name HN)	0 0 7.0
assign (resid 22 and name HE*)(resid 26 and name HA)	0 0 7.0
assign (resid 22 and name HA)(resid 27 and name HA)	0 0 3.5
assign (resid 22 and name HD*)(resid 27 and name HA)	0 0 4.7
assign (resid 22 and name HD*)(resid 27 and name HB1)	0 0 5.5
assign (resid 22 and name HD*)(resid 27 and name HB2)	0 0 5.5
assign (resid 22 and name HD*)(resid 27 and name HG*)	0 0 6.5
assign (resid 22 and name HD*)(resid 27 and name HD1)	0 0 7.0
assign (resid 22 and name HD*)(resid 27 and name HD2)	0 0 7.0
assign (resid 22 and name HE*)(resid 27 and name HA)	0 0 5.5
assign (resid 22 and name HE*)(resid 27 and name HB1)	0 0 5.5
assign (resid 22 and name HE*)(resid 27 and name HB2)	0 0 5.5
assign (resid 22 and name HE*)(resid 27 and name HG*)	0 0 5.7
assign (resid 22 and name HE*)(resid 27 and name HD1)	0 0 7.0
assign (resid 22 and name HE*)(resid 27 and name HD2)	0 0 7.0
assign (resid 22 and name HA)(resid 28 and name HN)	0 0 2.7
assign (resid 22 and name HE*)(resid 28 and name HN)	0 0 7.0
assign (resid 22 and name HD*)(resid 29 and name HA2)	0 0 7.0
assign (resid 23 and name HN)(resid 23 and name HA)	0 0 3.5
assign (resid 23 and name HA)(resid 24 and name HN)	0 0 2.7
assign (resid 23 and name HN)(resid 27 and name HA)	0 0 5.0
assign (resid 24 and name HN)(resid 24 and name HA)	0 0 2.7
assign (resid 24 and name HN)(resid 24 and name HB1)	0 0 5.0



assign (resid 24 and name HN)(resid 24 and name HB2)	0 0 5.0
assign (resid 24 and name HN)(resid 25 and name HN)	0 0 3.5
assign (resid 24 and name HA)(resid 25 and name HN)	0 0 2.7
assign (resid 24 and name HN)(resid 26 and name HN)	0 0 5.0
assign (resid 25 and name HN)(resid 25 and name HA1)	0 0 2.7
assign (resid 25 and name HN)(resid 25 and name HA2)	0 0 2.7
assign (resid 25 and name HN)(resid 26 and name HN)	0 0 3.5
assign (resid 25 and name HA1)(resid 26 and name HN)	0 0 3.5
assign (resid 25 and name HA2)(resid 26 and name HN)	0 0 3.5
assign (resid 26 and name HN)(resid 26 and name HA)	0 0 3.5
assign (resid 26 and name HN)(resid 26 and name HB)	0 0 2.7
assign (resid 26 and name HN)(resid 26 and name HG11)	0 0 5.0
assign (resid 26 and name HN)(resid 26 and name HG12)	0 0 5.0
assign (resid 26 and name HN)(resid 26 and name HG2*)	0 0 6.5
assign (resid 26 and name HN)(resid 26 and name HD1*)	0 0 6.5
assign (resid 26 and name HA)(resid 26 and name HB)	0 0 3.5
assign (resid 26 and name HA)(resid 26 and name HG11)	0 0 5.0
assign (resid 26 and name HA)(resid 26 and name HG12)	0 0 5.0
assign (resid 26 and name HA)(resid 26 and name HG2*)	0 0 5.0
assign (resid 26 and name HA)(resid 26 and name HD1*)	0 0 5.0
assign (resid 26 and name HB)(resid 26 and name HG11)	0 0 4.2
assign (resid 26 and name HB)(resid 26 and name HG12)	0 0 4.2
assign (resid 26 and name HB)(resid 26 and name HG2*)	0 0 4.2
assign (resid 26 and name HB)(resid 26 and name HD1*)	0 0 4.2
assign (resid 26 and name HA)(resid 27 and name HD1)	0 0 3.5
assign (resid 26 and name HA)(resid 27 and name HD2)	0 0 3.5
assign (resid 26 and name HG2*)(resid 27 and name HD*)	0 0 5.0
assign (resid 27 and name HN)(resid 28 and name HA)	0 0 2.7
assign (resid 27 and name HA)(resid 27 and name HB*)	0 0 3.7
assign (resid 27 and name HA)(resid 29 and name HN)	0 0 5.0
assign (resid 27 and name HB1)(resid 29 and name HN)	0 0 5.0
assign (resid 27 and name HB2)(resid 29 and name HN)	0 0 5.0
assign (resid 28 and name HN)(resid 28 and name HA)	0 0 3.5
assign (resid 28 and name HN)(resid 28 and name HB1)	0 0 3.5
assign (resid 28 and name HN)(resid 28 and name HB2)	0 0 3.5
assign (resid 28 and name HN)(resid 29 and name HN)	0 0 2.7
assign (resid 28 and name HA)(resid 29 and name HN)	0 0 5.0
assign (resid 28 and name HB1)(resid 29 and name HN)	0 0 5.0
assign (resid 28 and name HB2)(resid 29 and name HN)	0 0 5.0
assign (resid 28 and name HN)(resid 30 and name HN)	0 0 5.0
assign (resid 28 and name HB1)(resid 30 and name HN)	0 0 5.0
assign (resid 28 and name HB2)(resid 30 and name HN)	0 0 5.0
assign (resid 29 and name HN)(resid 29 and name HA1)	0 0 3.5
assign (resid 29 and name HN)(resid 29 and name HA2)	0 0 3.5
assign (resid 29 and name HN)(resid 30 and name HN)	0 0 2.7
assign (resid 29 and name HA1)(resid 30 and name HN)	0 0 3.5
assign (resid 29 and name HA2)(resid 30 and name HN)	0 0 3.5
assign (resid 30 and name HN)(resid 30 and name HA)	0 0 2.7
assign (resid 30 and name HN)(resid 30 and name HB1)	0 0 3.5
assign (resid 30 and name HN)(resid 30 and name HB2)	0 0 3.5
assign (resid 30 and name HN)(resid 30 and name HG1)	0 0 5.0
assign (resid 30 and name HN)(resid 30 and name HG2)	0 0 5.0

## APPENDIX V

List of dihedral restraints and hydrogen bond definitions applied in the simulated annealing protocol for the determination of the structure of circulin A in the aqueous environment (Chapter 6). The numbering of the residues is according to the sequence as presented in the first publication on circulins (Gustafson *et al.*, 1994).

### PHI ANGLES

phi angle	range	basis of application	residues
-120°	± 30°	$^3J_{\text{NH-H}\alpha}$ within $8.5 \pm 0.5$ Hz	1,2,5,8,15,17,20,26
-120°	± 15°	$^3J_{\text{NH-H}\alpha} > 9.5$ Hz	22,23
-65°	± 15°	$^3J_{\text{NH-H}\alpha} < 5.8$ Hz	7
-100	± 80°	intraresidue H $^\alpha$ -H $_N$ NOE is weaker than the sequential H $^\alpha$ -H $_N$ NOE	3,16,21
50°	± 40°	intense intraresidue H $^\alpha$ -H $_N$ NOE and $^3J_{\text{NH-H}\alpha} \sim 7$ Hz	18,19,24

### CHI 1 ANGLES

chi 1	range	residues
-60	± 30°	4,21,22,28
60	± 30°	2
180	± 30°	3,8,20,26

### HYDROGEN BONDS

CO Residue	NH Residue
21	28
19	2
22	15
17	20
30	21
15	22
26	23
23	26

## APPENDIX VI

Table of distance restraints used in the simulated annealing protocol for the determination of the structure of circulin B in the aqueous environment (Chapter 7). The numbering of the residues is according to that used for the first collective alignment of members of the inhibitor cystine knot family (Pallaghy *et al.*, 1994). The lower bound of the distance restraint given in the second distance column is not specified, and is, for the purposes of the calculations determined from the Van der Waal's forces. The upper bound distance values, as determined from the relative intensities of the NOE cross peaks and pseudoatom adjusted where relevant, are given in the third distance column.

	d	$\delta^-$	$\delta^+$
assign (resid 1 and name HN) (resid 1 and name HA)	0 0	3.5	
assign (resid 1 and name HN) (resid 1 and name HB*)	0 0	3.7	
assign (resid 1 and name HA) (resid 1 and name HB*)	0 0	3.7	
assign (resid 1 and name HN) (resid 2 and name HN)	0 0	2.7	
assign (resid 1 and name HN) (resid 2 and name HA1)	0 0	5.0	
assign (resid 1 and name HN) (resid 2 and name HA2)	0 0	5.0	
assign (resid 1 and name HA) (resid 2 and name HN)	0 0	3.5	
assign (resid 1 and name HB*) (resid 2 and name HN)	0 0	4.5	
assign (resid 1 and name HN) (resid 3 and name HN)	0 0	3.5	
assign (resid 1 and name HN) (resid 23 and name HG*)	0 0	6.5	
assign (resid 1 and name HN) (resid 25 and name HA)	0 0	2.7	
assign (resid 1 and name HN) (resid 25 and name HB1)	0 0	5.0	
assign (resid 1 and name HN) (resid 25 and name HB2)	0 0	5.0	
assign (resid 1 and name HN) (resid 25 and name HD*)	0 0	7.0	
assign (resid 1 and name HN) (resid 25 and name HE*)	0 0	7.0	
assign (resid 1 and name HN) (resid 26 and name HN)	0 0	3.5	
assign (resid 1 and name HN) (resid 31 and name HA)	0 0	2.7	
assign (resid 1 and name HN) (resid 31 and name HB*)	0 0	4.5	
assign (resid 1 and name HN) (resid 31 and name HG*)	0 0	6.0	
assign (resid 1 and name HN) (resid 31 and name HD1)	0 0	6.0	
assign (resid 1 and name HN) (resid 31 and name HD2)	0 0	6.0	
assign (resid 2 and name HN) (resid 2 and name HA1)	0 0	3.5	
assign (resid 2 and name HN) (resid 2 and name HA2)	0 0	3.5	
assign (resid 2 and name HA1) (resid 2 and name HA2)	0 0	2.7	
assign (resid 2 and name HN) (resid 3 and name HN)	0 0	3.5	
assign (resid 2 and name HA1) (resid 3 and name HN)	0 0	3.5	
assign (resid 2 and name HA2) (resid 3 and name HN)	0 0	3.5	
assign (resid 2 and name HN) (resid 25 and name HA)	0 0	5.0	
assign (resid 2 and name HN) (resid 25 and name HD*)	0 0	6.0	
assign (resid 2 and name HN) (resid 31 and name HA)	0 0	3.5	
assign (resid 2 and name HN) (resid 31 and name HB*)	0 0	4.5	
assign (resid 2 and name HN) (resid 31 and name HG*)	0 0	6.0	
assign (resid 3 and name HN) (resid 3 and name HA)	0 0	3.5	
assign (resid 3 and name HN) (resid 3 and name HB1)	0 0	3.5	
assign (resid 3 and name HN) (resid 3 and name HB2)	0 0	3.5	
assign (resid 3 and name HN) (resid 3 and name HG1)	0 0	3.5	
assign (resid 3 and name HN) (resid 3 and name HG2)	0 0	3.5	
assign (resid 3 and name HA) (resid 3 and name HB1)	0 0	3.5	
assign (resid 3 and name HA) (resid 3 and name HB2)	0 0	3.5	
assign (resid 3 and name HA) (resid 3 and name HG1)	0 0	3.5	

