

# **INTERACTIONS THAT AFFECT THE STABILITY OF THE $\alpha$ -HELIX**

A thesis submitted to The University of Manchester for the degree of  
Ph.D.  
In the Faculty of Life Sciences

**2005**

*Teuku Mohamad Iqbalsyah*

ProQuest Number: 10757326

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10757326

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

~~Th~~ 26222 ✓



# LIST OF CONTENTS

	Page
TITLE PAGE	1
LIST OF CONTENTS	2
LIST OF TABLES	5
LIST OF FIGURES	6
LIST OF ABBREVIATIONS	8
ABSTRACT	9
DECLARATION	10
COPYRIGHT STATEMENT	10
AKNOWLEDGEMENT	11
DEDICATION	12
 CHAPTER 1. INTRODUCTION TO THE STRUCTURE AND STABILITY OF THE $\alpha$ -HELIX	 13
1.1. Introduction	13
1.2. Structure of $\alpha$ -helix	15
1.2.1. $\alpha$ -Helix nomenclature	16
1.2.2. Helix dihedral angles	17
1.2.3. $\alpha$ -Helix distortion	18
1.2.4. Helix dipole	19
1.3. Structure of $3_{10}$ and $\pi$ -helix	19
1.4. Capping motifs	20
1.5. Factors that stabilise $\alpha$ -helix	24
1.5.1. Amino acids preference at helix interior	24
1.5.2. Amino acids preference at helix termini	28
1.5.3. Non-covalent side-chain interactions	32
1.5.3.1. Salt bridges	32
1.5.3.2. Hydrogen bonds	33
1.5.3.3. Hydrophobic interactions	34
1.5.3.4. Other non-covalent interactions	34
1.5.4. Covalent side-chain interactions	36
1.5.5. Metal binding	37
1.5.6. Phosphorylation	38
1.6. Ionic strength	38
1.7. Temperature	39
1.8. Trifluoroethanol	39
1.9. pKa values	40
1.10. Rotamers	41
1.11. Aims of the thesis	42
1.12. Organisation of the thesis	43
 CHAPTER 2. INTRODUCTION TO HELIX-COIL THEORY	 45
2.1. The helix model.	45
2.2. Zimm-Bragg (ZB) theory	46
2.3. Lifson-Roig (LR) theory	47



2.4. Comparison of ZB and LR theories	48
2.5. Partition function	48
2.6. Modification to the original Lifson-Roig theory	51
2.6.1. Modification to include N- and C- cappings	52
2.6.2. Modification to include side chain interactions	54
2.6.3. Modification to include N1, N2 and N3 preferences	55
2.6.4. Modification to include capping box interaction	59
2.6.5. Modification to include metal cross link	59
2.7. AGADIR.	60
2.8. Helix-coil programs on the web	62
 CHAPTER 3. MATERIALS AND METHODS	 63
3.1. Peptide design	63
3.2. Peptide synthesis and purification	67
3.2.1. Peptide synthesis	67
3.2.2. Acetylation and cleavage	72
3.2.3. Peptide purification	73
3.3. Verification of peptide mass	74
3.4. Peptide concentration determination	76
3.5. Circular Dicroism to determine helix content	76
3.6. pH titration	79
3.7. DTT titration of CAAC peptide	80
3.8. Zinc titration of CAAC peptide	81
3.9. Analysis using helix-coil theory	81
3.10. Protein crystals search	83
3.10.1. Probing the N3 rotamer distribution and energy calculation	83
3.10.2. Probing the CXXC motif	86
3.10.3. Probing the RFM motif	87
3.10.4. Probing the EKE motif	87
 CHAPTER 4. EFFECT OF THE N3 RESIDUE ON THE STABILITY OF THE $\alpha$ -HELIX	 88
4.1. Introduction	88
4.2. Design of N3 peptide	89
4.3. Aggregation tests of N3 peptides	90
4.4. Determination of N3 values ( $n_3$ ) and $\Delta G$ from helix contents	92
4.5. Determination of N3 values ( $n_3$ ) and $\Delta G$ from helix contents with different parameters	95
4.6. Comparison with other reported $n_3$ values	98
4.7. Comparison with helicities calculated using AGADIR	99
4.8. pH titrations of charged amino acids	99
4.9. Correlation of N3 preferences with N3 propensities	103
4.10. Comparison of N3 preferences with N-cap preferences	103
4.11. Comparison of N3 preferences with N1 and N2 preferences	103
4.13. Comparison of N3 preferences with helix interior preferences	106
4.14. Conformational entropy of residues at N3 position	106
4.15. Free energies of amino acid side-chain rotamers at N3 position	107
4.16. Discussion	111
4.17. Conclusions	114

CHAPTER 5. CXXC MOTIF AT THE N-TERMINUS OF AN ISOLATED $\alpha$ -HELICAL PEPTIDE	115
5.1. Introduction	115
5.2. Design of CAAC peptide	116
5.3. Aggregation and oxidations tests of the CAAC peptide	117
5.4. pH titration of the CAAC peptide	118
5.5. Redox properties of the CAAC peptide	121
5.6. Zinc titration on the CAAC peptide	123
5.7. Lifson-Roig helix-coil theory to calculate $\Delta G$ of interactions in CAAC peptide	124
5.8. The Capping box and CXXC motif	128
5.9. N-cap and helix probability of CAAC peptide	129
5.10. Discussion	131
5.11. Conclusions	135
CHAPTER 6. EFFECT OF PAIRWISE COUPLING IN AN ARG-PHE-MET TRIPLET IN AN ISOLATED $\alpha$ -HELICAL PEPTIDE	136
6.1. Introduction	136
6.2. Design of RFM peptides	137
6.3. Aggregation tests of the R5F5M peptide	138
6.4. Helicity of RFM peptides	139
6.5. Refitting $p$ values for RF and FM interactions	140
6.6. Adjustment of $p$ values for RF and FM interactions in the R4F4M peptide	141
6.7. Helix probability analysis of RFM peptides	143
6.8. Rotamer distribution of Arg, Phe, and Met residues	144
6.9. Discussion	145
6.10. Conclusions	148
CHAPTER 7. EFFECT OF PAIRWISE COUPLING IN A GLU-LYS-GLU TRIPLET IN AN ISOLATED $\alpha$ -HELICAL PEPTIDE	149
7.1. Introduction	149
7.2. Design of EKE peptides	151
7.3. Aggregation tests of EKE peptides	152
7.4. Salt screening experiment	152
7.5. Refitting of $p$ values for EK and KE interactions	154
7.6. pH titration of EKE peptides	158
7.7. Rotamer distribution of Glu and Lys residues	162
7.7.1. $\chi_1$ rotamer preference of Glu and Lys in the EK and KE pairs	162
7.7.2. $\chi_1$ rotamer preference of Glu and Lys in EKE triplet	162
7.8. Discussion	164
7.9. Conclusions	168
CHAPTER 8. GENERAL CONCLUSIONS AND FUTURE WORK	169
REFERENCES	172
APPENDICES	186
Appendix A. HPLC and MS results to check identity of peptides	186
Appendix B. Protein data set from the ASTRAL database	217
Appendix C. Perl script to search $\chi_1\chi_2$ rotamer at N3 position	220
Appendix D. Perl script to search amino acid pairs and their rotamers	224
Appendix E. Publications	

## LIST OF TABLES

	Page
Table 1.1. The most common capping motifs at the $\alpha$ -helix termini	22
Table 1.2. Amino acid propensities at internal positions of the $\alpha$ -helix	26
Table 1.3. Amino acid propensities at N- and C-terminal positions of the $\alpha$ -helix	30
Table 1.4. Energetic of side-chain $i, i+4$ interactions	35
Table 2.1. Assignment of statistical weights to the central residue in a triplet sequence in Lifson-Roig model	49
Table 2.2. Assignment of statistical weights to the fourth residue of a hexamer sequence to include the N1, N2 and N3 preferences	56
Table 3.1. Intrinsic helix parameters of all amino acids for helix-coil programs	66
Table 3.2. Fmoc-protected amino acids used for peptides synthesis	69
Table 3.3. Chemicals used in FastMoc 0.1mM and FastMoc 0.25mM programs	71
Table 3.4. Pre-program Fastmoc cycles on SynthAsist 2.0 run for the synthesis of N3 peptides.	72
Table 3.5. HPLC gradient elution program mainly used for peptide purification at a flow rate of 3.5ml/minute	74
Table 3.6. Mass to charge ratio ( $m/z$ ) of fragmented $[M+H]^+$ , $[M+2H]^+$ and $[M+3H]^+$ ions of the peptides used in the experiments	75
Table 4.1. Helix contents of N3 peptides and energetic preferences for the N3 position	94
Table 4.2. Comparison of $n3.w$ values and energetic preferences for the N3 amino acids calculated using different N1 and N2 values	96
Table 4.3. $\chi_1$ rotamer energies and conformational entropy for transferring a residue from coil to N3	109
Table 4.4. $\chi_1\chi_2$ rotamer energies for transferring a residue from coil to N3	110
Table 5.1. Energetics of the CAAC peptide	127
Table 5.2. Some selected N-cap-N3 pair propensities in the helix N-terminus	128
Table 6.1. Peptides used to study coupling effect in an RFM triplet	137
Table 6.2. Parameter values used for RFM peptides prediction using the Scint2 program	140
Table 6.3. Refitted $p$ values and energetics of interaction in the RFM peptides	141
Table 6.4. $\chi_1$ Rotamer populations of Arg, Phe and Met in $\alpha$ -helices	142
Table 7.1. Peptides used to study coupling effect in an EKE triplet	151
Table 7.2. Refitted $p$ values and energetics of interaction in the EKE peptides.	156
Table 7.3. $\chi_1$ rotamer population of Glu and Lys in $\alpha$ -helices	163
Table 7.4. $\chi_1$ rotamer distribution of the interacting residues in an EKE triplet spaced $i, i+4, i+8$	163
Table 7.5. Energetics of RER triplet	166
Table 7.6. $\chi_1$ rotamer distribution of Glu and Arg in the RE and ER pairs spaced $i, i+4$	166

## LIST OF FIGURES

	Page
Figure 1.1. (A) $\alpha$ -Helix structure (B) Helical wheel	16
Figure 1.2. The Ramachandran plot of protein secondary structures	18
Figure 1.3. Structural examples of (A) The capping box (B) The Schellman motif	21
Figure 2.1. Numbering, coding, and statistical weights assignment in a partially helical peptide for the original Zimm-Bragg and Lifson Roig helix-coil models	47
Figure 2.2. Assignment of statistical weights for residues in the LR model to include N1, N2 and N3 preferences	56
Figure 2.3. Matrix of statistical weights assignment in the N1N2N3 helix-coil model to include N-terminus preferences	57
Figure 3.1. Applied Biosystem 433A peptide synthesizer	67
Figure 3.2. Structure of Rink amide resin, 4-(2'4'-dimethoxyphenyl-Fmoc-aminomethyl)-phenoxyresin	68
Figure 3.3. Peptide chain assembly for Fmoc solid phase peptide synthesis using Rink amide resin	70
Figure 3.4. Circular dichroism spectra of "pure" secondary structures of poly-L-lysine	77
Figure 3.5. Jasco J810 spectropolarimeter	77
Figure 3.6. Illustration of rotamer energies in coil and helix for amino acid X at N3	86
Figure 4.1. Aggregation tests of N3(Glu) and N3(Ile) peptides	91
Figure 4.2. Aggregation tests of N3(Val) peptide	92
Figure 4.3. Plots of $n3.w$ values from different fittings	97
Figure 4.4. pH titrations of N3 peptides (A) N3(Glu), (B) N3(Asp), (C) N3(His), (D) N3(Cys)	101
Figure 4.5. Correlation between N3 propensity ( $P_g$ N3) and N3 statistical weight ( $n3.w$ )	104
Figure 4.6. Correlation between N3 propensity ( $P_g$ N3) and N-cap statistical weight ( $n$ )	104
Figure 4.7. Correlation between N1 statistical weight ( $n1$ ) and N3 statistical weight ( $n3.w$ )	105
Figure 4.8. Correlation between N2 statistical weight ( $n2.w$ ) and N3 statistical weight ( $n3.w$ )	105
Figure 4.9. Correlation between interior statistical weight ( $w$ ) and N3 statistical weight ( $n3.w$ )	106
Figure 5.1. Aggregation tests of CAAC peptide	118
Figure 5.2. Oxidation tests on 20 $\mu$ M CAAC peptide by adding various concentration of H <sub>2</sub> O <sub>2</sub>	119
Figure 5.3. pH titration of CAAC <sub>red</sub> (filled square) and CAAC <sub>ox</sub> (filled triangle)	120
Figure 5.4. Redox equilibrium between 20 $\mu$ M CAAC peptide and DTT of various concentrations	122
Figure 5.5. Zn titration of CAAC peptide at 20 $\mu$ M	123
Figure 5.6. Statistical weight assignments for residues in CAAC peptide	124
Figure 5.7. Fitting of $r$ values of CAAC peptide of different states to agree with experiment	126

Figure 5.8.	(A) N-cap probability (B) Predicted helicity as a function of positions of residues in CAAC peptide	130
Figure 5.9.	Structural examples of the CXXC motif in $\alpha$ -helices	134
Figure 6.1.	Aggregation test of the R5F5M peptide	138
Figure 6.2.	Adjustment of $p$ values for the RF and FM pairs when present simultaneously in the R4F4M peptide to get predicted helicity to agree with experiment	142
Figure 6.3.	Predicted helicity as a function of positions in RFM peptides using the Scint2 program	143
Figure 6.4.	Structural example of RFM triplet in the $\alpha$ -helix in pdb-1lw3	147
Figure 7.1.	(A) Aggregation test of EKE peptides. (B) Aggregation test of E4K4E peptide after prolonged incubation period	153
Figure 7.2.	Salt screening of EKE peptides	154
Figure 7.3.	Adjustment of $p$ values for the EK and KE pairs when present simultaneously in the E4K4E peptide to get the predicted helicity to agree with experiment	157
Figure 7.4.	pH titration of EKE peptides (A) E5K5E, (B) E4K5E, (C) E5K4E and (D) E4K4E	160
Figure 7.5.	Structural examples of (A) EK interaction between Glu62 and Lys66 in 1A48 (B) KE interaction between Lys175 and Glu179 in 1AOP (C) EKE interaction between Glu110 and Lys114-Glu118 in 1L2P	165
Figure 7.6.	Structural examples of R4E4R triplet in $\alpha$ -helix	167