assign (resid 3 and name HA) (resid 3 and name HG2)	0 0 3.5
assign (resid 3 and name HG1) (resid 3 and name HG2)	0 0 2.7
assign (resid 3 and name HN) (resid 4 and name HN)	0 0 5.0
assign (resid 3 and name HA) (resid 4 and name HN)	0 0 2.7
assign (resid 3 and name HB1) (resid 4 and name HN)	0 0 3.5
assign (resid 3 and name HB2) (resid 4 and name HN)	0 0 3.5
assign (resid 3 and name HG1) (resid 4 and name HN)	0 0 5.0
assign (resid 3 and name HG2) (resid 4 and name HN)	0 0 5.0
assign (resid 3 and name HB1) (resid 24 and name HB1)	0 0 3.5
assign (resid 3 and name HB1) (resid 24 and name HB2)	0 0 3.5
assign (resid 3 and name HB2) (resid 24 and name HB1)	0 0 3.5
assign (resid 3 and name HB2) (resid 24 and name HB2)	0 0 3.5
assign (resid 3 and name HG1) (resid 24 and name HB1)	0 0 5.0
assign (resid 3 and name HG1) (resid 24 and name HB2)	0 0 5.0
assign (resid 3 and name HG2) (resid 24 and name HB2)	0 0 5.0
assign (resid 3 and name HG2) (resid 24 and name HB1)	0 0 5.0
assign (resid 4 and name HN) (resid 4 and name HA)	0 0 3.5
assign (resid 4 and name HN) (resid 4 and name HB1)	0 0 5.0
assign (resid 4 and name HN) (resid 4 and name HB2)	0 0 5.0
assign (resid 4 and name HA) (resid 4 and name HB1)	0 0 2.7
assign (resid 4 and name HA) (resid 4 and name HB2)	0 0 2.7
assign (resid 4 and name HB1) (resid 4 and name HB2)	0 0 2.7
assign (resid 4 and name HN) (resid 5 and name HN)	0 0 5.0
assign (resid 4 and name HA) (resid 5 and name HN)	0 0 2.7
assign (resid 4 and name HB1) (resid 5 and name HN)	0 0 3.5
assign (resid 4 and name HB2) (resid 5 and name HN)	0 0 3.5
assign (resid 4 and name HB1) (resid 6 and name HN)	0 0 3.5
assign (resid 4 and name HB2) (resid 6 and name HN)	0 0 3.5
assign (resid 4 and name HN) (resid 23 and name HA)	0 0 6.0
assign (resid 5 and name HN) (resid 5 and name HA)	0 0 3.5
assign (resid 5 and name HN) (resid 5 and name HB1)	0 0 5.0
assign (resid 5 and name HN) (resid 5 and name HB2)	0 0 5.0
assign (resid 5 and name HA) (resid 5 and name HB1)	0 0 3.5
assign (resid 5 and name HA) (resid 5 and name HB2)	0 0 3.5
assign (resid 5 and name HB1) (resid 5 and name HB2)	0 0 2.7
assign (resid 5 and name HN) (resid 6 and name HN)	0 0 2.7
assign (resid 5 and name HN) (resid 6 and name HG*)	0 0 8.0
assign (resid 5 and name HA) (resid 6 and name HN)	0 0 2.7
assign (resid 5 and name HA) (resid 6 and name HB)	0 0 5.0
assign (resid 5 and name HA) (resid 6 and name HG*)	0 0 6.5
assign (resid 5 and name HB1) (resid 6 and name HN)	0 0 5.0
assign (resid 5 and name HB2) (resid 6 and name HN)	0 0 5.0
assign (resid 5 and name HA) (resid 7 and name HN)	0 0 3.5
assign (resid 5 and name HA) (resid 8 and name HN)	0 0 3.5
assign (resid 5 and name HA) (resid 8 and name HG11)	0 0 5.0
assign (resid 5 and name HA) (resid 8 and name HG12)	0 0 5.0
assign (resid 5 and name HA) (resid 8 and name HG2*)	0 0 6.5
assign (resid 5 and name HA) (resid 8 and name HD1*)	0 0 6.5
assign (resid 5 and name HB1) (resid 9 and name HA)	0 0 3.5
assign (resid 5 and name HB2) (resid 9 and name HA)	0 0 3.5
assign (resid 5 and name HA) (resid 10 and name HN)	0 0 5.0
assign (resid 5 and name HB1) (resid 10 and name HN)	0 0 3.5
assign (resid 5 and name HB2) (resid 10 and name HN)	0 0 3.5
assign (resid 5 and name HB1) (resid 19 and name HA)	0 0 5.0
assign (resid 5 and name HB2) (resid 19 and name HA)	0 0 5.0
assign (resid 5 and name HN) (resid 23 and name HA)	0 0 2.7

assign (resid 5 and name HN) (resid 23 and name HG*)	0 0 8.0
assign (resid 5 and name HN) (resid 24 and name HN)	0 0 3.5
assign (resid 6 and name HN) (resid 6 and name HG*)	0 0 5.7
assign (resid 6 and name HA) (resid 6 and name HB)	0 0 3.5
assign (resid 6 and name HA) (resid 6 and name HG1*)	0 0 4.2
assign (resid 6 and name HA) (resid 6 and name HG2*)	0 0 4.2
assign (resid 6 and name HN) (resid 4 and name HA)	0 0 3.5
assign (resid 6 and name HN) (resid 7 and name HN)	0 0 2.7
assign (resid 6 and name HA) (resid 7 and name HN)	0 0 3.5
assign (resid 6 and name HA) (resid 7 and name HE*)	0 0 8.0
assign (resid 6 and name HB) (resid 7 and name HN)	0 0 2.7
assign (resid 6 and name HB) (resid 7 and name HD*)	0 0 4.7
assign (resid 6 and name HB) (resid 7 and name HE*)	0 0 5.5
assign (resid 6 and name HG1*) (resid 7 and name HN)	0 0 5.0
assign (resid 6 and name HG2*) (resid 7 and name HN)	0 0 5.0
assign (resid 6 and name HG1*) (resid 7 and name HD*)	0 0 7.0
assign (resid 6 and name HG2*) (resid 7 and name HD*)	0 0 7.0
assign (resid 6 and name HG1*) (resid 7 and name HE*)	0 0 7.0
assign (resid 6 and name HG2*) (resid 7 and name HE*)	0 0 7.0
assign (resid 6 and name HN) (resid 8 and name HN)	0 0 5.0
assign (resid 6 and name HA) (resid 8 and name HN)	0 0 5.0
assign (resid 6 and name HB) (resid 8 and name HN)	0 0 3.5
assign (resid 6 and name HG*) (resid 8 and name HN)	0 0 6.5
assign (resid 6 and name HG*) (resid 22 and name HB1)	0 0 8.0
assign (resid 6 and name HG*) (resid 22 and name HB2)	0 0 8.0
assign (resid 6 and name HG1*) (resid 22 and name HE*)	0 0 7.5
assign (resid 6 and name HG2*) (resid 22 and name HE*)	0 0 7.5
assign (resid 6 and name HN) (resid 23 and name HA)	0 0 3.5
assign (resid 7 and name HN) (resid 7 and name HB1)	0 0 3.5
assign (resid 7 and name HN) (resid 7 and name HB2)	0 0 3.5
assign (resid 7 and name HN) (resid 7 and name HD*)	0 0 5.7
assign (resid 7 and name HA) (resid 7 and name HB1)	0 0 3.5
assign (resid 7 and name HA) (resid 7 and name HB2)	0 0 3.5
assign (resid 7 and name HA) (resid 7 and name HD*)	0 0 4.7
assign (resid 7 and name HA) (resid 7 and name HE*)	0 0 5.5
assign (resid 7 and name HB1) (resid 7 and name HB2)	0 0 2.7
assign (resid 7 and name HB1) (resid 7 and name HD*)	0 0 4.7
assign (resid 7 and name HB2) (resid 7 and name HD*)	0 0 4.7
assign (resid 7 and name HB1) (resid 7 and name HE*)	0 0 5.5
assign (resid 7 and name HB2) (resid 7 and name HE*)	0 0 5.5
assign (resid 7 and name HD*) (resid 7 and name HE*)	0 0 6.7
assign (resid 7 and name HB1) (resid 8 and name HN)	0 0 3.5
assign (resid 7 and name HB2) (resid 8 and name HN)	0 0 2.7
assign (resid 7 and name HD*) (resid 8 and name HN)	0 0 5.5
assign (resid 8 and name HN) (resid 8 and name HB)	0 0 3.5
assign (resid 8 and name HN) (resid 8 and name HG11)	0 0 2.7
assign (resid 8 and name HN) (resid 8 and name HG12)	0 0 2.7
assign (resid 8 and name HN) (resid 8 and name HG2*)	0 0 4.2
assign (resid 8 and name HN) (resid 8 and name HD1*)	0 0 5.0
assign (resid 8 and name HA) (resid 8 and name HB)	0 0 2.7
assign (resid 8 and name HA) (resid 8 and name HG11)	0 0 5.0
assign (resid 8 and name HA) (resid 8 and name HG12)	0 0 5.0
assign (resid 8 and name HA) (resid 8 and name HG2*)	0 0 4.2
assign (resid 8 and name HA) (resid 8 and name HD1*)	0 0 5.0
assign (resid 8 and name HB) (resid 8 and name HG11)	0 0 3.5

assign (resid 8 and name HB) (resid 8 and name HG12)	0 0 3.5
assign (resid 8 and name HB) (resid 8 and name HG2*)	0 0 4.2
assign (resid 8 and name HB) (resid 8 and name HD1*)	0 0 4.2
assign (resid 8 and name HN) (resid 9 and name HD1)	0 0 5.0
assign (resid 8 and name HN) (resid 9 and name HD2)	0 0 5.0
assign (resid 8 and name HA) (resid 9 and name HD1)	0 0 2.7
assign (resid 8 and name HA) (resid 9 and name HD2)	0 0 2.7
assign (resid 8 and name HG11) (resid 9 and name HD1)	0 0 5.0
assign (resid 8 and name HG12) (resid 9 and name HD1)	0 0 5.0
assign (resid 8 and name HG11) (resid 9 and name HD2)	0 0 5.0
assign (resid 8 and name HG12) (resid 9 and name HD2)	0 0 5.0
assign (resid 8 and name HG2*) (resid 9 and name HD2)	0 0 4.2
assign (resid 8 and name HD1*) (resid 9 and name HD1)	0 0 6.5
assign (resid 8 and name HD1*) (resid 9 and name HD2)	0 0 6.5
assign (resid 9 and name HA) (resid 9 and name HB1)	0 0 2.7
assign (resid 9 and name HA) (resid 9 and name HB2)	0 0 2.7
assign (resid 9 and name HA) (resid 9 and name HG*)	0 0 4.5
assign (resid 9 and name HA) (resid 9 and name HD1)	0 0 5.0
assign (resid 9 and name HA) (resid 9 and name HD2)	0 0 5.0
assign (resid 9 and name HB1) (resid 9 and name HB2)	0 0 2.7
assign (resid 9 and name HB1) (resid 9 and name HG1)	0 0 3.5
assign (resid 9 and name HB1) (resid 9 and name HG2)	0 0 3.5
assign (resid 9 and name HB1) (resid 9 and name HD1)	0 0 5.0
assign (resid 9 and name HB1) (resid 9 and name HD2)	0 0 5.0
assign (resid 9 and name HB2) (resid 9 and name HG1)	0 0 3.5
assign (resid 9 and name HB2) (resid 9 and name HG2)	0 0 3.5
assign (resid 9 and name HB2) (resid 9 and name HD1)	0 0 5.0
assign (resid 9 and name HB2) (resid 9 and name HD2)	0 0 5.0
assign (resid 9 and name HG1) (resid 9 and name HG2)	0 0 3.5
assign (resid 9 and name HG1) (resid 9 and name HD1)	0 0 3.5
assign (resid 9 and name HG1) (resid 9 and name HD2)	0 0 3.5
assign (resid 9 and name HG2) (resid 9 and name HD1)	0 0 3.5
assign (resid 9 and name HG2) (resid 9 and name HD2)	0 0 3.5
assign (resid 9 and name HD1) (resid 9 and name HD2)	0 0 2.7
assign (resid 9 and name HA) (resid 10 and name HN)	0 0 2.7
assign (resid 9 and name HB1) (resid 10 and name HN)	0 0 3.5
assign (resid 9 and name HB2) (resid 10 and name HN)	0 0 3.5
assign (resid 10 and name HN) (resid 10 and name HA)	0 0 3.5
assign (resid 10 and name HN) (resid 10 and name HB1)	0 0 5.0
assign (resid 10 and name HN) (resid 10 and name HB2)	0 0 3.5
assign (resid 10 and name HA) (resid 10 and name HB1)	0 0 3.5
assign (resid 10 and name HA) (resid 10 and name HB2)	0 0 3.5
assign (resid 10 and name HB1) (resid 10 and name HB2)	0 0 2.7
assign (resid 10 and name HB1) (resid 11 and name HN)	0 0 3.5
assign (resid 10 and name HB2) (resid 11 and name HN)	0 0 3.5
assign (resid 10 and name HA) (resid 12 and name HN)	0 0 5.0
assign (resid 10 and name HB1) (resid 12 and name HN)	0 0 5.0
assign (resid 10 and name HB2) (resid 12 and name HN)	0 0 6.0
assign (resid 10 and name HB1) (resid 12 and name HB1)	0 0 5.0
assign (resid 10 and name HN) (resid 13 and name HG2*)	0 0 6.5
assign (resid 10 and name HB1) (resid 13 and name HN)	0 0 5.0
assign (resid 10 and name HB2) (resid 13 and name HN)	0 0 5.0
assign (resid 10 and name HN) (resid 17 and name HB1)	0 0 6.0
assign (resid 10 and name HN) (resid 17 and name HB2)	0 0 6.0
assign (resid 10 and name HB1) (resid 24 and name HA)	0 0 6.0
assign (resid 11 and name HN) (resid 11 and name HA)	0 0 3.5

assign (resid 11 and name HN) (resid 11 and name HB)	0 0 3.5
assign (resid 11 and name HN) (resid 11 and name HG11)	0 0 2.7
assign (resid 11 and name HN) (resid 11 and name HG12)	0 0 2.7
assign (resid 11 and name HN) (resid 11 and name HG2*)	0 0 5.0
assign (resid 11 and name HN) (resid 11 and name HD1*)	0 0 5.0
assign (resid 11 and name HA) (resid 11 and name HB)	0 0 3.5
assign (resid 11 and name HA) (resid 11 and name HG11)	0 0 3.5
assign (resid 11 and name HA) (resid 11 and name HG12)	0 0 3.5
assign (resid 11 and name HA) (resid 11 and name HG2*)	0 0 4.2
assign (resid 11 and name HA) (resid 11 and name HD1*)	0 0 4.2
assign (resid 11 and name HB) (resid 11 and name HG11)	0 0 3.5
assign (resid 11 and name HB) (resid 11 and name HG12)	0 0 3.5
assign (resid 11 and name HB) (resid 11 and name HG2*)	0 0 4.2
assign (resid 11 and name HB) (resid 11 and name HD1*)	0 0 4.2
assign (resid 11 and name HN) (resid 12 and name HN)	0 0 3.5
assign (resid 11 and name HA) (resid 12 and name HN)	0 0 3.5
assign (resid 11 and name HB) (resid 12 and name HN)	0 0 5.0
assign (resid 11 and name HG1*) (resid 12 and name HN)	0 0 7.0
assign (resid 11 and name HG2*) (resid 12 and name HN)	0 0 5.0
assign (resid 11 and name HA) (resid 13 and name HN)	0 0 3.5
assign (resid 11 and name HG2*) (resid 13 and name HN)	0 0 5.0
assign (resid 11 and name HD1*) (resid 13 and name HN)	0 0 7.5
assign (resid 12 and name HN) (resid 12 and name HB1)	0 0 3.5
assign (resid 12 and name HN) (resid 12 and name HB2)	0 0 3.5
assign (resid 12 and name HA) (resid 12 and name HB1)	0 0 3.5
assign (resid 12 and name HA) (resid 12 and name HB2)	0 0 3.5
assign (resid 12 and name HB1) (resid 12 and name HB2)	0 0 2.7
assign (resid 12 and name HA) (resid 13 and name HN)	0 0 2.7
assign (resid 12 and name HB1) (resid 13 and name HN)	0 0 5.0
assign (resid 12 and name HB2) (resid 13 and name HN)	0 0 5.0
assign (resid 12 and name HA) (resid 14 and name HB2)	0 0 3.5
assign (resid 12 and name HA) (resid 15 and name HN)	0 0 6.0
assign (resid 12 and name HB1) (resid 15 and name HN)	0 0 5.0
assign (resid 12 and name HB1) (resid 15 and name HB*)	0 0 4.5
assign (resid 12 and name HB1) (resid 15 and name HG)	0 0 6.0
assign (resid 12 and name HB1) (resid 15 and name HD*)	0 0 8.0
assign (resid 12 and name HB2) (resid 15 and name HN)	0 0 5.0
assign (resid 12 and name HB2) (resid 15 and name HB*)	0 0 4.5
assign (resid 12 and name HB2) (resid 15 and name HG)	0 0 6.0
assign (resid 12 and name HB2) (resid 15 and name HD*)	0 0 8.0
assign (resid 12 and name HB1) (resid 17 and name HB1)	0 0 5.0
assign (resid 12 and name HB1) (resid 17 and name HB2)	0 0 5.0
assign (resid 12 and name HB2) (resid 17 and name HB1)	0 0 5.0
assign (resid 12 and name HB2) (resid 17 and name HB2)	0 0 5.0
assign (resid 13 and name HN) (resid 13 and name HG2*)	0 0 4.2
assign (resid 13 and name HN) (resid 14 and name HN)	0 0 2.7
assign (resid 13 and name HN) (resid 14 and name HG)	0 0 3.5
assign (resid 13 and name HG2*) (resid 14 and name HN)	0 0 5.0
assign (resid 13 and name HN) (resid 17 and name HB1)	0 0 6.0
assign (resid 13 and name HN) (resid 17 and name HB2)	0 0 6.0
assign (resid 13 and name HA) (resid 17 and name HN)	0 0 3.5
assign (resid 14 and name HA) (resid 14 and name HB1)	0 0 3.5
assign (resid 14 and name HA) (resid 14 and name HB2)	0 0 3.5
assign (resid 14 and name HA) (resid 14 and name HG)	0 0 3.5
assign (resid 14 and name HA) (resid 14 and name HD1*)	0 0 5.0
assign (resid 14 and name HA) (resid 14 and name HD2*)	0 0 5.0