## LIST OF ABBREVIATIONS

Ac	:	Acetyl
Ala	:	Alanine
Arg	:	Arginine
Asn	:	Asparagine
Asp	:	Aspartic acid
Boc	:	N- $\epsilon$ -t-butyloxycarbonyl
CD	:	Circular Dicroism
Cys	:	Cysteine
DCM	:	Dichloromethane
DIEA	:	N,N-Diisopropylethylamine
DMF	:	N,N-Dimethylformamide
EKE	:	Glu-Lys-Glu
ESI	:	Electrospray ionization
Fmoc	:	9-Fluorenylmethoxycarbonyl
g <sup>+</sup>	:	Gauche <sup>+</sup>
g <sup>-</sup>	:	Gauche <sup>-</sup>
Gln	:	Glutamine
Glu	:	Glutamic acid
Gly	:	Glycine
GuHCl	:	Guanidine hydrochloride
HBTU	:	2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
His	:	Histidine
HOBt	:	1-Hydroxy-1H-benzotriazole
HPLC	:	High Performance Liquid Chromatography
Ile	:	Isoleucine
Leu	:	Leucine
LR	:	Lifson-Roig
Lys	:	Lysine
MALDI-TOF	:	Matrix Assisted Laser Desorption Ionization Time-of-Flight
Met	:	Methionine
MS	:	Mass spectrometry
NMP	:	1-Methyl-2-pyrrolidinone
OtBu	:	$\alpha$ -t-butyl ester
Pbf	:	N <sup>G</sup> -2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl
Phe	:	Phenylalanine
Pro	:	Proline
RFM	:	Arg-Phe-Met
Ser	:	Serine
t	:	Trans
tBu	:	O-t butyl
TFA	:	Trifluoroacetic acid
Thr	:	Threonine
TIS	:	Triisopropylsilane
Trp	:	Tryptophan
Trt	:	N- $\beta$ -trityl
Tyr	:	Tyrosine
Val	:	Valine
ZB	:	Zimm-Bragg

## ABSTRACT

The aim of this thesis is to study some factors that affect  $\alpha$ -helix stability - as the starting point to understand protein structure, stability and folding - using a combined circular dichroism and PDB-derived structural analysis approach. Measurements of CD helicities of synthetic helical peptides were converted to free energies using a modified Lifson-Roig helix-coil theory. The results, in most cases, were rationalised from the protein crystal structures, in terms of frequency and side-chain rotamer preferences. Firstly, the effect of substituting all 20 amino acids in the N3 position in an  $\alpha$ -helical peptide was studied. The rank order of stabilising residues at N3 is Ala, Glu, Met/Ile, Leu, Lys, Ser, Gln, Thr, Tyr, Phe, Asp and Trp. The free energies for Cys, Gly, Asn, His, Arg and Pro could not be fitted, however. In general, the N3 position favours negatively charged residues, due to electrostatic interactions with the helix dipole. The results correlate with N1, N2 and helix interior preferences, but not with N-cap preferences. A complementary rotamer energy analysis gives a thorough understanding of the structural uniqueness of the N3 position. Then, thioredoxin and metal binding CXXC motifs in an  $\alpha$ -helical peptide was studied. This motif has a very high local propensity ( $26.3 \pm 6.2$ ) at the N-termini of  $\alpha$ -helices. The motif binds to zinc with a  $K_d$  of  $13.9 \mu\text{M}$ . The motif has a redox potential of  $E_0 = -235\text{mV}$ , which is very similar to native human thioredoxin. The motif is stabilising when bound to Zn with  $\Delta G = -1.32\text{kcal/mol}$ . Disulfide bond formation between the two Cys residues also stabilises the peptide by  $-0.47\text{kcal/mol}$ . Finally, a peptide containing a triplet of RFM was used to study cooperativity in pairwise coupling. The hydrophobic RF and FM interactions in an  $i, i+4$  spacing stabilise the  $\alpha$ -helix by  $-0.29\text{kcal/mol}$  and  $-0.59\text{kcal/mol}$ , respectively. When both interactions are present simultaneously, however, they stabilise the helix by an additional  $-0.75\text{kcal/mol}$ , energy nearly as much as the sum of its parts. This cooperativity effect is attributed to a shared rotamer preference, as the central Phe adopts the  $t$  rotamer in both pairs. In contrast, a peptide containing a EKE triplet shows an anti-cooperativity effect when the EK and KE pairs were coupled. The fully charged salt-bridges stabilise the  $\alpha$ -helix by  $-0.60\text{kcal/mol}$  and  $-1.02\text{kcal/mol}$ , respectively, when in an  $i, i+4$  spacing. When both interactions are present simultaneously, the second salt-bridge provides no additional stabilisation to the helix. This is attributed to the inability of the central Lys to form two salt-bridges simultaneously. Pairwise coupling is thus demonstrated to have large effects on protein stability. The results from this thesis may be used to rationally modify protein stability and help in *de novo* protein design. The free energies may be used to improve secondary structure prediction algorithms.

## DECLARATION

---

That no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning

## COPYRIGHT STATEMENT

---

- (i) Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made **only** in accordance with instructions given by the Author and lodged in the John Rylands University Library of Manchester. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.
- (ii) The ownership of any intellectual property right which may be described in this thesis is vested in The University of Manchester, subject to any prior agreement to the contrary, and may not be made available for use by third parties without written permission of the University, which will prescribe the terms and conditions of any such agreement.
- (iii) Further information of the conditions under which disclosures and exploitation may take place is available from the Dean of Faculty of Life Sciences.



## ACKNOWLEDGMENTS

---

I would like to express my best gratitude to the following people who helped me throughout my PhD program:

- My supervisor, Professor Andrew J. Doig, for his patience and invaluable guidance in the last three and a half years.
- Dr. Neil Errington, for his experience, assistance and discussion, especially for his help in writing the program for protein database search. I particularly have to mention that the Perl scripts for the search of CXXC motif was written solely by him.
- Dr. Jody Mason and Dr. Nicoleta Kokkoni for their technical advices that helped me to pass the most difficult part of the experiments in the early period of my PhD. Other Doig's group members are also thanked.
- The Michael Barber Mass Spectrometry at UMIST for verifying the peptides identity.
- The Government of Indonesia through the *Technological and Professional Skills Development Project* which made it possible for me to do my PhD at the University of Manchester.

*For my mother, Eka and Yasmin, who have always been with me in  
joy and sorrow*

## CHAPTER 1

### INTRODUCTION TO STRUCTURE AND STABILITY OF $\alpha$ -HELIX

#### 1.1. Introduction

Amino acid sequences determine the structure of proteins, at least for small proteins *in vitro*. This was first reported by Anfinsen (1973) in the folding and refolding experiment of ribonuclease. Molecular chaperones, which assist protein folding in cells, exist primarily to prevent aggregation of non-native structures, but do not contribute to conformational information in the folding process (Hartl and Hayer-Hartl, 2002). The mechanism of how a polypeptide sequence encoded from a gene, can fold into a distinct 3D shape and a fully functional, native protein remains elusive, however. Random sampling of the primary sequence to find the lowest energy conformation will take a tremendous amount of time (Levinthal, 1968). In reality, a polypeptide sequence folds in a time scale of milliseconds. The folding process needs to consider many features of the protein conformation, including the nature of every amino acid comprising the primary sequence of the protein, the possible interactions between amino acids within a secondary structure, the potential contacts among secondary structures and interactions with the environment, which are not yet fully understood.

There are several mechanisms proposed to explain the protein folding events. One of the earliest theories is the nucleation growth mechanism that suggests that tertiary structure propagates rapidly from an initial nucleus of local secondary structures (Wetlaufer, 1973). This mechanism is not fully accepted as it fails to explain folding intermediates. The alternative model is the hydrophobic collapse model in which hydrophobic collapse brings all possible contacts in the protein close together, so that folding can take place in a restricted volume to give a limited conformational search to the native structure (Gutin *et al.*, 1995; Zhou *et al.*, 2004). In this model, the earliest steps may or may not be accompanied the formation of secondary structure elements (Matthews, 1993). Other model is the so called framework model (Ptitsyn and Rashin, 1975; Kim and Baldwin, 1982), in which secondary structure is proposed to fold first, followed by docking of the pre-formed

secondary structural units to yield the native, folded protein. The nucleation-condensation mechanism is a model which unites features the hydrophobic collapse and framework models. It was proposed due to growing evidences that, in some proteins, secondary and tertiary structures are formed in parallel (reviewed by Daggett and Fersht (2003)).

In any proposed mechanism above, apart from hydrophobic collapse model, it is remarkable that folding always involves secondary structure formation as an early event. A widely accepted theory about protein folding is that secondary structure formation initiates the arrangement of native 3D structure of proteins. Molecules of the same protein can follow different pathways converging towards the same end, although the choices of pathways are limited (reviewed by Dobson (2003)). Experimental data and theoretical considerations support an argument that protein folding process is hierarchic (reviewed by Baldwin and Rose (1999a; 1999b)), starting by the formation of local structure with minimum energy that is determined largely by local sequence information. It has been long suggested that the folding nucleation centres, involving secondary structures like  $\alpha$ -helix,  $\beta$ -sheet and  $\beta$ -turn, can move in and out of the conformation that they occupy in the final protein. They form a relatively rigid structure, stabilised by a set of cooperative interactions (Anfinsen, 1973). This way, a totally random sampling of all conformations will be restricted due to steric hindrance and folding is thus a more directed process. Thus, understanding the secondary structure folding problem seems to be a reasonable pathway to begin to understand the protein-folding problem.

Secondary structures of proteins arise from specific main chain hydrogen bond pairings, adopting specific backbone torsion angles and stabilised by many types of interactions. One of the most abundant secondary structures, in which approximately 30% total amino acids in proteins are found, is the  $\alpha$ -helix (Barlow and Thornton, 1988). Other abundant secondary structures are  $\beta$ -sheets and turns, although they will not be discussed here. Helices are very often located parallel to the surface of globular proteins (Doig *et al.*, 1997), making them prime candidates for serving the function of a barrier between inside and out of the whole structure. Thus, the unique

packing of amino acid residues in the forms of helices is essential to attain the overall protein stability.

Although  $\alpha$ -helix structure has been extensively studied theoretically, statistically and experimentally, there are still areas that are not fully understood. For example, only some information on the energetics of all the 400 possible  $i, i+4$  interactions, resulting from 20x20 combinations of amino acids, within the helix are available. Studies on amino acid networks within helices in an isolated environment are also still in their early stages. Moreover, interactions between the secondary structures within tertiary structure are still unclear. Therefore, this thesis tries to address several forces that affect  $\alpha$ -helix stability, limited to the positional dependence of amino acids and side-chain interactions. The results may be important, not only to understand the folding mechanism at least at the secondary structure level, but also to improve algorithms for secondary structure prediction as well as in the area of protein modelling and *de novo* protein design.

Knowledge of the origins of  $\alpha$ -helix stability and amino acid preferences are important in understanding  $\alpha$ -helix formation. The following sections give a general perspective on  $\alpha$ -helix structure and stability. Many previous studies on the  $\alpha$ -helix, starting with an overview of the  $\alpha$ -helix structure, are presented. Factors that stabilise the  $\alpha$ -helix are also considered. At the end of this chapter, the aims of the study and organisation of the thesis are stated.

## 1.2. Structure of $\alpha$ -helix

Pauling *et al.* (1951) elegantly predicted an ideal  $\alpha$ -helix structure. An  $\alpha$ -helix is formed through a repeating pattern of hydrogen bonds between the backbone carbonyl CO of residue  $i$  to the backbone amide NH of residue  $i+4$  (Figure 1.1A). The distance between adjacent amino acids is 1.5 Å. Each turn rises through 5.4 Å, giving 3.6 amino acids per complete turn of the helix. The stable conformation of an  $\alpha$ -helix is raised by a succession of the sterically allowed Phi ( $\phi$ , N-C $\alpha$  bond) and Psi ( $\psi$ , C-C $\alpha$  bond) dihedral angles that naturally position the backbone NH and CO groups towards each other for hydrogen bond formation. All

amide protons point toward the N-terminus and all carbonyl oxygens point toward the C-terminus, giving a distinct dipole moment at each terminus. The repeating nature of the  $\phi$ ,  $\psi$  angles ensures this orientation. The interior of an  $\alpha$ -helix is filled with atoms, so all the side-chains project out from the helical axis. Adjacent side-chains are on nearly opposite sides of the helix, while every third or fourth side-chain is on the same side of the helix (Figure 1.1B). Side-chains of residues in positions  $i, i+3$  and  $i, i+4$  can thus interact with each other. This gives a distinct effect on  $\alpha$ -helix stability depending on the type of interactions.

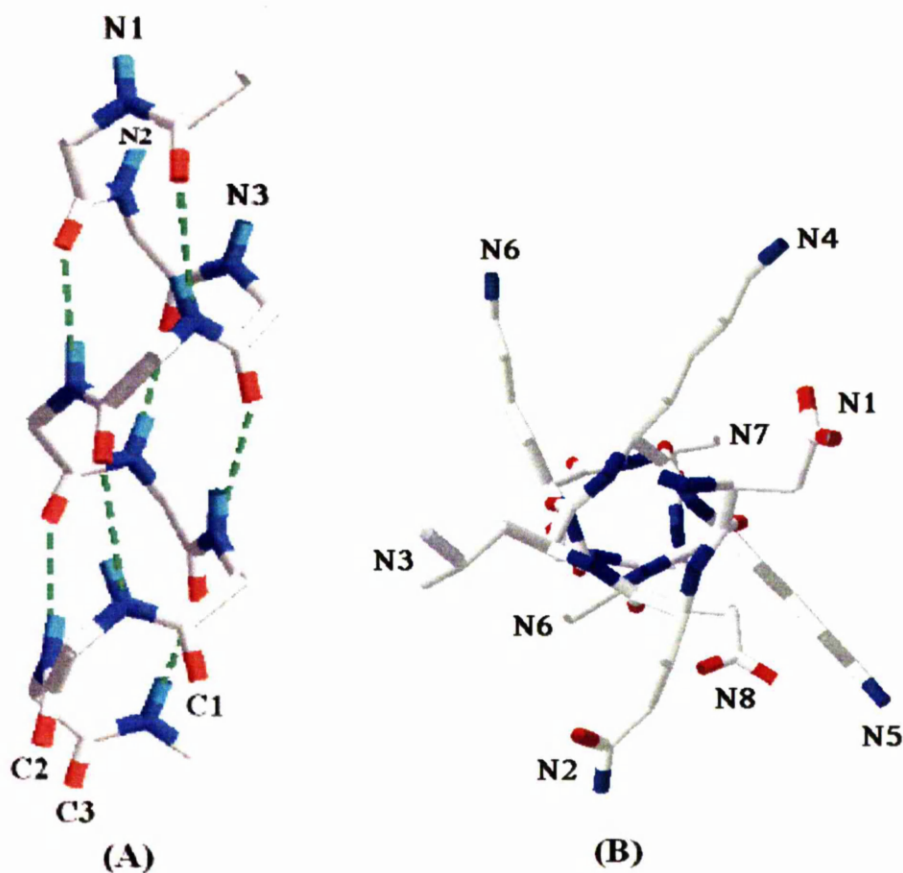


Figure 1.1. (A)  $\alpha$ -Helix structure (B) Helical wheel.

### 1.2.1. $\alpha$ -Helix nomenclature

The residues at the N-terminus of the  $\alpha$ -helix are called N''-N'-N-cap-N1-N2-N3 etc., where the N-cap is the residue with non-helical  $\phi$ ,  $\psi$  angles immediately

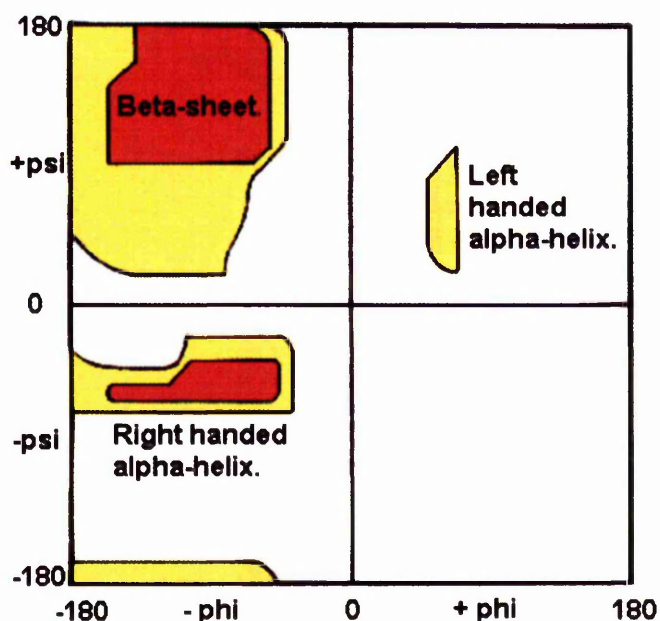
preceding the N-terminus of an  $\alpha$ -helix. Compatible N-cap residues can hydrogen-bond to residues in the following helical segment. N1, N2 and N3 are the first three amino acids with helical  $\phi$ ,  $\psi$  backbone dihedral angles at the helix N-terminus (Richardson and Richardson, 1988). They differ from interior positions as their amide NH groups do not participate in backbone-backbone  $i, i+4$  hydrogen bonds within the helix (Figure 1.1). The presence of these otherwise unsatisfied hydrogen bond donors has profound structural effects, most often satisfied by hydrogen bonds to side-chains local in sequence, such as to preceding N-cap side-chains. The C-terminal residues are similarly called C3-C2-C1-C-cap-C'-C'' etc. The C-cap is the residue with non-helical  $\phi$ ,  $\psi$  angles immediately following the C-terminus of an  $\alpha$ -helix and can also form intrahelical hydrogen bond with the preceding helical residues, provided compatible residues are selected. The C1, C2 and C3 residues are also unique because their amide groups participate in  $i, i+4$  backbone-backbone hydrogen bonds using only their NH groups. The need for these groups to form hydrogen bonds has powerful effects on helix structure and stability (Presta and Rose, 1988).

### 1.2.2. Helix dihedral angles

The  $\phi$  and  $\psi$  dihedral angles around the planes of the peptide bond theoretically can be rotated  $180^\circ$  in either direction of the  $\alpha$ -carbon. However, distances between atoms, charges, and bond angles prohibit this from occurring. The Ramachandran Plot (Figure 1.2) illustrates the relatively small fraction of all theoretical  $\phi$  and  $\psi$  rotations that are stable for the  $\alpha$ -helix (Ramachandran *et al.*, 1963). The mean  $\phi$  and  $\psi$  angles are  $-62$  degrees and  $-41$  degrees, respectively. However the  $\phi$  and  $\psi$  angles taken from helices identified in 3D structures are often slightly different from the ideal values.

The plot shows that both right- and left-handed polypeptide helices can be stabilising. Interestingly, the exact three-dimensional structure of a native protein is made from amino acids with the L-configuration, forming the right handed  $\alpha$ -helix. There is no obvious reason why nature chooses for all L instead of all D as the

proteins building block. There is no distinct area for  $\alpha$ -,  $3_{10}$  and  $\pi$ -helices and they are all positioned with  $-\phi$  and  $-\psi$  angles in the plot.



**Figure 1.2.** The Ramachandran plot of protein secondary structures (<http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html>).

### 1.2.3. $\alpha$ -Helix distortion

About 85% of  $\alpha$ -helices are in some way distorted. The distortion can be attributed to different factors such as solvent-side-chain interactions, local sequence and side-chain packing. A common distortion to naturally occurring  $\alpha$ -helices structure is a kink in the helical axis ( $\approx 25$  degrees from linear) caused by the presence of a Pro in the interior of long helices, normally  $>4$  turns (Barlow and Thornton, 1988). When non-Gly residues precede Pro, the angle of  $\psi$  is restricted to positive values ( $+30 < \psi < +180$ ) because Proline has its  $\phi$  and  $\psi$  restricted. However, there is no such effect if the preceding residue is a Gly. Therefore, sometimes Pro is referred to as a "helix breaker" due to this phenomenon. When this occurs, nevertheless, only one hydrogen bond (amide of the following residue,  $i+1$ , to the carbonyl oxygen of  $i-3$ ) is found to be broken due to the kink in the helix. Pro is often found in the first turn of helices since it has its own capping residue (imide)



and its "helix-like" backbone dihedral angles are suggested to have role to start and terminate the helix (Seale *et al.*, 1994; Viguera and Serrano, 1999).

#### 1.2.4. Helix dipole

In the  $\alpha$ -helix, all of the peptide units point in the same direction (roughly parallel to the helical axis) and therefore should show a cumulative effect, resulting in a helix macro dipole. The net effect of this helix macro dipole is the presence of partially positively charged amide nitrogen atoms of approximately +0.5 unit charge at the N-terminus and partially negatively charged oxygen atoms of approximately -0.5 unit charge at the C-terminus (Hol *et al.*, 1981). It destabilises the helix, but the presence of negatively charged and positively charged residues at the N- and the C-terminus, respectively, has been shown to be stabilising. For example, Huyghues-Despointes *et al.* (1993b) used a single Asp residue placed at various positions at the helix N-terminus. The interaction of Asp<sup>-</sup> with the helix dipole raises helix stability. Nevertheless the stabilising effect of these charged residues is associated with localised dipoles rather than the helix macro dipole (Aqvist *et al.*, 1991).

#### 1.3. Structure of $3_{10}$ - and $\pi$ -helix

Other helix structures are the  $3_{10}$ -helix and the  $\pi$ -helix. Only a brief introduction about these structures is given here. The  $3_{10}$ -helix is the fourth most common type of secondary structure in proteins after  $\alpha$ -helices,  $\beta$ -sheets and reverse turns (Barlow and Thornton, 1988; Karpen *et al.*, 1992). It is stabilised by  $i, i+3$  hydrogen bonds, instead of  $i, i+4$  found in  $\alpha$ -helices. Each turn rises through 6 Å, with 3 amino acids per complete turn. Most  $3_{10}$ -helices are short with only 3 or 4 residues (Barlow and Thornton, 1988), making it less stable than the  $\alpha$ -helix due to non-linear backbone hydrogen bonds.  $3_{10}$ -helices are commonly found as N- or C-terminal extensions to an  $\alpha$ -helix (Némethy *et al.*, 1967; Baker and Hubbard, 1984; Barlow and Thornton, 1988). Therefore, the  $3_{10}$ -helix is considered as a possible intermediate in  $\alpha$ -helix formation (Millhauser, 1995; Bolin and Millhauser, 1999). The mean  $\phi$  and  $\psi$  for the  $3_{10}$ -helix are  $-49^\circ$  and  $-26^\circ$ , respectively, and are found in a

similar region to the  $\alpha$ -helix in the Ramachandran plot. This may allow structural transformation from one helix to the other (Toniolo and Benedetti, 1991).

Low and Baybutt (1952) proposed the structure of the  $\pi$ -helix which is also formed through hydrogen bonds occurring between residues  $i$  and  $i+5$  giving a helical repeat unit of 4.4 amino acids per turn. The  $\pi$ -helix is extremely rare in protein crystal structures, suggesting that it is inherently unstable. This could be because its dihedral angles are energetically unfavourable relative to the  $\alpha$ -helix (Low and Grenville-Wells, 1953; Ramachandran and Sasisekharan, 1968). A 1 Å hole down the centre is too narrow to allow a water molecule to penetrate, resulting in the loss of van der Waals interactions. The  $\pi$ -helix initiation is also entropically unfavourable because the 4.4 residues per turn require correct orientation before the first  $i,i+5$  hydrogen bond is formed (Shirley and Brooks, 1997; Lee *et al.*, 2000; Weaver, 2000; Morgan *et al.*, 2001; Fodje and Al-Karadaghi, 2002). However, a synthetic peptide stabilised by cetyltrimethylammonium bromide (CTAB) micelles forms a  $\pi$ -helix when bound to  $\text{Zn}^{2+}$ . The peptide was found to form moderately stable monolayers at the air-water interface (Morgan *et al.*, 2001).

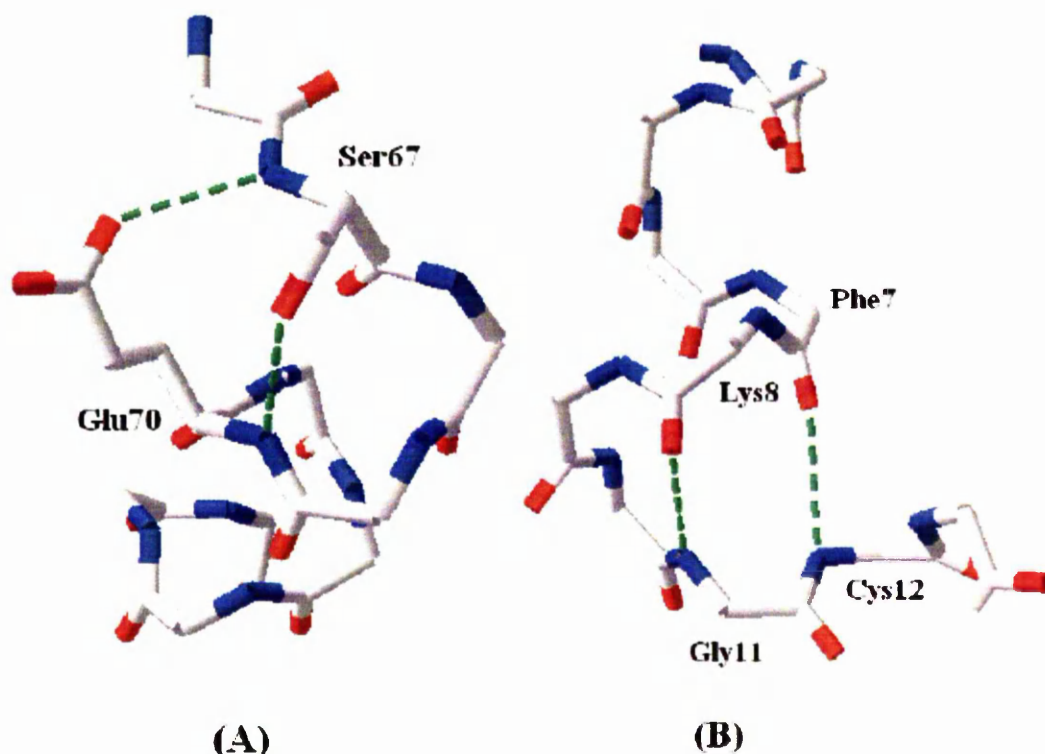
#### 1.4. Capping motifs

The amide NH groups at the helix N-terminus are satisfied predominantly by side-chain hydrogen bond acceptors. In contrast, carbonyl CO groups at the C-terminus are satisfied primarily by backbone NH groups from the sequence following the helix (Presta and Rose, 1988). The presence of such interactions would therefore stabilise helices. A well-known interaction is the helix capping, defined as specific patterns found at or near the ends of helices (Serrano and Fersht, 1989; Bell *et al.*, 1992; Chakrabartty *et al.*, 1993a; Dasgupta and Bell, 1993; Forood *et al.*, 1993). The patterns mostly involve hydrogen bonding and hydrophobic interactions as summarised in Table 1.1.

A well-known pattern of capping at the helix N-terminus is the capping box (Figure 1.3A). Here, the side-chain of the N-cap forms a hydrogen bond with the backbone of N3 and, reciprocally, the side-chain of N3 forms a hydrogen bond with the backbone of the N-cap (Harper and Rose, 1993). Ser or Thr at N-cap and Glu at

N3 are the most frequently observed interacting residues (Wan and Milner-White, 1999). The definition of the capping box was expanded by Seale *et al.* (1994) to include an associated hydrophobic interaction between residues N' and N4, known as a 'hydrophobic staple' (Muñoz and Serrano, 1995c). A variant of the capping box motif, termed the 'big box', contains a hydrophobic interaction between nonpolar side-chain groups in residues N4 and N'' (not N') (Seale *et al.*, 1994).

A capping motif involving three hydrophobic residues and a Pro residue at N-eap was proposed by the Serrano group (Viguera and Serrano, 1999), known as the Pro-box motif. The N-cap Pro is usually associated with Ile or Leu at position N'', Val at position N3 and a hydrophobic residue at position N4. As Pro has a low N-cap preference (Doig and Baldwin, 1995) and Val has a poor helix propagation parameter value (Rohl *et al.*, 1996), this motif possibly discourages helix elongation. The Pro-box motif may contribute more to the specificity of the protein fold rather than to the overall stability.



**Figure 1.3.** Structural examples of (A) The capping box (B) The Schellman motif. Both motifs are observed in *cytochrome 351*. Hydrogen bonds are shown as dashed green line.

The N-terminal capping box sequence stabilises helices by inhibiting N-terminal fraying. It seems, however, that it does not necessarily promote elongation unless accompanied by favourable hydrophobic interactions as in a “hydrophobic staple” motif (Jiménez *et al.*, 1994; Kallenbach and Gong, 1999). The nature of the capping box stabilising effect thus not only arises from reciprocal hydrogen bonds between compatible residues, but also from local interactions between side-chains, interactions between helix macrodipole and charged residue and solvation (Petukhov *et al.*, 1996).

**Table 1.1.** The most common capping motifs at  $\alpha$ -helix termini.

Sequence	Related position	Designation	Interactions	References
N-capping				
p-XXp	N-cap $\rightarrow$ N3	Capping box	Reciprocal H-bonds between residues at N-cap and N3, where S,T, N and D at N-cap and E at N3 are the most frequently observed residues.	(Dasgupta and Bell, 1993; Harper and Rose, 1993; Doig <i>et al.</i> 1997)
hp-XXph	N' $\rightarrow$ N4	Hydrophobic staple	Hydrophobic interactions between residues at N' and N4 accompanying capping box interaction.	(Seale <i>et al.</i> , 1994; Muñoz and Serrano, 1995c)
hP-XXVh	N' $\rightarrow$ N4	Pro-box motif	Pro at N-cap, associated with Ile and Leu at N', Val at N3 and a hydrophobic residue at N4.	(Viguera and Serrano, 1999)
C-capping				
hXp-XGh	C3 $\rightarrow$ C''	Schellman motif	H-bonds between C'' NH and C3 CO, accompanied by H-bonds between C' NH and C2 CO, further accompanied by hydrophobic interaction between C3 and C''	(Schellman, 1980; Viguera and Serrano, 1995a)
hXp-XGp	C2 $\rightarrow$ C''	$\alpha_L$ motifs	As Schellman motif but H-bond between C' NH and C3 CO where C'' polar.	(Aurora <i>et al.</i> , 1994)
X-Pro	C-cap $\rightarrow$ C'	Pro-capping motif	Aliphatic or aromatic residue at C-cap and Pro at C'.	(Prieto and Serrano, 1997)

p = polar amino acids, h = hydrophobic amino acids, X = any amino acids, P = Proline and G = Glycine

The two primary capping motifs found at helix C-termini are the Schellman and the  $\alpha_L$  motifs (Schellman, 1980; Aurora *et al.*, 1994; Aurora and Rose, 1998). The Schellman motif is defined by a doubly hydrogen-bonded pattern between backbone partners, consisting of hydrogen bonds between the amide NH at C'' and the carbonyl CO at C3 and between the amide NH at C' and the carbonyl CO at C2, respectively

(Figure 1.3B). The associated hydrophobic interaction is between C3 and C''. In the Schellman motif, polar residues are highly favoured at the C1 position and the C' residue is typically Glycine.