assign (resid 14 and name HB1) (resid 14 and name HN)	0 0 3.5
assign (resid 14 and name HB2) (resid 14 and name HN)	0 0 3.5
assign (resid 14 and name HG) (resid 14 and name HN)	0 0 3.5
assign (resid 14 and name HD*) (resid 14 and name HN)	0 0 6.5
assign (resid 14 and name HN) (resid 15 and name HN)	0 0 3.5
assign (resid 14 and name HA) (resid 15 and name HN)	0 0 3.5
assign (resid 14 and name HN) (resid 16 and name HN)	0 0 5.0
assign (resid 15 and name HN) (resid 15 and name HA)	0 0 3.5
assign (resid 15 and name HN) (resid 15 and name HB*)	0 0 3.7
assign (resid 15 and name HN) (resid 15 and name HD1*)	0 0 5.0
assign (resid 15 and name HN) (resid 15 and name HD2*)	0 0 5.0
assign (resid 15 and name HA) (resid 15 and name HB*)	0 0 3.7
assign (resid 15 and name HA) (resid 15 and name HG)	0 0 3.5
assign (resid 15 and name HA) (resid 15 and name HD1*)	0 0 5.0
assign (resid 15 and name HA) (resid 15 and name HD2*)	0 0 5.0
assign (resid 15 and name HN) (resid 16 and name HN)	0 0 2.7
assign (resid 15 and name HA) (resid 16 and name HN)	0 0 3.5
assign (resid 15 and name HB*) (resid 16 and name HN)	0 0 4.5
assign (resid 15 and name HD*) (resid 16 and name HN)	0 0 6.5
assign (resid 15 and name HB*) (resid 26 and name HD1)	0 0 4.5
assign (resid 15 and name HB*) (resid 26 and name HD2)	0 0 4.5
assign (resid 15 and name HD1*) (resid 26 and name HD1)	0 0 6.5
assign (resid 15 and name HD1*) (resid 26 and name HD2)	0 0 6.5
assign (resid 15 and name HD2*) (resid 26 and name HD1)	0 0 6.5
assign (resid 15 and name HD2*) (resid 26 and name HD2)	0 0 6.5
assign (resid 15 and name HD*) (resid 27 and name HD21)	0 0 6.5
assign (resid 15 and name HD*) (resid 27 and name HD22)	0 0 6.5
assign (resid 16 and name HN) (resid 16 and name HA*)	0 0 3.7
assign (resid 16 and name HN) (resid 17 and name HN)	0 0 2.7
assign (resid 16 and name HA*) (resid 17 and name HN)	0 0 3.7
assign (resid 17 and name HN) (resid 17 and name HA)	0 0 2.7
assign (resid 17 and name HN) (resid 17 and name HB1)	0 0 3.5
assign (resid 17 and name HN) (resid 17 and name HB2)	0 0 3.5
assign (resid 17 and name HA) (resid 17 and name HB1)	0 0 3.5
assign (resid 17 and name HA) (resid 17 and name HB2)	0 0 3.5
assign (resid 17 and name HB1) (resid 17 and name HB2)	0 0 2.7
assign (resid 17 and name HB1) (resid 18 and name HN)	0 0 3.5
assign (resid 17 and name HB2) (resid 18 and name HN)	0 0 3.5
assign (resid 17 and name HB*) (resid 24 and name HB1)	0 0 3.7
assign (resid 17 and name HB*) (resid 24 and name HB2)	0 0 3.7
assign (resid 18 and name HA) (resid 18 and name HB*)	0 0 3.7
assign (resid 18 and name HN) (resid 19 and name HN)	0 0 5.0
assign (resid 18 and name HB*) (resid 19 and name HN)	0 0 4.5
assign (resid 18 and name HB*) (resid 27 and name HN)	0 0 6.0
assign (resid 19 and name HN) (resid 19 and name HA)	0 0 2.7
assign (resid 19 and name HN) (resid 19 and name HB*)	0 0 3.7
assign (resid 19 and name HA) (resid 19 and name HB*)	0 0 3.7
assign (resid 19 and name HN) (resid 20 and name HN)	0 0 5.0
assign (resid 19 and name HN) (resid 24 and name HA)	0 0 3.5
assign (resid 20 and name HN) (resid 20 and name HB*)	0 0 3.7
assign (resid 20 and name HN) (resid 20 and name HG*)	0 0 4.5
assign (resid 20 and name HN) (resid 20 and name HD1)	0 0 3.5
assign (resid 20 and name HN) (resid 20 and name HD2)	0 0 3.5
assign (resid 20 and name HN) (resid 20 and name HE*)	0 0 6.0
assign (resid 20 and name HA) (resid 20 and name HB*)	0 0 3.7
assign (resid 20 and name HA) (resid 20 and name HG*)	0 0 4.5



assign (resid 20 and name HA) (resid 20 and name HD1)	0 0 5.0
assign (resid 20 and name HA) (resid 20 and name HD2)	0 0 5.0
assign (resid 20 and name HA) (resid 20 and name HE*)	0 0 6.0
assign (resid 20 and name HB*) (resid 20 and name HG*)	0 0 4.7
assign (resid 20 and name HB*) (resid 20 and name HD1)	0 0 3.7
assign (resid 20 and name HB*) (resid 20 and name HD2)	0 0 3.7
assign (resid 20 and name HG*) (resid 20 and name HE*)	0 0 5.5
assign (resid 20 and name HD1) (resid 20 and name HE*)	0 0 4.5
assign (resid 20 and name HD2) (resid 20 and name HE*)	0 0 4.5
assign (resid 20 and name HN) (resid 21 and name HN)	0 0 5.0
assign (resid 20 and name HA) (resid 21 and name HN)	0 0 3.5
assign (resid 20 and name HB*) (resid 21 and name HN)	0 0 4.5
assign (resid 20 and name HG*) (resid 21 and name HN)	0 0 4.5
assign (resid 20 and name HD1) (resid 21 and name HN)	0 0 5.0
assign (resid 20 and name HD2) (resid 21 and name HN)	0 0 5.0
assign (resid 20 and name HN) (resid 22 and name HN)	0 0 5.0
assign (resid 20 and name HN) (resid 22 and name HA)	0 0 6.0
assign (resid 20 and name HA) (resid 22 and name HN)	0 0 3.5
assign (resid 20 and name HN) (resid 23 and name HN)	0 0 3.5
assign (resid 20 and name HN) (resid 23 and name HG*)	0 0 9
assign (resid 20 and name HN) (resid 24 and name HA)	0 0 2.7
assign (resid 20 and name HN) (resid 25 and name HD*)	0 0 4.7
assign (resid 20 and name HN) (resid 25 and name HE*)	0 0 5.5
assign (resid 20 and name HB*) (resid 25 and name HD*)	0 0 5.7
assign (resid 20 and name HG*) (resid 25 and name HD*)	0 0 6.5
assign (resid 20 and name HG*) (resid 25 and name HE*)	0 0 6.5
assign (resid 20 and name HD1) (resid 25 and name HD*)	0 0 5.5
assign (resid 20 and name HD1) (resid 25 and name HE*)	0 0 7.0
assign (resid 20 and name HD2) (resid 25 and name HD*)	0 0 5.5
assign (resid 20 and name HD2) (resid 25 and name HE*)	0 0 7.0
assign (resid 21 and name HN) (resid 21 and name HA)	0 0 3.5
assign (resid 21 and name HN) (resid 21 and name HB1)	0 0 5.0
assign (resid 21 and name HN) (resid 21 and name HB2)	0 0 5.0
assign (resid 21 and name HA) (resid 21 and name HB1)	0 0 2.7
assign (resid 21 and name HA) (resid 21 and name HB2)	0 0 2.7
assign (resid 21 and name HB1) (resid 21 and name HB2)	0 0 2.7
assign (resid 21 and name HN) (resid 22 and name HN)	0 0 2.7
assign (resid 21 and name HN) (resid 22 and name HA)	0 0 5.0
assign (resid 21 and name HA) (resid 22 and name HN)	0 0 3.5
assign (resid 21 and name HB1) (resid 22 and name HN)	0 0 5.0
assign (resid 21 and name HB1) (resid 22 and name HG*)	0 0 6.0
assign (resid 21 and name HB2) (resid 22 and name HN)	0 0 5.0
assign (resid 21 and name HB2) (resid 22 and name HG*)	0 0 6.0
assign (resid 21 and name HN) (resid 23 and name HN)	0 0 6.0
assign (resid 21 and name HA) (resid 23 and name HG*)	0 0 8.0
assign (resid 22 and name HN) (resid 22 and name HA)	0 0 3.5
assign (resid 22 and name HN) (resid 22 and name HB1)	0 0 5.0
assign (resid 22 and name HN) (resid 22 and name HB2)	0 0 5.0
assign (resid 22 and name HN) (resid 22 and name HG*)	0 0 3.7
assign (resid 22 and name HN) (resid 22 and name HD1)	0 0 5.0
assign (resid 22 and name HN) (resid 22 and name HD2)	0 0 5.0
assign (resid 22 and name HN) (resid 22 and name HE*)	0 0 6.0
assign (resid 22 and name HA) (resid 22 and name HB1)	0 0 3.5
assign (resid 22 and name HA) (resid 22 and name HB2)	0 0 3.5
assign (resid 22 and name HA) (resid 22 and name HG*)	0 0 3.7
assign (resid 22 and name HA) (resid 22 and name HD*)	0 0 4.5

assign (resid 22 and name HA) (resid 22 and name HE*)	0 0 6.0
assign (resid 22 and name HB1) (resid 22 and name HG*)	0 0 3.7
assign (resid 22 and name HB1) (resid 22 and name HD1)	0 0 3.5
assign (resid 22 and name HB1) (resid 22 and name HD2)	0 0 3.5
assign (resid 22 and name HB2) (resid 22 and name HG*)	0 0 3.7
assign (resid 22 and name HB2) (resid 22 and name HD1)	0 0 3.5
assign (resid 22 and name HB2) (resid 22 and name HD2)	0 0 3.5
assign (resid 22 and name HE*) (resid 22 and name HB1)	0 0 7.0
assign (resid 22 and name HE*) (resid 22 and name HB2)	0 0 7.0
assign (resid 22 and name HE*) (resid 22 and name HG1)	0 0 3.7
assign (resid 22 and name HE*) (resid 22 and name HG2)	0 0 3.7
assign (resid 22 and name HE*) (resid 22 and name HD1)	0 0 3.7
assign (resid 22 and name HE*) (resid 22 and name HD2)	0 0 3.7
assign (resid 22 and name HN) (resid 23 and name HN)	0 0 2.7
assign (resid 22 and name HB1) (resid 23 and name HN)	0 0 5.0
assign (resid 22 and name HB2) (resid 23 and name HN)	0 0 5.0
assign (resid 22 and name HG*) (resid 23 and name HA)	0 0 7.0
assign (resid 23 and name HN) (resid 23 and name HA)	0 0 3.5
assign (resid 23 and name HN) (resid 23 and name HB)	0 0 2.7
assign (resid 23 and name HN) (resid 23 and name HG*)	0 0 5.7
assign (resid 23 and name HA) (resid 23 and name HB)	0 0 3.5
assign (resid 23 and name HA) (resid 23 and name HG1*)	0 0 4.2
assign (resid 23 and name HA) (resid 23 and name HG2*)	0 0 4.2
assign (resid 23 and name HB) (resid 23 and name HG1*)	0 0 4.2
assign (resid 23 and name HB) (resid 23 and name HG2*)	0 0 4.2
assign (resid 23 and name HA) (resid 24 and name HN)	0 0 2.7
assign (resid 23 and name HB) (resid 24 and name HN)	0 0 5.0
assign (resid 23 and name HG*) (resid 24 and name HA)	0 0 8.0
assign (resid 23 and name HG*) (resid 24 and name HB1)	0 0 8.0
assign (resid 23 and name HG*) (resid 24 and name HB2)	0 0 6.5
assign (resid 23 and name HG1*) (resid 24 and name HN)	0 0 5.0
assign (resid 23 and name HG2*) (resid 24 and name HN)	0 0 5.0
assign (resid 23 and name HA) (resid 25 and name HE*)	0 0 7.0
assign (resid 23 and name HB) (resid 25 and name HE*)	0 0 4.7
assign (resid 23 and name HG*) (resid 25 and name HA)	0 0 9.0
assign (resid 23 and name HG*) (resid 25 and name HD*)	0 0 7.7
assign (resid 23 and name HG*) (resid 25 and name HE*)	0 0 6.2
assign (resid 24 and name HN) (resid 24 and name HA)	0 0 3.5
assign (resid 24 and name HN) (resid 24 and name HB1)	0 0 3.5
assign (resid 24 and name HN) (resid 24 and name HB2)	0 0 3.5
assign (resid 24 and name HA) (resid 24 and name HB2)	0 0 3.5
assign (resid 24 and name HA) (resid 24 and name HB1)	0 0 3.5
assign (resid 24 and name HB1) (resid 24 and name HB2)	0 0 2.7
assign (resid 24 and name HN) (resid 25 and name HN)	0 0 5.0
assign (resid 24 and name HN) (resid 25 and name HD*)	0 0 5.5
assign (resid 24 and name HN) (resid 25 and name HE*)	0 0 5.5
assign (resid 24 and name HA) (resid 25 and name HN)	0 0 3.5
assign (resid 24 and name HA) (resid 25 and name HE*)	0 0 7.0
assign (resid 24 and name HB*) (resid 25 and name HN)	0 0 4.5
assign (resid 24 and name HB*) (resid 25 and name HA)	0 0 4.5
assign (resid 25 and name HN) (resid 25 and name HA)	0 0 3.5
assign (resid 25 and name HN) (resid 25 and name HB*)	0 0 4.5
assign (resid 25 and name HN) (resid 25 and name HD*)	0 0 4.7
assign (resid 25 and name HN) (resid 25 and name HE*)	0 0 7.0
assign (resid 25 and name HA) (resid 25 and name HB1)	0 0 3.5
assign (resid 25 and name HA) (resid 25 and name HB2)	0 0 3.5

assign (resid 25 and name HA) (resid 25 and name HE*)	0 0 5.5
assign (resid 25 and name HA) (resid 25 and name HD*)	0 0 4.7
assign (resid 25 and name HB1) (resid 25 and name HB2)	0 0 2.7
assign (resid 25 and name HB1) (resid 25 and name HD*)	0 0 4.7
assign (resid 25 and name HB1) (resid 25 and name HE*)	0 0 5.5
assign (resid 25 and name HB2) (resid 25 and name HD*)	0 0 4.7
assign (resid 25 and name HB2) (resid 25 and name HE*)	0 0 5.5
assign (resid 25 and name HD*) (resid 25 and name HE*)	0 0 6.7
assign (resid 25 and name HN) (resid 26 and name HB*)	0 0 6.0
assign (resid 25 and name HA) (resid 26 and name HN)	0 0 2.7
assign (resid 25 and name HA) (resid 26 and name HB*)	0 0 6.0
assign (resid 25 and name HA) (resid 26 and name HG1)	0 0 5.0
assign (resid 25 and name HA) (resid 26 and name HG2)	0 0 5.0
assign (resid 25 and name HA) (resid 26 and name HE)	0 0 6.0
assign (resid 25 and name HB1) (resid 26 and name HN)	0 0 3.5
assign (resid 25 and name HB2) (resid 26 and name HN)	0 0 3.5
assign (resid 25 and name HB*) (resid 26 and name HA)	0 0 6.0
assign (resid 25 and name HD*) (resid 26 and name HN)	0 0 5.5
assign (resid 25 and name HE*) (resid 26 and name HN)	0 0 7.0
assign (resid 25 and name HB1) (resid 28 and name HN)	0 0 5.0
assign (resid 25 and name HB1) (resid 28 and name HA*)	0 0 6.0
assign (resid 25 and name HB2) (resid 28 and name HN)	0 0 5.0
assign (resid 25 and name HD*) (resid 28 and name HN)	0 0 7.0
assign (resid 25 and name HB1) (resid 29 and name HA)	0 0 3.5
assign (resid 25 and name HB2) (resid 29 and name HA)	0 0 3.5
assign (resid 25 and name HD*) (resid 29 and name HA)	0 0 5.5
assign (resid 25 and name HD*) (resid 29 and name HB)	0 0 7.0
assign (resid 25 and name HE*) (resid 29 and name HA)	0 0 5.5
assign (resid 25 and name HD*) (resid 29 and name HG*)	0 0 10.0
assign (resid 25 and name HE*) (resid 29 and name HG*)	0 0 8.5
assign (resid 25 and name HB1) (resid 30 and name HN)	0 0 3.5
assign (resid 25 and name HB2) (resid 30 and name HN)	0 0 3.5
assign (resid 25 and name HD*) (resid 30 and name HN)	0 0 5.5
assign (resid 25 and name HA) (resid 31 and name HA)	0 0 3.5
assign (resid 25 and name HD*) (resid 31 and name HA)	0 0 4.7
assign (resid 25 and name HD*) (resid 31 and name HD1)	0 0 4.7
assign (resid 25 and name HD*) (resid 31 and name HD2)	0 0 5.5
assign (resid 25 and name HD*) (resid 31 and name HG*)	0 0 5.7
assign (resid 25 and name HE*) (resid 31 and name HA)	0 0 4.7
assign (resid 25 and name HE*) (resid 31 and name HG*)	0 0 5.7
assign (resid 25 and name HE*) (resid 31 and name HD1)	0 0 7.0
assign (resid 25 and name HE*) (resid 31 and name HD2)	0 0 7.0
assign (resid 26 and name HN) (resid 26 and name HA)	0 0 3.5
assign (resid 26 and name HN) (resid 26 and name HB*)	0 0 4.5
assign (resid 26 and name HN) (resid 26 and name HG1)	0 0 2.7
assign (resid 26 and name HG2) (resid 26 and name HN)	0 0 2.7
assign (resid 26 and name HN) (resid 26 and name HD1)	0 0 5.0
assign (resid 26 and name HN) (resid 26 and name HD2)	0 0 5.0
assign (resid 26 and name HN) (resid 26 and name HE)	0 0 6.0
assign (resid 26 and name HA) (resid 26 and name HB*)	0 0 3.7
assign (resid 26 and name HA) (resid 26 and name HG1)	0 0 5.0
assign (resid 26 and name HA) (resid 26 and name HG2)	0 0 5.0
assign (resid 26 and name HA) (resid 26 and name HD1)	0 0 6.0
assign (resid 26 and name HA) (resid 26 and name HD2)	0 0 6.0
assign (resid 26 and name HA) (resid 26 and name HE)	0 0 5.0
assign (resid 26 and name HB*) (resid 26 and name HG1)	0 0 3.7

assign (resid 26 and name HB*) (resid 26 and name HG2)	0 0 3.7
assign (resid 26 and name HB*) (resid 26 and name HD1)	0 0 3.7
assign (resid 26 and name HB*) (resid 26 and name HD2)	0 0 3.7
assign (resid 26 and name HB*) (resid 26 and name HE)	0 0 3.7
assign (resid 26 and name HG1) (resid 26 and name HD1)	0 0 3.5
assign (resid 26 and name HG1) (resid 26 and name HD2)	0 0 3.5
assign (resid 26 and name HG1) (resid 26 and name HE)	0 0 3.5
assign (resid 26 and name HG2) (resid 26 and name HD1)	0 0 3.5
assign (resid 26 and name HG2) (resid 26 and name HD2)	0 0 3.5
assign (resid 26 and name HG2) (resid 26 and name HE)	0 0 3.5
assign (resid 26 and name HD1) (resid 26 and name HD2)	0 0 3.5
assign (resid 26 and name HD1) (resid 26 and name HE)	0 0 3.5
assign (resid 26 and name HD1) (resid 26 and name HE)	0 0 3.5
assign (resid 26 and name HN) (resid 27 and name HN)	0 0 5.0
assign (resid 26 and name HA) (resid 27 and name HN)	0 0 2.7
assign (resid 26 and name HB*) (resid 27 and name HN)	0 0 3.7
assign (resid 26 and name HG1) (resid 27 and name HN)	0 0 5.0
assign (resid 26 and name HG2) (resid 27 and name HN)	0 0 5.0
assign (resid 26 and name HD*) (resid 27 and name HN)	0 0 6.0
assign (resid 26 and name HN) (resid 28 and name HN)	0 0 5.0
assign (resid 26 and name HA) (resid 28 and name HN)	0 0 3.5
assign (resid 26 and name HG1) (resid 28 and name HN)	0 0 6.0
assign (resid 26 and name HG2) (resid 28 and name HN)	0 0 6.0
assign (resid 26 and name HN) (resid 29 and name HN)	0 0 6.0
assign (resid 26 and name HN) (resid 29 and name HA)	0 0 5.0
assign (resid 26 and name HN) (resid 30 and name HN)	0 0 3.5
assign (resid 26 and name HN) (resid 30 and name HB)	0 0 2.7
assign (resid 26 and name HN) (resid 30 and name HG2*)	0 0 5.0
assign (resid 26 and name HA) (resid 30 and name HN)	0 0 0.5
assign (resid 26 and name HG*) (resid 30 and name HN)	0 0 4.5
assign (resid 26 and name HN) (resid 31 and name HD*)	0 0 7.0
assign (resid 27 and name HN) (resid 27 and name HB1)	0 0 5.0
assign (resid 27 and name HN) (resid 27 and name HB2)	0 0 5.0
assign (resid 27 and name HB1) (resid 27 and name HB2)	0 0 2.7
assign (resid 27 and name HN) (resid 28 and name HN)	0 0 2.7
assign (resid 27 and name HB1) (resid 28 and name HN)	0 0 3.5
assign (resid 27 and name HB2) (resid 28 and name HN)	0 0 3.5
assign (resid 27 and name HB1) (resid 29 and name HG*)	0 0 8.0
assign (resid 27 and name HB2) (resid 29 and name HG*)	0 0 8.0
assign (resid 27 and name HN) (resid 30 and name HB)	0 0 6.0
assign (resid 27 and name HB1) (resid 30 and name HN)	0 0 5.0
assign (resid 27 and name HB1) (resid 30 and name HG2*)	0 0 5.0
assign (resid 27 and name HB2) (resid 30 and name HN)	0 0 5.0
assign (resid 27 and name HB2) (resid 30 and name HG2*)	0 0 5.0
assign (resid 28 and name HN) (resid 28 and name HA*)	0 0 3.7
assign (resid 28 and name HA1) (resid 28 and name HA2)	0 0 2.7
assign (resid 28 and name HN) (resid 29 and name HN)	0 0 2.7
assign (resid 28 and name HN) (resid 29 and name HG1*)	0 0 6.5
assign (resid 28 and name HN) (resid 29 and name HG2*)	0 0 6.5
assign (resid 28 and name HA1) (resid 29 and name HN)	0 0 3.5
assign (resid 28 and name HA*) (resid 29 and name HA)	0 0 6.0
assign (resid 28 and name HA1) (resid 29 and name HG1*)	0 0 6.5
assign (resid 28 and name HA1) (resid 29 and name HG2*)	0 0 6.5
assign (resid 28 and name HA2) (resid 29 and name HN)	0 0 3.5
assign (resid 28 and name HA2) (resid 29 and name HG1*)	0 0 6.5
assign (resid 28 and name HA2) (resid 29 and name HG2*)	0 0 6.5