If C'' is polar, the alternative  $\alpha_L$  motif is observed. The  $\alpha_L$  motif is defined by a hydrogen bond between the amide NH at C' and the carbonyl CO at C3. As in the Schellman motif, the C' residue is typically Glycine, which adopts a positive value of  $\phi$  dihedral angle. However, the hydrophobic interaction in an  $\alpha_L$  is heterogeneous, occurring between C3 and any of several residues external to the helix (C<sup>3'</sup>, C<sup>4'</sup>, or C<sup>5'</sup>) (Aurora *et al.*, 1994).

Despite statistical analyses revealing that Schellman motifs are observed more frequently than expected at the helix C-terminus, this motif populates only transiently in aqueous solution but it is formed in 30% TFE (Viguera and Serrano, 1995a). This might be due to the C-terminus being frayed and the increase of helical content contributed from this motif is small. Energetically this motif is not very favourable due to the entropic cost of fixing a Gly residue at the position C'. The Schellman motif is thus a consequence of helix formation and does not involve  $\alpha$ -helix nucleation (Gong *et al.*, 1995). The  $\alpha_L$  motif seems to be more stable than the alternative Schellman motif (Kallenbach and Gong, 1999).

A statistical analysis of protein crystals detected another possible C-cap local motif, designated Pro C-capping motif (Prieto and Serrano, 1997). Certain combinations of X-Pro pairs in which residue X is the C-cap and the Pro is at position C' are more abundant than expected. The aliphatic (Ile, Val and Leu) and aromatic residues (Phe, Tyr, Trp), together with Asn, His and Cys are the most favoured residues accompanying Pro at position C'.

A notable difference between the N- and C-terminal motifs is that at the N-terminus, helix geometry favours side-chain to backbone hydrogen bonding and selects for compatible polar residues (Doig *et al.*, 1997; Penel *et al.*, 1999). Accordingly, the N-terminus promotes selectivity in all polar positions, especially for N-cap and N3 in the capping box. In contrast, side-chain to backbone hydrogen bonding is disfavoured at the C-terminus. Backbone hydrogen bonds are satisfied

instead by post-helical backbone groups. The C-terminus needs only to select for C' residues that can adopt positive values of the backbone dihedral angle  $\phi$ , most notably Gly (Aurora *et al.*, 1994).

### 1.5. Factors that stabilise $\alpha$ -helix

In addition to the backbone hydrogen bonds, helix formation involves amino acid propensities, side-chain interactions and many other forces that contribute to the conformational stability. Two experimental approaches are principally used to study the factors that influence  $\alpha$ -helices, namely site-directed mutagenesis on intact proteins and site substitutions on peptide models. The latter is superior to the former because it provides a relatively context-free environment for substitutions. However, it involves a complex equilibrium and needs to use helix-coil theory for energetics analysis. A comparison between helix propensities in peptides and proteins shows a good correlation (Myers *et al.*, 1997). Therefore the following sections give more attention to the studies using the peptide models approach.

#### 1.5.1. Amino acids preference at helix interior

Statistical studies on protein crystal structures show that each amino acid at each position in the  $\alpha$ -helix has its own preference (Barlow and Thornton, 1988; Kumar and Bansal, 1998). For example, Ala and Leu are abundant, whereas Pro and Gly are rare. One of the early empirical studies to determine the helical preference of individual amino acids was done by Scheraga and co-workers using a host-guest strategy (Wojcik *et al.*, 1990). The host-guest strategy employs long random copolymers of water soluble, non-ionic guest (poly[N<sup>5</sup>-(3-hydroxypropyl)-L-glutamine] (PHPG) or poly[N<sup>5</sup>-(4-hydroxybutyl)-L-glutamine] (PHBG)), together with a low (10-50%) content of the guest residue of all 20 naturally occurring amino acids. The propagation ( $s$ ) and initiation ( $\sigma$ ) parameter values of the Zimm-Bragg helix-coil model (see Chapter 2) of the host homopolymer are used to compute those for the guest. The results from the host-guest work are in disagreement with most of those from short peptides of fixed sequence (Table 1.2). This is because the host side-chains can interact with each other. The introduction of a guest residue thus removes

host-host interactions and replaces them with PHBG-guest or PHPG-guest side-chain interactions that may obscure the intrinsic helix propensities.

Another approach to determine the helical preference of individual amino acids is by employing short alanine-based peptides of fixed sequence. Kallenbach and co-workers used synthetic peptides with a sequence of succinyl-YSEEEKAKKAXAEAEKKKK-NH<sub>2</sub>. Substitutions at X allow determination of helix stabilising energies for common amino acids (Lyu *et al.*, 1990). Stellwagen and co-workers made substitutions in position 9 of Ac-Y(EAAAK)<sub>3</sub>A-NH<sub>2</sub> (Park *et al.*, 1993). Chakrabarty *et al.* (1994) made substitutions at internal positions of various homologous poly-Ala peptides with a general sequences of Ac-(AAKAA)<sub>m</sub>Y-NH<sub>2</sub> (or with Q instead of K) and yielded a complete energetic preference for amino acids in the interior. Rohl *et al.* (1996) adjusted the interior helix propensities of Chakrabarty by measuring them both in water and 40% (v/v) trifluoroethanol (TFE) to allow a complete calculation of the Lifson-Roig  $w$  parameter and stabilisation energy for all 20 amino acids.

Helical propensities and stabilisation of amino acids in the interior have also been studied using site directed mutagenesis methods. Blaber *et al.* (1993) substituted site 44 of T4 lysozyme, yielding an energetic scale relative to Gly for the substituting amino acids. The scale correlates well with model peptide studies and with studies on the frequency of amino acid occurrence in protein structures (Chou and Fasman, 1978). Fersht and co-workers used a similar method by replacing Ala32 of the second helix in Barnase with the other 19 naturally common amino acids (Horovitz *et al.*, 1992). This site, however, is at a C2 position and may not be comparable with the internal sites.

O'Neil and DeGrado (1990) used substitutions into an  $\alpha$ -helical two-stranded coiled-coil system to deduce helix-forming tendencies of common amino acids. The designed peptide, containing a single solvent exposed site, forms a non-covalent  $\alpha$ -helical dimer, which is in equilibrium with a randomly coiled monomeric state. Substitution of amino acids into the guest site gave equilibrium constants for the monomer-dimer equilibrium, which were then used to determine the free energy. The

values show good agreement with those of other values obtained from different model systems.

**Table 1.2.** Amino acid propensities at internal positions of  $\alpha$ -helix.

Amino acid	$s$ (20°C) (Wojcik <i>et al.</i> , 1990)	$\Delta G^{\circ}_{H_2O}$ (kcal/mol) (Rohl <i>et al.</i> , 1996)	$\Delta G^{\circ}_{TFE}$ (kcal/mol) (Rohl <i>et al.</i> , 1996)	$\Delta\Delta G_{X-Gly}$ (kcal/mol) (O'Neil and DeGrado, 1990)	$\Delta\Delta G_{X-Ala}$ (kcal/mol) (Petukhov <i>et al.</i> , 1998; Petukhov <i>et al.</i> , 1999)	$\Delta\Delta G_{X-Gly}$ of T4L site 44 (kcal/mol) (Blaber <i>et al.</i> , 1993)	$\Delta\Delta G_{X-Gly}$ of Barnase site 32 (kcal/mol) (Horovitz <i>et al.</i> , 1992)	$\Delta\Delta G_{X-Ala}$ (kcal/mol) (Pace and Scholtz, 1998)
A	1.07	-0.27	-0.52	-0.77		-0.96	-0.91	0
C <sup>-</sup>								na
C <sup>o</sup>	0.99*	0.64	0.49	-0.23		-0.43*	-0.09*	0.68
D <sup>-</sup>	0.78*	0.54	0.52	-0.15*		-0.43	-0.2	0.69
D <sup>o</sup>		0.52						0.43
E <sup>-</sup>	1.35*	0.35	-0.27	-0.27*		-0.53*	-0.36*	0.4
E <sup>o</sup>		0.21						na
F	1.09	0.73	0.02	-0.41		-0.59	-0.22	0.54
G	0.59	1.7	1.3	0	1.26	0	0	1
H <sup>+</sup>	0.85*	0.84	0.33	-0.06*		-0.57*	-0.13*	0.66
H <sup>o</sup>		0.57	0.19					0.56
I	1.14	0.44	-0.17	-0.23	0.54	-0.84	-0.1	0.41
K <sup>+</sup>	0.94	0.019	-0.06	-0.66		-0.73	-0.72	0.26
L	1.14	0.095	-0.29	-0.62	0.21	-0.92	-0.56	0.21
M	1.2	0.25	-0.15	-0.5	0.16	-0.86	-0.6	0.24
N	0.78	0.69	0.03	-0.07	0.6	-0.39	-0.25	0.65
P	0.19	>3.8	3.5	~3		2.5	na	3.16
Q	0.98	0.28	-0.3	-0.33	0.42	-0.8	-0.43	0.39
R <sup>+</sup>	1.03	-0.052	-0.08	-0.68		-0.77	-0.77	0.21
S	0.76	0.52	0.6	-0.35	0.55	-0.53	-0.5	0.5
T	0.82	0.95	0.27	-0.11	0.83	-0.54	-0.12	0.66
V	0.95	0.77	-0.05	-0.14	0.56	-0.63	-0.03	0.61
W	1.11	0.69	na	-0.45		-0.58	-0.07	0.49
Y	1.02	0.42	na	-0.17		-0.72	-0.09	0.53

\*Glu and Asp are assumed as negatively charged while Cys is neutral and His, Arg and Lys are positively charged.

Pace and Scholtz (1998) compared the propensity of each amino acid in the helix interior derived from previous studies, including both in proteins and peptides.



The  $\Delta\Delta G$  of the amino acids were then recalculated relative to Ala, as shown in Table 1.2. Ala is a strong helix former because it has a small side-chain that cannot interact significantly with other side-chains; hence Ala has the highest helical propensity. Helix formation by Ala is thus stabilised predominantly by the backbone, i.e. peptide hydrogen bonds (Chakrabartty *et al.*, 1994; Pace and Scholtz, 1998). In contrast to Ala, Pro is mostly avoided in the interior.

Parameter values of amino acids at internal positions from some previous studies are compared in Table 1.2. The  $s$  values from the Zimm-Bragg model as derived by the Scheraga group (Wojcik *et al.*, 1990) do not agree with the amino acid helical propensities. For example, the  $s$  value of Wojcik *et al.* is the highest for Glu, suggesting Glu as a helix former, while Ala is a helix neutral. In contrast, the data of Rohl *et al.* (1996) shows that Ala is a helix former while Glu is a helix neutral, particularly when negatively charged. However, both series agree that Pro and Gly are helix breakers (highest  $\Delta\Delta G$  value or lowest  $s$  value). The discrepancy of Ala preference from both approaches was initially attributed to solubility (Vila *et al.*, 1992; Vila *et al.*, 2000). In fact, there is a length-dependent term in the helicity of polyalanine. For example, the helix propensity of Ala is low in six-residue poly-Ala. Extending the length beyond six residues gives a dramatic increase in helix propensity (Kennedy *et al.*, 2002). This could explain the discrepancy of the Ala preference in the interior.

Comparison between values obtained using a series of short helical peptides of Rohl *et al.* (1996) and using directed mutagenesis of Blaber *et al.* (1993) shows a good agreement, with Ala the strongest helix former residue, while Gly and Pro are helix breakers in the interior. In general, the helix propagation propensities of the amino acids increase when in 40% TFE, with the nonpolar amino acids showing the largest increases, although not uniformly. Pro and Gly are still helix breakers in 40% TFE (see section 1.8).

The relationship of amino acid substitutions in the isolated  $\alpha$ -helical peptides and in proteins shows a very good correlation. Myers *et al.* (1997) compared the helix propensities of nonpolar amino acids in an  $\alpha$ -helix in ribonuclease T1 and in a 17-residue model peptide with identical sequence. The results show that helix

propensities of the nonpolar amino acids are identical in peptide helices and intact proteins. This suggests that studying factors that affecting  $\alpha$ -helix stability could be used, to some extent, to predict the behaviour of helix formation in the native proteins.

### 1.5.2. Amino acids preference at helix termini

The unique structural preferences for the first four residues at the N terminus, N-cap, N1, N2 and N3 can be explained as a function of hydrogen bonding interactions to the free N1, N2 and N3 backbone NH groups. The N-cap, which is found to initiate and stabilise the helix, is the most selective in its choice of residues. The preferences at the N-cap are determined by hydrogen bonding of side-chains to the exposed backbone NH groups (Penel *et al.*, 1999).

Serrano and Fersht studied the capping preferences at the N-cap by mutating Thr residues at the N-cap of two helices in Barnase (Serrano and Fersht, 1989). They found that negatively charged residues were favoured at the N-cap with a rank order of Asp > Thr > Glu > Ser > Asn > Gly > Gln > Ala > Val. The presence of a negative charge at N-cap adds to the stabilisation energy because of the interaction mainly with the helix dipole. Experimentally, Asn destabilises the helices by 1.3kcal/mol relative to Thr, contradicting the statistical survey result that Asn is one of the most frequently found N-caps in proteins (Richardson and Richardson, 1988 ; Penel *et al.*, 1999). Bell *et al.* (1992) found the N-cap preferences in T4 lysozyme in the rank order of Thr > Ser > Asn > Asp > Val = Ala > Gly. They suggested that Asn can be inherently as good N-cap as Ser or Thr, but it requires a change in backbone dihedral angles of N-cap residues which might be altered in native proteins as a result of tertiary contacts. Indeed, Asn is the most stabilising residue at the N-cap in a peptide model in the absence of tertiary contacts and other side-chain interactions (see below).

The Kallenbach group substituted several amino acids at the N-cap position in peptide models in the presence of a capping box (Lyu *et al.*, 1992). They found that Ser and Arg are the most stabilising residues with  $\Delta\Delta G$  relative to Ala of -0.74 and -0.58kcal/mol, respectively, whilst Gly and Ala are less stabilising. The results are in

agreement with the results of Forood *et al.* (1993) that the stabilising trend in the  $\alpha$ -helix is in the order of Asp > Asn > Ser > Glu > Gln > Ala. A more comprehensive work to determine the preferences for all 20 amino acids at the N-cap position used isomeric peptides with a sequence of XAKAAAKAAAKAAGY-CONH<sub>2</sub>, where X is the substituting N-cap (Doig and Baldwin, 1995). N-capping free energies range from Asn (best) to Gln (worst) (Table 1.3). The acetyl group is nearly as effective as Asn to cap the helices at the N-terminus. Short polar residues such as Thr, Ser, Asp, and Asn have a strong tendency to form side-chain to backbone hydrogen bonds. When these residues are capping the N-terminus, their side-chains form hydrogen bond with neighbouring backbone amides of the N3 or N2 residue. However, this is not the case when they are in the middle of the helix since their side-chains compete with the backbone CO carbonyl for the same hydrogen-bonding partner (i.e. the backbone amide) and tend to break backbone carbonyl-amide hydrogen bonds (Vijayakumar *et al.*, 1999).