assign (resid 29 and name HN) (resid 29 and name HA)	0 0 3.5
assign (resid 29 and name HN) (resid 29 and name HB)	0 0 5.0
assign (resid 29 and name HN) (resid 29 and name HG1*)	0 0 5.0
assign (resid 29 and name HN) (resid 29 and name HG2*)	0 0 5.0
assign (resid 29 and name HA) (resid 29 and name HB)	0 0 2.7
assign (resid 29 and name HA) (resid 29 and name HG1*)	0 0 4.2
assign (resid 29 and name HA) (resid 29 and name HG2*)	0 0 4.2
assign (resid 29 and name HB) (resid 29 and name HG1*)	0 0 4.2
assign (resid 29 and name HB) (resid 29 and name HG2*)	0 0 4.2
assign (resid 29 and name HN) (resid 30 and name HN)	0 0 2.7
assign (resid 29 and name HA) (resid 30 and name HN)	0 0 3.5
assign (resid 29 and name HB) (resid 30 and name HN)	0 0 5.0
assign (resid 29 and name HG1*) (resid 30 and name HN)	0 0 5.0
assign (resid 29 and name HG2*) (resid 30 and name HN)	0 0 5.0
assign (resid 29 and name HG*) (resid 30 and name HB)	0 0 8.0
assign (resid 30 and name HN) (resid 30 and name HA)	0 0 3.5
assign (resid 30 and name HN) (resid 30 and name HB)	0 0 2.7
assign (resid 30 and name HN) (resid 30 and name HG11)	0 0 3.5
assign (resid 30 and name HN) (resid 30 and name HG12)	0 0 3.5
assign (resid 30 and name HN) (resid 30 and name HG2*)	0 0 5.0
assign (resid 30 and name HA) (resid 30 and name HB)	0 0 3.5
assign (resid 30 and name HA) (resid 30 and name HG11)	0 0 5.0
assign (resid 30 and name HA) (resid 30 and name HG12)	0 0 5.0
assign (resid 30 and name HA) (resid 30 and name HD1*)	0 0 4.2
assign (resid 30 and name HA) (resid 30 and name HG2*)	0 0 4.2
assign (resid 30 and name HB) (resid 30 and name HG11)	0 0 3.5
assign (resid 30 and name HB) (resid 30 and name HG12)	0 0 3.
assign (resid 30 and name HB) (resid 30 and name HG2*)	0 0 4.2
assign (resid 30 and name HN) (resid 31 and name HD*)	0 0 6.0
assign (resid 30 and name HA) (resid 31 and name HD*)	0 0 3.7
assign (resid 30 and name HG2*) (resid 31 and name HD1)	0 0 4.2
assign (resid 30 and name HG2*) (resid 31 and name HD2)	0 0 4.2
assign (resid 31 and name HA) (resid 31 and name HB1)	0 0 3.5
assign (resid 31 and name HA) (resid 31 and name HB2)	0 0 3.5
assign (resid 31 and name HA) (resid 31 and name HG1)	0 0 5.0
assign (resid 31 and name HA) (resid 31 and name HG2)	0 0 5.0
assign (resid 31 and name HA) (resid 31 and name HD*)	0 0 4.5
assign (resid 31 and name HB1) (resid 31 and name HD1)	0 0 5.0
assign (resid 31 and name HB1) (resid 31 and name HD2)	0 0 5.0
assign (resid 31 and name HB2) (resid 31 and name HD1)	0 0 5.0
assign (resid 31 and name HB2) (resid 31 and name HD2)	0 0 5.0
assign (resid 31 and name HG1) (resid 31 and name HD1)	0 0 3.5
assign (resid 31 and name HG1) (resid 31 and name HD2)	0 0 3.5
assign (resid 31 and name HG2) (resid 31 and name HD1)	0 0 3.5
assign (resid 31 and name HG2) (resid 31 and name HD2)	0 0 3.5
assign (resid 31 and name HD1) (resid 31 and name HD2)	0 0 2.7

## APPENDIX VII

List of dihedral restraints and hydrogen bond definitions applied in the simulated annealing protocol for the determination of the structure of circulin B in the aqueous environment (Chapter 7). The numbering of the residues is according to that used for the first collective alignment of members of the inhibitor cystine knot family (Pallaghy *et al.*, 1994).

### PHI ANGLES

phi angle	range	basis of application	residues
-120°	± 30°	$^3J_{\text{NH-H}\alpha}$ within 8.0 - 9.5 Hz	8,11,18,23,24,26,29
-120°	± 15°	$^3J_{\text{NH-H}\alpha} > 9.5$ Hz	3,5,7,15,20,25
-65°	± 15°	$^3J_{\text{NH-H}\alpha} < 5.8$ Hz	10
-100	± 80°	intraresidue H $^{\alpha}$ -H $_N$ NOE is weaker than the sequential H $^{\alpha}$ -H $_N$ NOE	6
50°	± 40°	intense intraresidue H $^{\alpha}$ -H $_N$ NOE and $^3J_{\text{NH-H}\alpha} \sim 7$ Hz	21,22

### CHI 1 ANGLES

chi 1	range	residues
-60	± 30°	7,10,24,14,25
60	± 30°	-
180	± 30°	6,11,12,23,27,30

### HYDROGEN BONDS

CO Residue	NH Residue
3	24
18	25
20	23
24	1
25	18
26	30
30	26

## REFERENCES

Anfinsen, C. B., (1973) Principles that govern the folding of protein chains. *Science* **181**, 223-230.

Atkinson, A. H., Heath, R. L., Simpson, R. J., Clarke, A. E., Anderson, M. A. (1993) Proteinase inhibitors in *Nicotiana glauca* stigmas are derived from a precursor protein which is processed into five homologous inhibitors. *The Plant Cell* **5**, 203-213.

Bradley, E. K., Thomason, J. F., Cohen, F. E., Kosen, P. A., Kuntz, I. D. (1990) Studies of synthetic helical peptides using circular dichroism and nuclear magnetic resonance. *J. Mol. Biol.* **215**, 607-622.

Bax, A., Davis, D. J. (1985) MLEV-17-based two-dimensional homonuclear magnetisation transfer spectroscopy. *J. Magn. Reson.* **65**, 355-360.

Bax, A., Subramanian, S. (1986) Sensitivity enhanced two-dimensional heteronuclear shift correlation NMR spectroscopy. *J. Magn. Reson.* **67**, 565-569.

Bax, A., Lerner, L. J. (1986) Sensitivity-enhanced two-dimensional heteronuclear relay coherence transfer NMR spectroscopy. *J. Magn. Reson.* **69**, 375-380.

Benham, C. J., Jafri, M. S. (1993) Disulfide bonding patterns and protein topologies. *Protein Sci.* **2**, 41-54.

Braunschweiler L., Ernst R. R. (1983) Coherence transfer by isotopic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521-528.



Anfinsen, C. B., (1973) Principles that govern the folding of protein chains. *Science* **181**, 223-230.

Atkinson, A. H., Heath, R. L., Simpson, R. J., Clarke, A. E., Anderson, M. A. (1993) Proteinase inhibitors in *Nicotiana glauca* stigmas are derived from a precursor protein which is processed into five homologous inhibitors. *The Plant Cell* **5**, 203-213.

Bradley, E. K., Thomason, J. F., Cohen, F. E., Kosen, P. A., Kuntz, I. D. (1990) Studies of synthetic helical peptides using circular dichroism and nuclear magnetic resonance. *J. Mol. Biol.* **215**, 607-622.

Bax, A., Davis, D. J. (1985) MLEV-17-based two-dimensional homonuclear magnetisation transfer spectroscopy. *J. Magn. Reson.* **65**, 355-360.

Bax, A., Subramanian, S. (1986) Sensitivity enhanced two-dimensional heteronuclear shift correlation NMR spectroscopy. *J. Magn. Reson.* **67**, 565-569.

Bax, A., Lerner, L. J. (1986) Sensitivity-enhanced two-dimensional heteronuclear relay coherence transfer NMR spectroscopy. *J. Magn. Reson.* **69**, 375-380.

Benham, C. J., Jafri, M. S. (1993) Disulfide bonding patterns and protein topologies. *Protein Sci.* **2**, 41-54.

Braunschweiler L., Ernst R. R. (1983) Coherence transfer by isotopic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521-528.

- Broekaert, W. F., Terras, F. R. G., Cammue, B. P. A., Osborn, R. W. (1995) Plant defensins: novel antimicrobial peptides as components of the host defence system. *Plant Physiol.* **108**, 1353-1358.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187-217.
- Brünger, A. T. (1992) *X-PLOR Manual Version 3.1*, Yale University Press, New Haven, CT.
- Bystrov, V. F. (1976) Spin-spin coupling and the conformational states of peptide systems. *Prog. NMR. spectrosc.* **10**, 41-81.
- Chang, J. Y., Canals, F., Schindler, P., Querol, E., Aviles, F. X. (1994) The disulphide folding pathway of potato carboxypeptidase inhibitor. *J. Biol Chem.* **269**, 22087-22094.
- Chothia, C. (1992) One thousand families for the molecular biologist. *Nature* **357**, 543-544.
- Claeson, P., Göransson, U., Johansson, S., Lijndijk, T., Bohlin, L. (1998) Fractionation protocol for the isolation of polypeptides from plant biomass. *J. Nat. Prod.* **61**, 77-81.
- Clore, G. M., Nilges, M., Sukuraman, D. K., Brünger, A. T., Karplus, M., Gronenborn, A. M. (1986a) The three-dimensional structure of  $\alpha$ 1l-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and dynamics calculations. *EMBO J.* **5**, 2729-2735.

- Clore, G. M., Brünger, A. T., Karplus, M., Gronenborn, A. M. (1986b) Application of molecular dynamics with interproton distance restraints to three-dimensional protein structure determination. A model study of crambin. *J. Mol. Biol.* **191**, 523-551.
- Clore, G. M., Robien, M. A., Gronenborn, A. M. (1993) Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **231**, 82-102.
- Clubb, R. T., Ferguson, S. B., Walsh, C. T., Wagner, G. (1994) Three-dimensional solution structure of *Escherichia coli* periplasmic cyclophilin. *Biochemistry* **33**, 2761-2722.
- Craik, D. J. (1994) Table of NOE connectivities of kalata-B1, private communication.
- Craik, D. J., Daly, N. L., Bond, T., Waite, W. (1999) Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.* **294**, 1327-1336.
- Craik, D. J., Daly, N. L., Waite, W. (2001) The cystine knot motif in toxins and implications for drug design. *Toxicon* **39**, 43-60.
- Creighton, T. E. (1988) Disulphide bonds and protein stability. *BioEssays* **8**, 57-63.
- Daly, N. L., Love, S., Alewood, P. F., Craik, D. J. (1999) Chemical synthesis and folding of large cyclic polypeptides: studies of the cystine knot polypeptide kalata-B1. *Biochemistry* **38**, 10606-10614.

Daly, N. L., Craik, D. J. (2000) Acyclic permutants of naturally occurring cyclic peptides. *J. Biol. Chem.* **275**, 19068-19075.

de Young, L. R., Burton, L. E., Liu, J. P., Powell, M. F., Schmelzer, C. H., Skelton, N. J. (1996) RhNGF slow unfolding is not due to proline isomerization: possibility of a cystine knot loop-threading mechanism. *Protein Sci.* **5**, 1554-1566.

DeMarco, A., Llinas, M., Wüthrich, K. (1978) Analysis of the <sup>1</sup>H-NMR spectra of ferrichrome peptides I. The non-amide protons. *Biopolymers* **17**, 617-636.

Derua, R., Gustafson, K. R., Pannell, L. K. (1996) Analysis of the disulphide linkage pattern in circulin A and B, HIV-inhibitory macrocyclic peptides. *Biochem. Biophys. Res. Commun.* **228**, 632-638.

De Wit, P. J. G. M. (1995) Fungal avirulence genes and plant resistance genes: unravelling the molecular basis of gene-for-gene interactions. *Advan. Bot. Res.* **21**, 147-185.

Dyson, H. J., Wright, P. E. (1991) Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 519-538.

Favel, A., Mattras, H., Coletti-Previero, M. A., Zwillling, R., Robinson, E. A., Castro, B. (1989) Protease inhibitors from *Ecballium elaterium* seeds. *Int. J. Peptide Protein. Res.* **33**, 202-208.

Flor, H. H. (1971) Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* **9**, 275-296.

Fujikawa, K., Suketa, Y., Hayashi, K., Suzuki, T. (1965) Chemical structure of circulin A. *Experientia* **21**, 307-308.

Gilbert, H. F. (1994) The formation of disulphide bonds. In: "Mechanisms of Protein folding." (ed. R. H. Pain) Oxford University Press, New York. pp 104-130.

Göransson, U., Luijendijk, T., Johansson, S., Bohlin, L., Claeson, P. (1999) Seven novel macrocyclic polypeptides from *Viola arvensis*. *J. Nat. Prod.* **62**, 283-286.

Gran, L. (1970) An oxytocic principal found in *Oldenlandia affinis* DC. *Medd. Nor. Farm. Selsk.* **32**, 173-180.

Gran, L. (1972) 5-hydroxytryptamin (serotonin) isolated from *Oldenlandia affinis*. *Medd. Nor. Farm. Selsk.* **35**, 125-135.

Gran, L. (1973a) On the effect of a polypeptide isolated from Kalata-Kalata (*Oldenlandia affinis* DC) on the oestrogen dominated uterus. *Acta pharmacol. et toxicol.* **33**, 400-408.

Gran, L. (1973b) Oxytocic principalsof *Oldenlandia affinis*. *Lloydia* **36**, 174-178.

Gran, L. (1973c) Isolation of oxytocic peptides from *Oldenlandia affinis* by solvent extraction of tetraphenylborate complexes and chromatography on Sephadex LH-20. *Lloydia* **36**, 207-208.

Gran, L. (1973d) On the isolation of tetramethylputrescine from *Oldenlandia affinis*. *Lloydia* **36**, 209-210.

Griesinger, C., Sørensen, O. W., Ernst, R. R. (1987) Practical aspects of the E. COSY technique, measurement of scalar spin-spin coupling constants in peptides. *J. Magn. Reson.* **75**, 474-492.

Gustafson, K. R., Walton, L. K., Sowder II, R. C., Johnson, D. G., Pannell, L. K., Cardellina II, J. H., Boyd, M. R. (2000) New circulin macrocyclic polypeptides from *Chassalia parvifolia*. *J. Nat. Prod.* **63**, 176-178.

Gustafson, K. R., Sowder II, R. C., Henderson, L. E., Parsons, I. C., Kashman, Y., Cardellina II, J. H., McMahon, J. B., Buckheit Jr., R. W., Pannell, L. K., Boyd, M. R. (1994) Circulins A and B: novel HIV-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*. *J. Am. Chem. Soc.* **116**, 9337-9338.

Hallock, Y. F., Sowder II, R. C., Pannell, L. K., Hughes, C. B., Johnson, D. G., Gulakowski, R., Cardellina II, J. H., Boyd, M. R. (2000) Cycloviolins A-D, anti-HIV macrocyclic polypeptides from *Leonia cymosa*. *J. Org. Chem.* **65**, 124-128.

Hancock, R. E. W., Falla, T., Brown, M. (1995) Cationic Bactericidal Peptides. *Adv. Microb. Physiol.* **37**, 135-175.

Hayashi, K., Suketa, Y., Suzuki, T. (1968) Chemical structure of circulin B. *Experientia* **24**, 656-657.

Holak, T. A., Gondol, D., Otlewski, J., Wilusz, T. (1989b) Determination of the complete three-dimensional structure of the trypsin inhibitor from squash seeds in aqueous solution by nuclear magnetic resonance and a combination of distance geometry and dynamical simulated annealing. *J. Mol. Biol.* **210**, 635-648.

- Holak, T. A., Bode, W., Huber, R., Otlewski, J., Wilusz, T. (1989a) Nuclear magnetic resonance solution and X-ray structures of squash trypsin inhibitor exhibit the same conformation of the proteinase binding loop. *J. Mol. Biol.* **210**, 649-654.
- Hore, P. J. (1983) Solvent suppression in Fourier transform nuclear magnetic resonance. *J. Magn. Reson.* **55**, 283-300.
- Howarth, O. W., Lilley, D. M. J. (1978) Carbon-13 NMR of peptides and proteins. *Prog. NMR Spectr.* **12**, 1-40.
- Hruby, V. (1974) In: "Chemistry and Biochemistry of Peptides." (ed. W. Z. Dekker) vol 3, p16.
- Hutchinson, E. G., Thornton, J. M. (1996) PROMOTIF - A program to identify and analyse structural motifs in proteins. *Protein Sci.* **5**, 212-220.
- Hucho, F. (1995) Toxins as tools in neurochemistry. *Angew. Chem. Int. Ed. Engl.* **34**, 39-50.
- Hyberts, S. G., Goldberg, M. S., Havel, T. S., Wagner, G. (1992) The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736-751.
- Isaacs, N. (1995) Cystine Knots. *Curr. Opin. Struc. Biol.* **5**, 391-395.
- IUPAC-IUB Commission on Biochemical Nomenclature (1970) *J. Mol. Biol.* **52**, 1.
- Jeener, J., Alewaeters, G. (1971) unpublished work; G. Alewaeters Doctoral Thesis, Free University of Brussels.

Jeener, J., Meier, B. H., Bachmann, P., Ernst, R. R. (1979) Investigation of exchange process by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546-4553.

Karplus, M. (1959) Contacts electron spin coupling of nuclear magnetic resonance resonance moments. *J. Chem. Phys.* **30**, 11-15.

Kessler, H., Seip, S. (1994) NMR of peptides. In: "Two-Dimensional NMR Spectroscopy.: Applications for Chemists and Biochemists." (ed. W. R. Coasmun, R. M. K. Carlson) VCH Publishers Inc., U.S.A.

Kessler, H., Konat, R. K., Schmitt, W. (1996) Conformational analysis of peptides: application to drug design. In: "NMR in Drug Design." (ed. D. J. Craik) CRC Press Inc., U.S.A.

King, G. F., Mackay, J. P. (1996) Protein structure determination using NMR spectroscopy. In: "NMR in Drug Design." (ed. D. J. Craik) CRC Press Inc., U.S.A.

Klaus, W., Broger, C., Gerber, P., Senn, H. (1993) Determination of the disulphide bonding pattern in proteins by local and global analysis of nuclear magnetic resonance data. *J. Mol. Biol.* **232**, 897-906.

Kumar, A., Ernst, R. R., Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Comm.* **95**, 1-6.