In the same way as the N-cap study, the Doig group has used a series of peptide models to probe the preferences at N1 (Cochran *et al.*, 2001) and N2 (Cochran and Doig, 2001) using peptides with sequences of Ac-XAAAAQAAAAQAAGY-NH<sub>2</sub>, and Ac-AXAAAAKAAAKAAGY-NH<sub>2</sub>, respectively. The results have given N1 and N2 preferences for most amino acids for these positions (Table 1.3) and these agree well with preferences seen in protein structures (Kumar and Bansal, 1998; Penel *et al.*, 1999), with the interesting exception of Pro at N1, where Pro is preferred statistically but not empirically. The Serrano group used a similar approach to obtain the N1, N2 and N3 preferences for non-polar and uncharged polar residues by applying AGADIR to experimental helical peptide data, and gives almost identical results (Petukhov *et al.*, 1998; Petukhov *et al.*, 1999).

**Table 1.3.** Amino acid propensities at N- and C-terminal positions of  $\alpha$ -helix.

Amino Acid	$\Delta\Delta G$ relative to Ala for transition from coil to the position (kcal/mol)											
	N-cap	N1		N2		N3	C3	C2		C1	C-cap	C'
	(Doig and Baldwin, 1995) <sup>1</sup>	(Cochran <i>et al.</i> , 2001) <sup>2</sup>	(Petukhov <i>et al.</i> , 1998); (Petukhov <i>et al.</i> , 1999) <sup>3</sup>	(Cochran and Doig, 2001) <sup>4</sup>	(Petukhov <i>et al.</i> , 1998); (Petukhov <i>et al.</i> , 1999) <sup>5</sup>	(Petukhov <i>et al.</i> , 1998); (Petukhov <i>et al.</i> , 1999) <sup>6</sup>	(Petukhov <i>et al.</i> , 2002) <sup>7</sup>	(Ermolenko <i>et al.</i> , 2003) <sup>8</sup>	(Petukhov <i>et al.</i> , 2002) <sup>9</sup>	(Petukhov <i>et al.</i> , 2002) <sup>10</sup>	(Doig and Baldwin, 1995) <sup>11</sup>	(Thomas <i>et al.</i> , 2001) <sup>12</sup>
A	0	0	0	0	0	0	0	0	0	0	0	0
C <sup>o</sup>											0.2	
C <sup>-</sup>	-1.4	1.0		0.9								
D <sup>o</sup>		0.5		0.7							0.2	0.3
D <sup>-</sup>	-1.6	0		-0.2								
E <sup>o</sup>		1.0		-0.2							-0.4	0.3
E <sup>-</sup>	-0.7	0.1		-0.4							-0.5	
F	-0.7	1.4		0.9				0.6				0.1
G	-1.2	1.0	0.7		0.4	0.8	2.1	1.0	0.6	0.4	0.1	-1.1
H <sup>o</sup>	-0.7	0.7		0.8								
H <sup>+</sup>											-0.2	-0.9
I	-0.5	0.5	0.4	0.6	0.5	0.5	0.2	0.2	0.4	0.5		1.5
K <sup>+</sup>	0.1	0.7		0.9							-0.1	-0.1
L	-0.7	0.4	0.2	0.5	0.5	0.4		0.1			-0.1	
M	-0.3	0.5	0.1	0.7	0.3	0.4		0.1			-0.3	0.1
N	-1.7		0.6	1.7	0.7	0.7	0.5	0.7	0.4	0.3	0.1	-0.4
P	-0.4	0.6	0.5									1.2
Q	2.5	0.5	0.3	0.5	0.3	0.2	0.2	-0.02	0.2	0.05	-0.5	-0.1
R <sup>+</sup>	-0.1	0.7		0.8							-0.4	-0.2
S	-1.2	0.4	0.4	0.7	0.5	0.6	0.6	0.5	0.7	0.5	0.8	0.3
T	-0.7	0.5	0.5	0.5	0.5	0.6	0.8	0.6	0.5	0.8		1.1
V	-0.1	0.6		0.5	0.4	0.4	0.3	0.4	0.7	0.6	0.9	1.6
W	-1.3	0.4		0.8								0.7
Y	-0.9										-2.2	

1. XAKAAAAKAAAAKAAGY-NH<sub>2</sub>.
2. CH<sub>3</sub>CO-XAAAAQAQAAGY-NH<sub>2</sub>
3. CH<sub>3</sub>CO -XAAAAAAAAARAAAGGY-NH<sub>2</sub>
4. CH<sub>3</sub>CO-AXAAAAKAAAAKAAGY-NH<sub>2</sub>
5. CH<sub>3</sub>CO -AXAAAAAAAAARAAAGGY-NH<sub>2</sub>
6. CH<sub>3</sub>CO -AAXAAAAAAAAARAAAGGY-NH<sub>2</sub>
7. YGGSAAKEAAAAAAXAA-NH<sub>2</sub>
8. Substitution of residue 32 (C2 position) of  $\alpha$ -helix of ubiquitin.
9. YGGSAAKEAAAAAAXA-NH<sub>2</sub>
10. YGGSAAKEAAAAAAX-NH<sub>2</sub>
11. CH<sub>3</sub>CO-YGAAKAAAAKAAKAX
12. Substitution of residue 35 (C' position) of  $\alpha$ -helix of ubiquitin.

N1, N2 and N3 preferences are distinct from preferences for the preceding N-cap and  $\alpha$ -helix interior. N2 preferences are similar to those at the N1. Glu<sup>-</sup>, Asp<sup>-</sup> and Ala are among the best residues at N1 and N2 both statistically and energetically. Notably the good N2 amino acid residues preferentially form *i,i* or *i,i+1* hydrogen bonds to the backbone though this is hindered by good N-caps (Asp, Asn, Ser, Thr and Cys) that compete for these hydrogen bond donors. These hydrogen bonds are weak due to poor geometry (Penel *et al.*, 1999).

A study of statistical preferences for amino acid residues at the N3 position reveals that Glu, Gln, Asp and Ala are favoured at this position (Penel *et al.*, 1999). Energetic preferences for non-polar and non-charged polar residues at this position in helical peptides have previously been studied (Petukhov *et al.*, 1998; Petukhov *et al.*, 1999). There are a number of specific side-chain to side-chain interactions involving this position that might alter the energetic preferences including the capping box and N-Pro box. Charged residues also show a unique preference at this site as they can potentially interact with free backbone amides and helix dipole. Therefore, a complete survey of all 20 residues in context-free peptides of similar sequence is needed. The energetic preference for this position is one of the objectives of this thesis. The results are shown in Chapter 4.

Although it is also unique in terms of the presence of unsatisfied backbone hydrogen bonds, the C-terminus region is less explored experimentally. The C-terminus tends to fray more than the N-terminus, making C-terminal measurements less accurate. Preferences at the C-cap position differ from those at the N-cap. At the N-terminus, the helix geometry favours side-chain to backbone hydrogen bonding, so polar residues are preferred (Doig *et al.*, 1997; Penel *et al.*, 1999). At the C-terminus unsatisfied backbone hydrogen bonds are fulfilled by interactions with backbone groups upstream of the helix (see Section 1.4). Zhou *et al.* (1994) found that Asn is the most favoured residue at the C-cap followed by Gln > Ser~Ala > Gly~Thr. Forood *et al.* (1993) substituted several amino acids at C1 position of the C-terminus and found a rank order of Arg > Lys > Ala. Doig and Baldwin (1995) determined the C-capping preferences for all 20 amino acids in  $\alpha$ -helical peptides. The thermodynamic propensities of some amino acids at C', C-cap, C1, C2 and C3 (Petukhov *et al.*, 2002; Ermolenko *et al.*, 2003) are also included in Table 1.3.

### 1.5.3. Non-covalent side-chain interactions

Many efforts have been done on studying the stabilising effects of interactions between amino acid side-chains in  $\alpha$ -helices, although some have only been analysed qualitatively. As described earlier, residue side-chains spaced  $i, i+3$  and  $i, i+4$  are on the same face of the  $\alpha$ -helix (see Figure 1.1B), so they can interact with each other. The  $i, i+4$  is stronger than the  $i, i+3$  spacing, so they are considered more important. A summary of stabilising energies for  $i, i+4$  side-chain interactions is given in Table 1.4. Only those that have been measured in helical peptides with the side-chain interaction energies are presented. The energies are grouped according to the type of interactions (see below).

#### 1.5.3.1. Salt bridges

When amino acid side-chains of opposite charges are in close proximity, they can form an ion pair, which is also called a salt bridge. Among the interactions that stabilise the native proteins, the role of electrostatic interactions has been difficult to quantify precisely. This is partly because salt bridges are frequently accompanied by hydrogen bonding that complicates the energetic analysis. However, the importance of ionic components can be measured by pH titrations, by analysing the difference in protein/peptide stability at high and low pH (see section 1.9).

The energetic role of salt-bridges in protein structure stability is still controversial. Solvent-exposed salt-bridges contribute only marginally to protein stability (Horovitz *et al.*, 1990; Daopin *et al.*, 1991; Spek *et al.*, 1998). In contrast, a buried salt-bridge stabilises the native state by up to 5kcal/mol (Anderson *et al.*, 1990; Tissot *et al.*, 1996). Studies on solvent exposed salt bridges in isolated  $\alpha$ -helical peptides give only a modest stabilising energy (Marqusee and Baldwin, 1987; Lyu *et al.*, 1989; Horovitz *et al.*, 1990; Gans *et al.*, 1991; Merutka and Stellwagen, 1991; Stellwagen *et al.*, 1992; Huyghues-Despointes *et al.*, 1993; Scholtz *et al.*, 1993; Huyghues-Despointes and Baldwin, 1997), as shown in table 1.4. The strength of salt-bridges has been found to be determined by several factors: (a) The geometry and distance of the interactions (Kumar and Nussinov, 2002); (b) The degree of exposure to solvent (Takano *et al.*, 2000); (c) The effect of neighbouring residues

(Makhatadze *et al.*, 2003). Some of energetic of single salt bridges so far studied are listed in table 1.4.

Salt bridges in proteins are more abundant in thermophiles rather than in mesophiles (Elcock, 1998; Kumar *et al.*, 2000). For example, salt bridges in the thermophile *P. furiosus* glutamate dehydrogenase are more abundant and highly stabilising compared with those in the mesophile *C. symbiosum* glutamate dehydrogenase that are only marginally stabilising (Kumar *et al.*, 2000). These results suggest that salt bridges play a role in stabilising protein structures at high temperature.

#### 1.5.3.2. Hydrogen bonds

The contribution of hydrogen bonds to protein stability was initially regarded as negligible. This is due to the internal hydrogen bonding energy being largely cancelled by the large free energy cost associated with removing polar groups from water and also with fixing backbone dihedral angles in the helical conformation (Yang and Honig, 1995). Nevertheless, this hypothesis has been challenged. Although they are small in term of energy, hydrogen bonds can significantly stabilise proteins, provided that their geometric requirements can be achieved (Myers and Pace, 1996).

Some experiment utilising isolated  $\alpha$ -helical peptides have yielded energetics of the hydrogen bond interactions (Scholtz *et al.*, 1993; Huyghues-Despointes *et al.*, 1995; Huyghues-Despointes and Baldwin, 1997; Stapley and Doig, 1997b). A pair of Gln-Asn spaced  $i, i+4$  in an Ala-based helical peptides stabilises the helix through the side-chain to side-chain interaction when placed either in the middle, N- or C-terminus. However, Stapley and Doig (1997) found that the interaction energy of Asn-Gln is lower than that of Gln-Asn, providing evidence that the free energy of interaction is sensitive to the geometry of the interacting groups. An analogous experiment on a Gln and Asp  $i, i+4$  pair shows that the Gln-Asp orientation was preferred over the Asp-Gln orientation (Huyghues-Despointes *et al.*, 1993b). Side-chain interactions between Gln and Asp stabilise the helix by hydrogen-bond, which

is specific for both the orientation and spacing of the residues (Huyghues-Despointes *et al.*, 1995).

### 1.5.3.3. Hydrophobic interactions

Hydrophobic interactions are arguably the most important forces in the formation and stabilisation of secondary structure elements. There is a tendency that hydrophobic residues are involved in initiation and termination of an  $\alpha$ -helical segment since they are observed frequently in capping interactions. For example, the association of Ile and Leu at N' with hydrophobic residues at the N4 position is seen in the N-Pro box motif (see section 1.3). There is also a significant correlation between the burial of residues at the C-cap position (C-cap is on average 70% buried) and their hydrophobicity. Thus, the hydrophobic interactions are important for C-capping. The interactions between helical residues C3 or C4 and residues outside the helical segment toward the C-terminus are frequently enhanced by the hydrophobic interactions of C-cap residues (Ermolenko *et al.*, 2002).

Some hydrophobic interactions previously measured are listed in Table 1.4 (Padmanabhan and Baldwin, 1994a; Padmanabhan and Baldwin, 1994b; Stapley *et al.*, 1995; Viguera and Serrano, 1995b). Andrew *et al.* (2001) showed that interactions involving nonpolar residues (Val or Ile) bonding to polar residues (Lys or Arg) placed  $i, i+4$  in  $\alpha$ -helices are stabilising. The interactions are believed to be hydrophobic with contacts between Val or Ile and the alkyl groups in Arg or Lys.

### 1.5.3.4. Other non-covalent interactions

The cation- $\pi$  interaction is increasingly recognised as important in protein stability. This non-classical non-covalent interaction occurs when positively charged side-chain groups interact with delocalised  $\pi$ -electrons. Considering this, they potentially occur when Phe, Tyr or Trp are close to amino acids with a positive charge (His, Arg or Lys). The stabilising effect is seen when Trp and Arg are spaced  $i, i+4$  apart only in the N  $\rightarrow$  C direction in an isolated peptide model. This interaction is not seen for a pair of Phe-Arg (Olson *et al.*, 2001b). Similarly, the interaction



between Trp and His spaced  $i, i+4$  apart, especially when His is protonated, is also stabilising (Fernández-Recio *et al.*, 1997).

**Table 1.4.** Energetics of side-chain  $i, i+4$  interactions.

Interactions	Pairs ( $i, i+4$ )	$\Delta G$ (kcal/mol)	$p$ values*	Notes	Reference
Salt bridge	Glu-Lys	-0.47	2.4	Pairs fully charged	(Scholtz <i>et al.</i> , 1993)
	Lys-Glu	-0.46	2.3		
	Glu-Lys	-0.42	2.15	Pairs fully charged	(Smith and Scholtz, 1998)
	Lys-Glu	-0.40	2.09		
	Asp-Lys	-0.24	1.6		
	Lys-Asp	-0.58	2.9		
Salt bridge/H-bond	Glu-His	-0.19	1.4		(Smith and Scholtz, 1998)
	His-Glu	-0.65	3.3		
	His-Asp	-2.38	80.4		(Luo and Baldwin, 1999)
H-bond	Gln-Asn	0.55	0.4		(Stapley and Doig, 1997b)
	Asn-Gln	0	1.0		
	Gln-Asp	-0.97	6.0		(Huyghues-Despointes <i>et al.</i> , 1995)
	Gln-Glu	-0.31	1.8		(Scholtz <i>et al.</i> , 1993)
Hydrophobic/electrostatic	Phe-Met	-0.65	3.3		(Viguera and Serrano, 1995b)
	Met-Phe	-0.2	1.4		
	Phe-Cys	2	0.0		
	Cys-Phe	-0.6	3.0		
Hydrophobic	Phe-Met	-0.75	4.0		(Stapley <i>et al.</i> , 1995)
	Met-Phe	-0.54	2.7		
	Tyr-Leu	-0.44	2.3		(Shalongo and Stellwagen, 1995)
	Tyr-Val	-0.31	1.8		
	Leu-Tyr	-0.65	3.3		
	Ile-Lys	-0.22	1.5		(Andrew <i>et al.</i> , 2001)
	Val-Lys	-0.25	1.6		
	Ile-Arg	-0.22	1.5		
Cation- $\pi$	Trp-Arg	-0.4	2.1		(Shi <i>et al.</i> , 2002b)
	Arg-Trp	na		< Trp-Arg	
	Phe-Lys	-0.14	1.3		(Andrew <i>et al.</i> , 2002a)
	Lys-Phe	-0.1	1.2		
	Phe-Arg	-0.18	1.4		
	Arg-Phe	-0.1	1.2		
	Tyr-Lys	-0.1	1.2		
	Phe-Lys	0	1.0		(Tsou <i>et al.</i> , 2002)
	Trp-His	-1.0	6.3	pH 5.0	(Fernández-Recio <i>et al.</i> , 1997)
		0	1.0	pH 9.5	
	Phe-Arg	0	1.0		(Olson <i>et al.</i> , 2001a)
CH---OH	Glu-Phe	-0.53	2.7		(Olson <i>et al.</i> , 2001a)
Aromatic	Phe-Phe	-0.27	1.6	Interior	(Butterfield <i>et al.</i> , 2002)
		-0.8	4.4	C-terminus	
Charge repulsion	Lys-Lys	0.17	0.7		(Stellwagen <i>et al.</i> , 1992)

\* $p$  is the equilibrium constant for  $i, i+4$  side-chain interactions calculated using  $\Delta G = -RT \ln p$  (see Chapter 2).

Residues containing aromatic rings (Phe, Tyr, Trp) can interact with hydrogen backbone amides to form weak polar interactions, known as Ar-NH interaction. The occurrence of these interactions at the N-terminus is higher than in the other parts of helices because nonpolar side-chains are buried in the interior. Ar-HN interactions favour the orientation where Ar residues precede the successive residue in the direction of N  $\rightarrow$  C (Toth *et al.*, 2001). Another non-conventional non-covalent interaction is CH - OH bond. For instance, A pair of Glu-Phe in the N  $\rightarrow$  C direction contribute to the stability of the helix by the contact of OH Glu with HC from cyclic ring of Phe (Olson *et al.*, 2001a).

#### 1.5.4. Covalent side-chain interactions

Disulfide bonds and lactam bridges are examples of covalent side-chain interactions affecting helix stability. In theory, given the correct side-chain spacing between the interacting residues, disulphide bonds are helix stabilising by constraining the side-chains to be close, reducing the entropy of non-helical states. The effects of disulphide formation on helix stability, however, have been rarely studied. A model  $\alpha$ -helical peptide was stabilised by an  $i,i+7$  disulphide bond (Jackson *et al.*, 1991). The disulphide bond formation requires oxidation of the Cys residues, promoted by oxidising agent such as  $H_2O_2$ . Excess oxidising agents, however, could perturb the analysis. For example, the presence of  $H_2O_2$  perturbs CD spectra, increasing errors in the sample measurement. The rate of disulphide formation is strongly dependent on the temperature and pH of the solution. Higher pH and temperature require less time for the reaction to complete (Berezhkovskiy *et al.*, 1999).

Lactam (amide) bonds formed between  $NH_3^+$  and  $CO_2^-$  side-chains can stabilise  $\alpha$ -helix structure. Lactam bridges between Lys-Asp, Lys-Glu and Glu-Orn spaced  $i, i+4$  have been introduced into analogues of human growth hormone releasing factor (Campbell *et al.*, 1995), and proved to be stabilising with Lys-Asp most effective. The same Lys-Asp  $i,i+4$  lactam was stabilising in other helical peptide systems (Chorev *et al.*, 1991; Bouvier and Taylor, 1992; Osapay and Taylor, 1992; Kapurniotu and Taylor, 1995; Mayne *et al.*, 1998), while Lys-Glu  $i,i+4$  lactam

bridges were less effective (Osapay and Taylor, 1990). Two overlapping Lys-Asp lactams were even more stabilising (Bracken *et al.*, 1994). Lactams between side-chains spaced  $i, i+7$  (Chen *et al.*, 1993; Zhang and Taylor, 1996) or  $i, i+3$  (Luo *et al.*, 1994; Zhang and Taylor, 1996), spanning two or one turns of the helix have also been reported.

### 1.5.5. Metal binding

Metal ions play a key structural and functional role in many proteins. The stabilisation of folded forms of proteins can be achieved via metal binding which subsequently reduces the entropy. There is therefore interest in using peptide models containing metal complexes. Again, the side-chains that will bind to the metal should be spaced appropriately, e.g.  $i, i+4$  or  $i, i+3$  spacing so that metals can be bound in the helical conformation. The rule of thumb in metal-ligand binding is that "hard metals" prefers "hard ligands". For example Ca and Mg prefer ligands with oxygen as the coordinating atoms (Asp, Glu) (Jernigan *et al.*, 1994). In contrast, soft metals, such as Cu and Zn, bind mostly to His, Cys and Trp ligands and sometimes indirectly via water molecules (Alberts *et al.*, 1998).

In studies of helical peptide models, soft ligands have been mostly used. In the presence of Cd ions, a synthetic peptide containing Cys-His ligands  $i, i+4$  apart at the C-terminal region promoted helicity from 54% to 90%. The helicity of a similar peptide containing His-His ligands increased by up to 90% as a result of Cu and Zn binding (Ghadiri and Choi, 1990). The addition of a cis-Ru(III) ion to a 6-mer peptide, Ac-AHAAHA-NH<sub>2</sub>, changed the peptide conformation from random coil to 37% helix (Kise and Bowler, 2002). An 11-residue peptide was converted from random coil to 80% helix content by the addition of Cd ions, although the ligands used were not natural amino acids but aminodiacetic acids (Ruan *et al.*, 1990). As(III) stabilizes helices when bound to Cys side-chains spaced  $i, i+4$  by -0.7 to -1.0 kcal/mol (Cline *et al.*, 2003). Siedlecka *et al.* (1999) suggested that helix structure of the calcium-binding loop from EF-hand proteins saturated with a lanthanide ion promotes a rigid short helical conformation at its C-terminus region. This was done by introducing artificial ligands composed of a few residues fixed in a helical conformation.

Many common biological metal ligands such as Zn and Cu have a tetrahedral binding geometry. Therefore, there is a strong possibility that they can induce helical peptides to dimerise when present, which will cause problems with measurements of helicity when trying to derive energetic parameters.

### 1.5.6. Phosphorylation

Phosphorylation is one of the control mechanisms in biological systems. For example protein kinase C, whose function is to transfer phosphate groups to other proteins, is also phosphorylated in an ordered phosphorylation cascade to recover its function (Newton, 2003). The energetics of serine phosphorylation in short helical peptide has been studied. In general, in comparison with serine, phosphoserine is stabilising at the helix N-terminus positions, but destabilising at the interior of the helix, most notably for the -2 charge state (Szalik *et al.*, 1997; Liehr and Chenault, 1999; Andrew *et al.*, 2002b). The difference in stabilising effect can be explained in terms of the balance of protein solvation, favorable interactions, and dehydration. Favourable electrostatic interactions between the phosphate group and positively charged residues most likely enhance helix stability. Errington and Doig, (2005) investigated the effect of phosphoserine-Lys  $i, i+4$  interactions in the interior of an  $\alpha$ -helix. The free energy of interaction can overcome the intrinsically low preference of phosphoserine for in the interior. The stabilisation contribution is proposed as the strongest yet recorded in an isolated  $\alpha$ -helix.

### 1.6. Ionic strength

Electrostatic interactions between charged side-chains and the helix macrodipole with opposite charges are helix stabilising (Aqvist *et al.*, 1991; Armstrong and Baldwin, 1993; Huyghues-Despointes *et al.*, 1993b). The interactions are potentially quite strong, but are alleviated by the screening effects of water, ions, and nearby protein atoms. In theory, increasing ionic strength of the solvent (up to 1.0M) should shift the equilibrium between the  $\alpha$ -helix (with a large dipole moment) and the random coil (with a small dipole moment due to almost randomly oriented dipoles), in favour of helix formation (Scholtz *et al.*, 1991a).

The energetics of the interaction between fully charged ion pairs can be diminished by added salt and completely screened at 2.5M NaCl (Huyghues-Despointes *et al.*, 1993b; Smith and Scholtz, 1998). In peptides containing side-chain to side-chain interactions, the effect of ion pairs and charge-helix dipole interactions cannot be clearly separated. There are, however, indications that the interactions of charged residues with the helix macrodipole are less affected than those between charged side-chains (Lockhart and Kim, 1993; Smith and Scholtz, 1998).

### 1.7. Temperature

Thermal unfolding experiments show that the helix unfolds with increasing temperature (Schellman, 1980; Scholtz *et al.*, 1991b; Yoder *et al.*, 1997; Huang *et al.*, 2001), indicating that helix formation is an enthalpy driven process. Enthalpy and entropy changes for the helix-coil transition are difficult to determine as the helix-coil transition is very broad, precluding accurate determination of high and low temperature baselines by calorimetry (Scholtz *et al.*, 1991b). Nevertheless, isothermal titration calorimetric studies of a series of peptides that form helix when binding a nucleating  $\text{La}^{3+}$ , find  $\Delta H$  for helix formation to be -1.0kcal/mol (Siedlecka *et al.*, 1999; Goch *et al.*, 2003), in good agreement with the earlier work.

### 1.8. Trifluoroethanol

There is a dependency of helix interior propensities on the solvent environment (see table 1.2). Peptides with sequences of helices in proteins usually show low helix contents in water. However, adding 2,2,2-trifluoroethanol (TFE) induces helix formation (Nelson *et al.*, 1986; Nelson and Kallenbach, 1989; Sonnichsen *et al.*, 1992; Jasanoff and Fersht, 1994; Waterhous and Johnson, 1994; Albert and Hamilton, 1995). For many peptides, the concentration of TFE used to increase the helix content is only effective up to 40%.

The difference in helix propensity of amino acids in water and TFE is apparent. The propagation propensities of all amino acids increase by 40% in TFE relative to water. In 40% TFE, the interior propensities of the non-polar amino acids increase greatly, leaving charged and neutral polar,  $\beta$  substituted amino acids as helix breakers

(Rohl *et al.*, 1996). In addition, 40% TFE dramatically alters electrostatic (and polar) interactions and increases the dependence of helix propensities on the sequence (Myers *et al.*, 1998).

TFE may act by shielding CO and NH groups from water molecules while leading to hydrogen bond formation between them. The conformational equilibrium thus shifts toward more compact structures, such as the  $\alpha$ -helical conformation (Vila *et al.*, 2000; Starzyk *et al.*, 2005). The mechanism involves interaction between TFE and water with several interpretations. One view suggests that TFE indirectly disrupts the solvent shell on  $\alpha$ -helices (Luo and Baldwin, 1998; Walgers *et al.*, 1998). Another view proposes that TFE destabilises the unfolded species and thereby indirectly enhances the kinetics and thermodynamics of folding of the coiled coil (Kentsis and Sosnick, 1998). A more compromising view suggests that TFE forms clusters in water solution, which at lower concentration pulls the water molecules from the surface of proteins. At higher concentration, TFE clusters associate with appropriate hydrophobic side-chains reducing their conformational entropy and switch the conformation at TFE concentration >40% (Reiersen and Rees, 2000).

### 1.9. pKa values

pKa shifts of charged residues in the helix are significant as they can be used to evaluate the strength of possible interactions they form in water. This is most important for amino acids at the helix termini, because they can potentially interact with unsatisfied hydrogen bonds of the NH groups and CO group at the N-terminus and C-terminus, respectively, or with the helix dipole. The pKa values can be measured accurately from the change in helix stability across a broad range of pH. The asymptotic values of the ellipticities for the different protonation states can be fitted to a Henderson-Hasselbach equation to calculate the pKa.

In general, the negatively charged residues at the N-cap have lower pKa values (Doig and Baldwin, 1995). This may be because side-chains at the N-cap can form strong hydrogen bonds to NH groups of N2 and N3, while the bonds formed by side-chains at N1, N2 and N3 are much weaker (Doig *et al.*, 1997; Penel *et al.*, 1999). At the N1, N2 and N3, the pKa values of Glu and His are normal compared to those in

model compounds, but those for Asp and Cys are shifted to slightly lower values. The strength of salt bridge interactions can also be measured using pH titrations. (Nozaki and Tanford, 1967; Armstrong and Baldwin, 1993; Huyghues-Despointes *et al.*, 1993b; Scholtz *et al.*, 1993; Campbell *et al.*, 1995; Kortemme and Creighton, 1995; Smith and Scholtz, 1998; Cochran and Doig, 2001; Cochran *et al.*, 2001; Miranda, 2003). The negatively charged residues at higher pH destabilise helices when at the C-cap (Doig and Baldwin, 1995). The increased pKa may result from an unfavourable electrostatic interaction with the C-terminal dipole or partial negative charges on the terminal CO groups.

### 1.10. Rotamers

The side-chains of amino acids, except Gly and Ala, have different conformations due to the difference in  $\beta$ -carbon substitution. Energetically, arrangements of two tetrahedrally coordinated carbon atoms in which the substitution of one carbon is in between those of the other when seen along the axis of rotation, are more favoured. The staggered substitution at the  $C_\beta$  avoids being close to larger groups of the  $C_\alpha$ . This is what most is frequently found in real proteins.

The rotation angle between  $C_\alpha$  and  $C_\beta$  is called chi1 ( $\chi_1$ ), between  $C_\beta$  and  $C_\gamma$  is called chi 2 ( $\chi_2$ ) and so on. Except for Ala, Gly and Pro, a  $sp^3$ - $sp^3$  bond of other amino acids has possible rotation angles, called rotamers. The general definitions of these rotamers (Val is excluded) are as follows: *gauche*<sup>+</sup> ( $g^+$ ,  $-60\pm60^\circ$ ), *trans* ( $t$ ,  $180\pm60^\circ$ ), and *gauche*<sup>-</sup> ( $g^-$ ,  $60\pm60^\circ$ ) (Janin *et al.*, 1978). These definitions apply for every  $\chi$  angle with a  $sp^3$  -  $sp^3$  bond.

Based on statistical surveys, all side-chains in  $\alpha$ -helix interiors strongly avoid the  $g^-$  conformation. This restricts the  $\beta$ -branched residues like Thr, Val and Ile to a single conformer,  $g^+$ . Asp, Asn, Met and Ser also overwhelmingly prefer the  $g^+$  conformation. Arg, Cys, Gln, Glu, Leu and Lys can adopt the  $t$  and  $g^+$  conformations roughly equally. Only the aromatic residues, His, Tyr, Trp and Phe prefer the  $t$  conformation (Blaber *et al.*, 1994; Stapley and Doig, 1997a; Penel *et al.*, 1999).