Laskowski R. A., MacArthur M. W., Moss, D. S., Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structure coordinates. *J. Appl. Crystallogr.* **26**, 283-291.

Liang C., Mislow K. (1994a) Topological chirality of proteins. *J. Am. Chem. Soc.* **116**, 3588-3592.

Liang C., Mislow K. (1994b) Knots in proteins. *J. Am. Chem. Soc.* **116**, 11189-11192.

Liang C., Mislow K. (1995) Topological features of protein structures: knots and links. *J. Am. Chem. Soc.* **117**, 4201-4213.

Liang C., Mislow K. (1996) Knotted structures in chemistry, biochemistry and molecular biology. *Croatica Chemica Acta* **69**, 1385-1403.

Ludvigsen, S., Poulsen, F. M. (1992) Positive theta-angles in proteins by nuclear magnetic resonance spectroscopy. *J. Biomol. NMR* **2**, 227-233.

Narasimhan, L., Singh, J., Humblet, C., Guruprasad, K., Blundell, T. (1994) Snail and spider toxins share a similar tertiary structure and 'cystine knot'. *Nature Struct. Biol.* **1**, 850-852.

Mao, B. (1989) Molecular topology of multiple-disulphide polypeptide chains. *J. Am. Chem. Soc.* **111**, 6132-6136.

Mao, B. (1993) Topological chirality of proteins. *Protein Sci.* **2**, 1057-1059.

Marion, B., Wüthrich K. (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurement of  $^1\text{H}$ - $^1\text{H}$  spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967-974.

McDonald, N. Q., Hendrickson, W., A. (1993) A structural superfamily of growth factors containing a cystine knot motif. *Cell* **73**, 421-424.

Meitinger, T., Meindl, A., Bork, P., Rost, B., Sander, C., Haasemann, M., Murken, J. (1993) Molecular modelling of the Norrie disease protein predicts a cystine knot growth factor tertiary structure. *Nature Genet.* **5**, 376-380.

Merutka, G., Dyson, H. J., Wright, P. E. (1995) "Random Coil"  $^1\text{H}$  chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J. Biomol. NMR* **5**, 14-24.

Murray-Rust, J., McDonald, N. Q., Blundell, T. L., Hosang, M., Oefner, C., Winkler, F., Bradshaw, R. A. (1993) Topological similarities in TGF- $\beta$ 2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure* **1**, 153-159.

Nagayama, K., Wüthrich, K. (1986) Structural interpretation of vicinal proton-proton coupling constants  $^3J_{\text{H}_\alpha\text{-H}_\beta}$  in the basic pancreatic trypsin inhibitor measured by two-dimensional J-resolved NMR spectroscopy. *Eur. J. Biochem.* **115**, 653-657.

Neuhaus, D., Williamson, P. (1989) In: "The Nuclear Overhauser Effect in Structural and Conformational Analysis." VCH Publishers Inc..

Nilges, M., Gronenborn, A. M., Brünger, A. T., Clore, G. M. (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2**, 27-38.

Nielsen, K. J., Alewood, D., Andrews, J., Kent, S. B. H., Craik, D. J. (1994a) An  $^1\text{H}$  NMR determination of the three-dimensional structures of mirror-image forms of a Leu-5 variant of the trypsin inhibitor from *Ecballium elaterium* (EETI-II). *Protein Sci.* **3**, 291-302.

Nielsen, K. J., Heath, R. L., Anderson, M. A., Craik, D. J. (1994b) The three-dimensional solution structure by  $^1\text{H}$  NMR of a 6-kDa proteinase inhibitor isolated from the stigma of *Nicotiana glauca*. *J. Mol. Biol.* **242**, 231-243.

Nielsen, K. J., Heath, R. L., Anderson, M. A., Craik, D. J. (1995) Structures of a series of 6-kDa trypsin inhibitors isolated from the stigma of *Nicotiana glauca*. *Biochemistry* **34**, 14304-14311.

Omecinsky, D. O., Holub, K. E., Adams, M. E., Reilly, M. D. (1996) Three-dimensional structure analysis of  $\mu$ -agatoxins: further evidence for common motifs among neurotoxins with diverse ion channel specificities. *Biochemistry* **35**, 2836-2844.

Pallaghy, P. K., Duggan, B. M., Pennington, M. W., Norton, R. S. (1993) Three-dimensional structure in solution of the calcium channel blocker  $\omega$ -conotoxin. *J. Mol. Biol.* **234**, 405-420.

Pallaghy, P. K., Nielsen, K.J., Craik, D.J., Norton, R. S. (1994) A common structural motif incorporating a cystine knot and a triple-stranded  $\beta$ -sheet in toxic and inhibitory polypeptides. *Protein Sci.* **3**, 1833-1839.

Pardi, A., Billeter, M., Wüthrich, K. (1984) Calibration of the angular dependence of the amide proton- $\text{C}\alpha$  proton coupling constants,  $^3J_{\text{HN}\alpha}$ , in a globular protein. Use of  $^3J_{\text{HN}\alpha}$  for identification of helical secondary structure. *J. Mol. Biol.* **180**, 741-751.

Piotto, M., Saudek, V., Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661-665.

Plateau, P., Gueron, M. (1982) Exchangeable proton NMR without baseline distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.* 104, 7310-7311.

Ramachandran, G. M., Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv. Prot. Chem.* 23, 283-437.

Rance, M., Sørensen, O. W., Bodenhausen G., Wagner G., Ernst, R. R., Wüthrich, K. (1983) Improved spectral resolution in COSY <sup>1</sup>H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* 117, 479-485.

Richardson, J. S. (1981) The anatomy and taxonomy of protein structure. In: "Advances in Protein Chemistry, Volume 34." (ed. C. B. Anfinsen, J. T. Edsall, F. M. Richards) Academic Press, U.S.A. pp 167-339.

Richardson, D. C., Richardson, J. S. (1992) The kinemage: a tool for scientific communication. *Protein Sci.* 1, 3-9.

Shaka, A. J., Freeman, R. (1983) Simplification of NMR spectra by filtration through multiple quantum coherence. *J. Magn. Reson.* 51, 169-173.

Shaka, A. J., Barker, P. B., Freeman, R. (1985) Computer optimised decoupling scheme for wideband applications and low level operation. *J. Magn. Reson.* 64, 547-552.

Saether, O., Craik D. J., Campbell, I. D., Sletten, K., Juul, J., Norman, D. G. (1995) Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide Kalata B1. *Biochemistry* **34**, 4147-4158.

Schill, G. (1971) Catenanes, rotaxanes and knots. In: "Organic Chemistry: A Series of Monographs, Vol. 22." Academi Press, New York.

Schöpke, Th., Hasan Agha, M. I., Kraft, R., Otto, A., Hiller, K. (1993) Hämolytisch aktive Komponenten aus *Viola tricolor* L. und *Viola arvensis* Murray. *Sci. Pharm.* **61**, 144-153.

Sletten, K., Gran, L. (1973) Some molecular properties of kalata peptide B-1. *Medd. Nor. Farm. Selsk.* **35**, 69-82.

Srinivasan, N., Sowdhamini, R., Ramakrishnan, C., Balaram, P. (1990) Conformation of disulphide bridges in proteins. *Int. J. Pept. Protein Res.* **36**, 147-155.

Steudel, R. (1975) Properties of sulfur-sulfur bonds. *Angew. Chem. Int. Ed. Engl.* **14**, 655-664.

Tam, J. P., Lu, Y. (1998) A biomimetic strategy in the synthesis and fragmentation of cyclic protein. *Protein Sci.* **7**, 1583-1592.

Tam, J. P., Lu, Y., Yu, Q. (1999a) Thia zip reaction for synthesis of large cyclic peptides: mechanisms and applications. *J. Am. Chem. Soc.* **121**, 4316-4324.

Tam, J. P., Lu, Y. -A., Yang, J. -L., Chiu, K. -W. (1999b) An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot motif. *Proc. Natl. Acad. Sci. USA* **96**, 8913-8918.

Thornton, J. M. (1981) Disulphide bridges in globular proteins. *J. Mol. Biol.* **151**, 261-287.

Van den Ackerveken, G. F. J. M., Vossen, P., De Wit, P. J. G. M. (1993) The AVR9 race-specific elicitor of *Cladosporium fulvum* is processed by endogenous and plant proteases. *Plant. Physiol.* **103**, 91-96.

Vervoort, J., Van den Hooven, H. W., Berg, A., Vossen, P., Vogelsang, R., Joosten, M. H. A. J., De Wit, P. J. G. M. (1997) The race-specific elicitor AVR9 of the tomato pathogen *Cladosporium fulvum*: a cystine-knot protein. Sequence specific <sup>1</sup>H NMR assignments, secondary structure and global fold of the protein. *FEBS Letters* **404**, 153-158.

Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., Wüthrich, K. (1987) Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611-639.

Walba, D. M. (1985) Topological Stereochemistry. *Tetrahedron* **41**, 3161-3212.

Walton, A. G. (1981) In: "Polypeptides and Protein Structure." Elsevier North Holland Inc., New York, U.S.A.

Williamson, M. P., Havel, T. F., Wüthrich, K. (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by <sup>1</sup>H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295-315.

Witherup, K. M., Bogusky, M. J., Anderson, P. S., Ramjit, H., Ransom, R. W., Wood, T., Sardana, M. (1994) Cyclopsychotride A, a biologically active, 31-residue cyclic peptide isolated from *Psychotria longipes*. *J. Nat. Prod.* **57**, 1619-1625.

Wishart, D. S., Sykes, B. D., Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure using through NMR spectroscopy. *Biochemistry* **31**, 1647-1651

Wishart, D. S., Sykes, B. D. (1994) The  $^{13}\text{C}$  chemical-shift index: a simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data. *J. Biomol. NMR* **4**, 171-180..

Wright, P. E., Dyson, H. J., Lerner, R. A. (1988) Conformations of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. *Biochemistry* **27**, 7167-7175.

Wüthrich, K., Billeter, M., Braun, W. (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949-961.

Wüthrich, K. (1986) In: "NMR of Proteins and nucleic acids." Willey and Sons, New York.

Zuiderweg, E. R., Boelens, R., Kaptein, R. (1985) Stereospecific assignments of  $^1\text{H}$ -NMR methyl lines and conformation of valyl residues in the *lac* repressor headpiece. *Biopolymers*. **24**, 601-611.