There are unique rotamer preferences at  $\alpha$ -helix N-caps. Thr and Ser at the N-cap adopt the  $g^+$  rotamer and hydrogen bond to the N3 NH. Asp and Asn N-cap side-chains either adopt the  $g^+$  rotamer and hydrogen bond to the N3 NH, or adopt the  $t$  rotamer and hydrogen bond to both the N2 and N3 NH groups. With all other N-caps, the side-chain is found in  $g^+$  (Doig *et al.*, 1997; Penel *et al.*, 1999). A complete new rotamer library has been made available to improve the prediction accuracy of protein structures (Lovell *et al.*, 2000).

Rotamer energy can be calculated from rotamer populations assuming they are in a Boltzmann distribution using  $E = -RT \sum \ln p_i$ , where  $p_i$  is the fraction of rotamer  $i$  when all rotamers are populated (see section 3.10 for details). The rotamer energy is important in opposing protein folding as it has a similar magnitude as conformational entropy (Stapley and Doig, 1997a). In proteins, however, the rotamer preferences of amino acids are not always in their lowest energy state, as they are affected by many factors such as tertiary contacts. The difference between ideal and actual rotamer energy gives strain, called rotameric strain energy, which is also as important as rotamer energy. The rotameric strain energy of amino acids in the helix interior has been previously quantified as 0.42kcal/mol per residue upon folding (Penel and Doig, 2001).

### 1.11. Aim of the thesis

The aim of this thesis is to accurately characterise a range of important interactions, in terms of energy and structure, using  $\alpha$ -helical peptides and survey of protein crystal structures, more specifically:

1. Amino acid preferences at the N3 position of the  $\alpha$ -helix.
2. The behaviour of the CXXC redox and metal binding motif in the context of isolated  $\alpha$ -helical peptides.
3. The energetics of pairwise coupling in an Arg-Phe-Met triplet.
4. The energetics of pairwise coupling in a Glu-Lys-Glu triplet.



The experimental approach is to study poly-Alanine peptides that fold into  $\alpha$ -helices in aqueous solution. Poly-Ala peptides provide a context free environment for substitutions. Circular Dichroism is used mainly for data collection, as it is sufficient to follow the helix-coil transition. As helix-coil transition is not a two-state process, empirical data from CD measurement is to be analysed using Lifson-Roig based helix-coil theory. The theory considers the stability and structure of every possible peptide conformation. The energetics can thus be calculated from the equilibrium constant of helix formation, which relates to peptide helicity. Protein crystal structure survey is used to complement the measurement of free energies of interactions. Observations of frequencies of amino acids in proteins are a useful indication of energetic preferences, but cannot substitute for free energy measurements, however.

### **1.12. Organisation of the thesis**

The thesis is grouped into eight chapters with the following organisation:

1. Chapter 1 is a review about  $\alpha$ -helix structure and factors that affect its stability. This gives a background not only for understanding the helix behaviour, but also for the prediction and data analysis in the later chapters.
2. Chapter 2 explains the helix-coil theory with which the energetic and behaviour of helix formation can be analysed.
3. Chapter 3 deals with the materials and methods used for the experiment.
4. Chapter 4 presents and discusses results on amino acids substitution at the N3 position.
5. Chapter 5 presents and discusses results on the CXXC motif in the context of an isolated  $\alpha$ -helix.
6. Chapter 6 presents and discusses results on pairwise coupling in an Arg-Phe-Met triplet that exhibits cooperative behaviour.
7. Chapter 7 presents and discusses results on pairwise coupling in a Glu-Lys-Glu triplet that exhibits anti-cooperative behaviour.

8. Chapter 8 concludes all the results presented in chapters 4-7 as well as giving suggestions for future work.

## CHAPTER 2

### INTRODUCTION TO HELIX-COIL THEORY

#### 2.1. The helix model

If there are  $N$  residues in a polypeptide, there will be  $2^N$  combinations of structures adopted by the chain. These structures result from two different conformations of the residues in the sequence, e.g. helix ( $h$ ) and coil ( $c$ ) states. The  $h$ -state is the residue with helical  $\phi$  and  $\psi$  dihedral angles, according to the Ramachandran plot (see Figure 1.2 in Chapter 1). The  $c$ -state is defined as the remainder. In aqueous solution,  $\alpha$ -helices cannot be considered as a two-state model, in which the equilibrium is assumed to be 100% helix or 100% coil. This is because they adopt a large number of structures of fully helical, fully coil, and partly helical conformations which are in an equilibrium. The two-state model will thus give serious flaws in the prediction of helix-coil transition.

The thermodynamics of helix formation is based on the geometry of the  $\alpha$ -helix in which the carbonyl CO group  $i$  is hydrogen-bonded to the amide NH group  $i+4$ . Consequently, the formation of a helical hydrogen bond requires the fixing of the  $\phi$  and  $\psi$  angles of three consecutive peptide units. The helix formation involves a free energy ( $\Delta G$ ) which is associated with the initiation and the propagation parameters. The former is responsible for the cooperativity and is usually considered purely entropic. The latter includes the hydrogen bond interaction between peptide groups, the entropy loss resulting when the residue is fixed in the helix and side-chain interactions. These two parameters are essential for thermodynamic formulation of the  $\alpha$ -helix statistical weights (see later sections). The difference in magnitude of both parameters is important in the helix-coil transition. The initiation of a new helix is much more difficult than the propagation of an existing helix. This is because three residues need to be fixed in a helical geometry to form the first hydrogen bond, while adding an additional hydrogen bond to an existing helix requires only one extra residue. Consequently, the value of the initiation parameter will be much lower than that of the propagation parameter.

There is a very important assumption that needs considerations in modelling helix-coil transition. Since helix initiation is difficult, conformations with multiple helical segments are expected to be rare in short helices. Thus polypeptide conformations containing more than one helical segment are sometimes assumed not to be populated. The approach using this assumption is known as the 'single sequence approximation'. When a residue with a high preference for a helix terminus is in the middle of a long sequence, the polypeptide tends to be fragmented. The approximation is thus no longer valid and continuing using it could give serious errors for modelling a sequence that has a high probability to populate more than one helix simultaneously. Conformations with two or more helices may also include tertiary interactions that are not included in all helix-coil models developed to date.

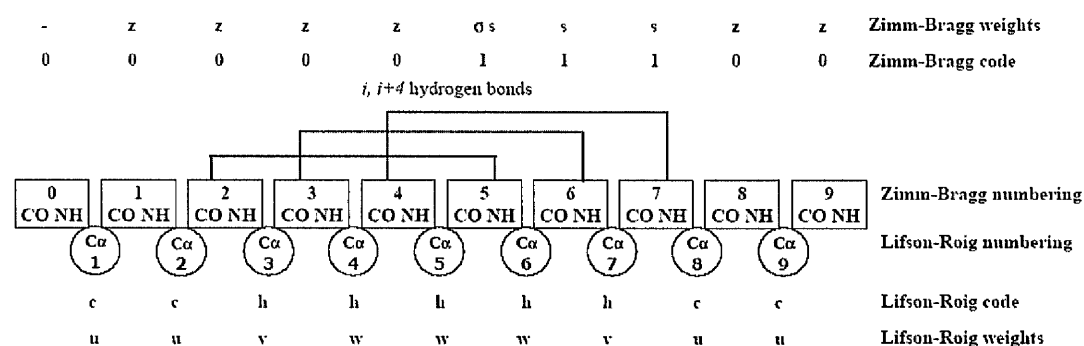
In general, there are two main theories describing the helix-coil transition, namely Zimm-Bragg (Zimm and Bragg, 1959) and Lifson-Roig (Lifson, 1961) theories. The differences of the two main theories originate in the way statistical weights assigned to different residue conformations. Both theories are discussed below.

## 2.2. Zimm-Bragg (ZB) theory

The Zimm-Bragg (ZB) theory (Zimm and Bragg, 1959) is based on whether their NH groups participate in hydrogen bonds within the helix to form peptide groups. A peptide unit is given a code of 1 if its NH group  $i$  forms a hydrogen bond with CO group  $i-3$ , and 0 otherwise. The first hydrogen-bonded unit proceeding from the N-terminus has a statistical weight of  $\sigma$ , successive hydrogen bonded units have weights of  $s$ , and non-hydrogen-bonded units have weights of zero (0). The  $\sigma$  and  $s$  are the initiation and propagation parameters, respectively. The cost of initiation,  $\sigma$ , is paid only once for each helix, while one additional  $s$  value is simply multiplied to include the weight for each extra hydrogen bond in the extending helix. The  $\sigma$  value physically means the probability of three successive residues being restricted into an  $\alpha$ -helical geometry with no intra helix hydrogen bond formation.

The procedure for assigning statistical weights to a peptide sequence is as follows. The polypeptide in Figure 2.1 contains 10 peptide units (numbered from 0-

9) with units number 2 to 7 forming a helix. Although there are six helical peptide units, there are only three hydrogen bonds formed by the NH groups as indicated for positions number 5-7 and therefore only three '1's in the coding. Because the formation of the first hydrogen bond requires three preceding helical residues, the first '1' must be preceded by at least three consecutive '0's. These are the three peptide units with helical dihedral angles at the N-terminal end of the helix which are not NH hydrogen-bonded, as indicated for positions number 2-4.



**Figure 2.1.** Numbering, coding, and statistical weights assignment in a partially helical peptide for the original Zimm-Bragg and Lifson Roig helix-coil models.

### 2.3. Lifson-Roig (LR) theory

Lifson-Roig-based model (Lifson, 1961) describes the helix-coil transition as two-state model on the level of individual residues. In the original theory, each residue in a polypeptide sequence is considered to exist in either helix (*h*) or coil (*c*) state. The numbering is centred on C $\alpha$  atom and each residue has a statistical weight based upon its middle state in a triplet sequence. There are only two non-unity statistical weights assigned for the residues in the sequence. The residue in the *h*-state with at least one of its nearest neighbours is in the non *h*-state has a weight of *v*. The residue in the *h*-state whose nearest N- and C-terminal neighbours are also in the *h*-state has a weight of *w*. The weights for residues in coil are *u*, usually given a value of unity (Table 2.1). The *v* and *w* are the initiation and propagation parameters, respectively.

## 2.4. Comparison between ZB and LR theories

The ZB and LR theories differ in how units in the polypeptide chain are classified. The units in the ZB theory are peptide groups, rather than residues, and they are classified on the basis of whether their NH groups participate in hydrogen bonds within the helix, rather than whether the residue adopts helical  $\phi$ ,  $\psi$  angles. The  $\sigma$  and  $s$  values in the ZB theory are similar (but not identical) to  $\nu^2$  and  $w$  in LR theory, respectively, and can be related by the following formulae:  $s = w/(1+\nu)$ ;  $\sigma = \nu^2/(1+\nu)^4$  (Qian and Schellman, 1992).

In the ZB model, the non helical '0' units are not the same as the  $c$  units of the LR theory. The former represent coil units, which may comprise the helical conformations without hydrogen bonding. Hence, the helical '0' units sometimes have similar weight as the coil '0' units. In contrast, helical conformations are excluded from  $c$  coding in the latter. A further difference is that the ZB model assigns weights of '0' to all conformations that contain a *chc* or *chhc* sequence. This excludes a very large number of conformations that contain a residue with helical  $\phi$ ,  $\psi$  angles but with no hydrogen bond. In the LR theory, these are all considered.

As the  $\sigma$  is not a property of a single peptide group or residue and the value of the  $s$  parameter depends on several residues, it is not possible to assign a  $\sigma$  value to a particular amino acid in a heteropolymer. If a single residue substitution is made in the middle of a peptide sequence, it is not clear whether this leads to a change in the  $s$  value of the preceding residue, the following residue, or both. Substitution for heteropolymers is therefore easier to incorporate in the LR model since the  $w$  and  $\nu$  parameters are assigned to individual residues. The substitution of one amino acid at a certain position only changes the  $w$ - and  $\nu$ -values at that position. For this reason, most recent work has been based on the LR model. The following sections, therefore, give a description of the development of the LR theory.

## 2.5. Partition function

The partition function is defined as the sum of the statistical weights for every conformation in a polypeptide sequence. The role of partition function is central in

the statistical mechanics of helix-coil equilibrium. Any property of the equilibrium can be extracted from the partition function by applying appropriate mathematical functions. The properties include the mean number of residues, the maximum number of hydrogen bonds, the mean helix length, the probability that each residue is within a helix, etc. The way to generate the partition function according to the LR coding and numbering is given below.

**Table 2.1.** Assignment of statistical weights for the central residue in the triplet in Lifson-Roig model.

Conformation	Weight in original theory	Weight in the theory that includes capping (Doig <i>et al.</i> , 1994)
ccc	1	1
cch	1	$n$
chc	$v$	$v$
chh	$v$	$v$
hcc	1	$c$
hch	1	$\sqrt{nc}$
hhc	$v$	$v$
hhh	$w$	$w$

Each residue in the sequence is assigned a statistical weight based on its conformation in the triplet as shown in Table 2.1. The weight assignment in the table above can be rewritten in the form of a matrix (M):

$$M = \begin{matrix} & \begin{matrix} \bar{h}h & \bar{h}c & \bar{c}h & \bar{c}c \end{matrix} \\ \begin{matrix} h\bar{h} \\ h\bar{c} \\ c\bar{h} \\ c\bar{c} \end{matrix} & \begin{pmatrix} w & v & 0 & 0 \\ 0 & 0 & 1 & 1 \\ v & v & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{matrix}$$

For each triplet, the state of the leftmost residue is shown at the start of each row in the matrix while the state of the rightmost residue is shown at the end of the top of each column. The state of the middle residue of each triplet is shown as the

barred residue in both the rows and columns. When the states of the residues in the middle differ (i.e. one is  $c$  while the other is  $h$ ), the entry in the matrix is zero. Otherwise, the matrix gives the weight, taken from Table 2.1. This then allows the generation of the partition function ( $Z$ ) for a polypeptide of  $N$  residues as:

$$Z = \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix} M^N \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

The end vectors are present to ensure that the first and last residues in the peptide cannot have a  $w$ -weighting because the middle residue in the triplet is between two residues in an  $h$ -state which is impossible for terminal residues. Further extensions to helix-coil theory are dealt with in the same way (see later sections), by defining residue weight assignments, rewriting as a matrix and determining end vectors. Identical columns in the matrix above can be combined to simplify the calculation. For example, the third and fourth columns are identical so can be simplified to a  $3 \times 3$  matrix, instead of  $4 \times 4$ .

$$M = \begin{matrix} & \bar{h}h & \bar{h}c & \bar{c}(h \cup c) \\ \begin{matrix} \bar{h}h \\ \bar{h}c \\ \bar{c}(h \cup c) \end{matrix} & \begin{pmatrix} w & v & 0 \\ 0 & 0 & 1 \\ v & v & 1 \end{pmatrix} \end{matrix}$$

Here,  $h \cup c$  shows the residues being combined. The matrix entries for the combined positions are the entries for the non-combined entries, with zero being discounted if combined with a non-zero entry. New end vectors are now required, as their lengths must be the same as the order of the square matrix. This gives a partition function of:

$$Z = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} M^N \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

For a homopolymer with  $N$  hydrogen bonds, the statistical weight of a 100% helix is given by  $v^2 w^{N-2}$ , according to the LR theory (this is  $\alpha^{N-1}$  in the ZB theory).



The population of each conformation is given by the statistical weight of that conformation divided by the sum of the statistical weights for every conformation (i.e. the partition function). To clarify this, the following example is given:

In the triplets without capping preferences, the partition function is given by  $Z = w + 3v + 4$ . This is the sum of all weights for all possible conformations in the triplet, shown in column 2 Table 2.1. The population of the  $hhh$  triplet is thus  $w/(w + 3v + 4)$ . In this way the population of every conformation can be calculated and all properties of the helix-coil equilibrium can be evaluated. For example, the average number of helical hydrogen bonds per molecule is given by:

$$n_h = \frac{\partial \ln Z}{\partial \ln w}$$

Circular dichroism (CD) is commonly used to give fractional helicity,  $f_H$ , (or average helix content) of a peptide, namely the fraction of residues that have a weight of  $w$ . LR based models can thus be related to experimental data by equating the measured mean helix content to:

$$f_H = n_h / N, \text{ where } N \text{ is the number of residues in the peptide.}$$

The greater the statistical weight, the more stable the conformation. Statistical weights can be regarded as equilibrium constants for the equilibrium between coil and helix. They can, therefore, be converted to free energies as  $-RT \ln(\text{weight})$ .

## 2.6. Modification to the original Lifson-Roig theory

As described in Chapter 1, the capping region of helices plays an important role for  $\alpha$ -helix stability. The original LR theory is incapable of predicting the effects of modification in this region. In addition, the effect of side-chain interactions that give changes in helix stability is not considered in the original theory either. The following sub-sections describe the development of the LR model to include capping and side-chain interactions.



### 2.6.1. Modification to include N- and C- cappings

The N-cap is defined as the residue without the  $\phi$ ,  $\psi$  helical backbone angles immediately preceding a helical segment, hydrogen-bound to a residue ahead in the sequence. Similarly, the C-cap is defined as the residue without  $\phi$ ,  $\psi$  helical backbone angles following the last residue following a helical segment. In LR theory, the number of residues is defined as the number of  $\alpha$ -carbons flanked by peptide (CONH) units on both sides. Thus, in a peptide  $\text{NH}_3^+ \text{-ABCDEF-COO}^-$ , only B, C, D, and E count as residues. A and F are, therefore, the N-cap and C-cap, respectively. If the N-terminus is acetylated and the C-terminus is amidated ( $\text{CH}_3\text{CO-ABCDEF-COONH}_2$ ), A and F count as residues.

Because the N-cap residues are considered as a *c*-state in the original LR theory, they make no contribution to the partition function. Substitutions at this position therefore give no effect on helix content. In fact, there are large differences in helix content when the first amino acid is altered in an unacetylated helical peptide (Chakrabartty *et al.*, 1993a). The LR model is incapable of accounting for this phenomenon, and a modified theory is thus necessary. Doig *et al.* (1994) included the N- and C-capping effects in the original LR theory.

The modification principally altered the assignment of weights of the cap residues to give a contribution to the partition function. The N-cap and C-cap residues are assigned weights  $n$  and  $c$ , respectively, instead of  $u$  in the original theory. *cch*- and *hcc*-states, therefore, have weights  $n$  and  $c$ , respectively. A *hch*-state is assumed to have the geometric mean of the  $n$ - and  $c$ -weightings of  $\sqrt{n \cdot c}$  (Table 2.1). This may not be accurate because the *hch*-state is unlikely to occur in reality. Physically, this corresponds to a single residue in the middle of a helix not forming a helical hydrogen bond, thus introducing a kink in the helix. This has been observed with proline (Barlow and Thornton, 1988). Pro may have a relatively high weighting when it is in the centre residue in an *hch*-triplet. It may also occur with Gly as it has a low  $w$ -value. Other weightings are unchanged. If  $n = 1$  and  $c = 1$ , the modified theory reduces to the original theory, where all *c*-states have the same weightings (i.e. columns 2 and 3 in Table 2.1 will have identical weightings). These modified definitions give a new matrix:

$$M = \begin{matrix} & \bar{h}h & \bar{h}c & \bar{c}h & \bar{c}c \\ \begin{matrix} h\bar{h} \\ h\bar{c} \\ c\bar{h} \\ c\bar{c} \end{matrix} & \begin{pmatrix} w & v & 0 & 0 \\ 0 & 0 & \sqrt{n.c} & c \\ v & v & 0 & 0 \\ 0 & 0 & n & 1 \end{pmatrix} \end{matrix}$$

This matrix is not singular and cannot be reduced to a  $3 \times 3$  matrix. The partition function  $Z$  for an  $N$  residue sequence is:

$$Z = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} M^N \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

The initial vector ensures that the N-terminal residue can only take the weightings  $n$  (if it precedes an  $h$ -state) or 1 (if it precedes a  $c$ -state). Similarly, the end vector allows weightings of only  $c$  or 1 for the final residue. The end units cannot be in an  $h$ -state as this requires both  $\phi$  and  $\psi$  dihedral angles to be restricted into an  $\alpha$ -helical geometry. The modified theory has increased the number of units in the partition function by two compared to the original theory. Because these additional units, one at each end, can only be in a  $c$ -state, the maximum number of  $h$ -states is therefore unchanged.

The problem with this model is when dealing with a triplet containing a string of residues with high N-cap preference, high interior preference and low interior preference, for example a triplet of N-cap(Asp) - N1(Ala) - N2(Pro). The triplet will have a high population for the  $chc$  conformation. This is incorrect, however, as the Asp N-cap preference results from hydrogen bonding to the N3 position, so should not apply to  $chc$  and  $chhc$  conformations. The definition of the N-cap weight, therefore, applies only to the  $c$  residue in a  $chhh$  quartet as proposed by Andersen and Tong (1997) and Rohl and Baldwin (1998). Consequently, these modifications give different N-cap and helix interior energies after fitting.

### 2.6.2. Modification to include side-chain interactions

Some attempts have also been done to include side-chain interactions in the LR model (Stapley *et al.*, 1995; Sun and Doig, 1998). Only the model of Stapley *et al.* is discussed below as it assumes polypeptide sequences adopt pure  $\alpha$ -helix conformations. The model of Sun and Doig deals with side-chain interactions in  $3_{10}$ -helix and a mixture of  $3_{10}$ - and  $\alpha$ -helix structures.

As the  $i, i+4$  and  $i, i+3$  side-chain interactions are the most relevant interactions in the  $\alpha$ -helix, the original LR model incorporate parameters  $p$  and  $q$  for the  $i, i+4$  and  $i, i+3$  interactions, respectively. The free energy of interaction is related to these parameters via the relations  $\Delta G_{(i, i+4)} = -RT \ln p$  and  $\Delta G_{(i, i+3)} = -RT \ln q$ , where  $p$  and  $q$  are equilibrium constants for formation of an interaction in an  $\alpha$ -helix. For  $i, i+4$  interactions to occur, the model requires five consecutive residues to be in the  $h$  state. One residue contributes  $w.p$  to the partition function instead of  $w$ . For mathematical simplicity, the weight  $w.p$  is assigned to the central residue in the  $h h h h h$  quintet, where the first and last residues are the interacting residues. For  $i, i+3$  interactions, four residues must be in the  $h$  state. The central residue in the quintet  $h h h h c$ , where the first and fourth residues are the interacting ones, is assigned weight  $w.q$ . When  $i, i+3$  and an  $i, i+4$  interactions are present simultaneously, the central residue of the  $h h h h h$  quintet is assigned weight  $w.q.p$ . The statistical weight of each residue in the helix can thus be assigned by correlating the conformations of the residue of interest and the two nearest N-terminal and two nearest C-terminal neighbours. A  $16 \times 16$  matrix can be generated to include all possible conformations of five consecutive residues. As many of the quintets have statistical weights of '0', the matrix can be simplified to  $6 \times 6$ :

$$M = \begin{matrix} & \overline{hhhh} & \overline{hhhc} & \overline{hhc(h \cup c)} & \overline{(h \cup c)ch(h \cup c)} & \overline{ch(h \cup c)(h \cup c)} & \overline{(h \cup c)cc(h \cup c)} \\ \begin{matrix} \overline{hhhh} \\ \overline{hhhc} \\ \overline{hhc(h \cup c)} \\ \overline{(h \cup c)ch(h \cup c)} \\ \overline{ch(h \cup c)(h \cup c)} \\ \overline{(h \cup c)cc(h \cup c)} \end{matrix} & \begin{pmatrix} w.p.q & w.q & 0 & 0 & 0 & 0 \\ 0 & 0 & v & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{n.c} & 0 & c \\ 0 & 0 & 0 & 0 & v & 0 \\ w & w & v & \sqrt{n.c} & 0 & c \\ 0 & 0 & 0 & n & 0 & 1 \end{pmatrix} \end{matrix}$$

Entries in the matrix are for residue  $i$ . Row labels are residues  $i-2$ ,  $i-1$ ,  $i$  and  $i+1$ . Column labels are residues  $i-1$ ,  $i$ ,  $i+1$ , and  $i+2$ .  $h \cup c$  indicates the union of the helix and coil conformations. The partition function,  $Z$ , for the system is given by:

$$Z = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \left( \prod_{i=1}^N M \right) \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

When all the interaction parameters are set equal to unity, the partition function is equivalent to the model of Doig *et al.* (1994) in the LR model to include N- and C-cappings above.

### 2.6.3. Modification to include N1, N2 and N3 preferences

As the N1, N2 and N3 positions at the helix N-terminus are unique (see section 1.4), these positions need to be treated explicitly in the helix coil theory. In the original LR model, the N1 and C1 residues are both assigned the same weight,  $v$ , although they are sometimes differentiated as  $v_N$  and  $v_C$  (Shalongo and Stellwagen, 1995; Andersen and Tong, 1997).

A LR-based helix-coil model N1N2N3 (Sun *et al.*, 2000) adds the amino acid preferences for the N1, N2, and N3 positions. This model also incorporates the N- and C-capping preferences. To account for the N1 preference, the  $v$  weight is changed to  $n1$  in the new model. The penalty for helix initiation is now  $n1.v$ , instead of  $v^2$ . An N2 helical residue is assigned a weight of  $n2.w$ , instead of  $w$ . The  $n2$ -value is an adjustment to the weight  $w$  at the N2 position. Similarly, an N3 residue is now assigned the weight  $n3.w$ , instead of  $w$ . Figure 2.2 shows these assignments for a short helix.

Position	N''	N'	N-cap	N1	N2	N3	N4-C4	C3	C2	C1	C-cap	C'	C''
Conformation	c	c	c	h	h	h	h	h	h	h	c	c	c
Weight	<i>u</i>	<i>u</i>	<i>n</i>	<i>n1</i>	<i>n2.w</i>	<i>n3.w</i>	<i>w</i>	<i>w</i>	<i>w</i>	<i>v</i>	<i>c</i>	<i>u</i>	<i>u</i>

**Figure 2.2.** Assignment of statistical weights for residues in the LR model to include N1, N2 and N3 preferences.

In this model, six consecutive residues from N-cap to N5 are needed for statistical weight assignments. Table 2.2 lists all possible conformations and the weight of the fourth residue in the hexamer. Note that the weight is for the amino acid found at the fourth position with the exception of *n*, which is the weight of the first residue in the hexamer. These weights can be written in the form of a 32 x 32 matrix that is simplified by combining identical columns to a 16 x 16 matrix (Figure 2.3):

**Table 2.2.** Assignment of statistical weights to the fourth residue in a hexamer peptide to include the N1, N2 and N3 preferences.

Conformation	Weight	Conformation	Weight	Conformation	Weight	Conformation	Weight
cccccc	1	chcccc	1	hcccc	1	hhcccc	1
ccccch	1	chccch	1	hccch	1	hhccch	1
cccchc	1	chcchc	1	hccchc	1	hhcchc	1
cccchh	1	chcchh	1	hccchh	1	hhcchh	1
ccchcc	<i>n1</i>	chchcc	<i>n1</i>	hchcc	<i>n1</i>	hhchcc	<i>n1</i>
ccchch	<i>n1</i>	chchch	<i>n1</i>	hchch	<i>n1</i>	hhchch	<i>n1</i>
ccchhc	<i>n1</i>	chchhc	<i>n1</i>	hchhch	<i>n1</i>	hhchhc	<i>n1</i>
ccchhh	<i>n1</i>	chchhh	<i>n1</i>	hchhhh	<i>n1</i>	hhchhh	<i>n1</i>
cchccc	1	chhccc	1	hchccc	1	hhhccc	<i>c</i>
cchcch	1	chhcch	1	hchcch	1	hhhccch	<i>c</i>
cchchc	1	chhchc	1	hchchc	1	hhhchc	1
cchchh	1	chhchh	1	hchchh	1	hhhchh	1
cchhcc	<i>n2.v</i>	chhhcc	<i>n3.v</i>	hchhcc	<i>n2.v</i>	hhhchcc	<i>v</i>
cchhch	<i>n2.v</i>	chhhch	<i>n3.v</i>	hchhch	<i>n2.v</i>	hhhchch	<i>v</i>
cchhhc	<i>n2.w</i>	chhhhch	<i>n3.w</i>	hchhhc	<i>n2.w</i>	hhhchhc	<i>w</i>
cchhhh	<i>n2.w</i>	chhhhh	<i>n3.w</i>	hchhhh	<i>n2.w</i>	hhhchhh	<i>w</i>

$$M =$$

**Figure 2.3.** Matrix of statistical weights assignment in the N1N2N3 helix-coil model to include N-terminus preferences.

The partition function of the  $\alpha$ -helix with  $N$  amino acids ( $Z$ ) is given by:

$$Z = U.M.V$$

where:

$$U = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0) \text{ and}$$

$$V = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Residues in the centre of the rare *hch* conformation have a weight of 1. The assignment of weights to very short helical segments requires careful consideration of the structures that can be adopted. An isolated helical residue, in the centre of a *chc* conformation, is assigned a weight  $n1$ , instead of  $v$ . This is because the most important structures adopted by residues at N1 are hydrogen bonds between the N1 side-chain and the N1 backbone NH group (Penel *et al.*, 1999). These interactions can form even if only a single residue is in an *h* conformation. All N1 residues commencing any helical segment have  $n1$ -weightings. The N2 residue in a *chhc* sequence is assigned  $n2.v$ -weighting for the same reason. N2 structures are typically hydrogen bonds between the N2 side-chain and the N2 backbone NH group (Penel *et al.*, 1999). Therefore, these can form even if only two consecutive helical residues



are present. The N3 residue in a *chhhc* sequence has  $n3.v.n$ -weighting since both form a capping box interaction. A capping box requires a minimum of three consecutive helical residues (Harper and Rose, 1993). Interaction between the N3 side-chain and the N-cap backbone is thus assigned a  $n3$ -weighting. Similarly, interaction between the N-cap side-chain and N3 backbone is assigned a  $n$ -weighting. A weight of  $v$  is assigned, instead of  $w$ , as the final helical residue in a sequence does not have an associated hydrogen bond.

#### 2.6.4. Modification to include capping box interaction

The N-terminal capping box includes a side-chain-backbone hydrogen bond between N-cap and N3 (see Chapter 1, Figure 1.2). Rohl and Baldwin (1998) included this interaction in the LR model by assigning a weight of  $w.r$  to the *chhh* conformation, where  $r$  is the weight for the N-cap backbone to the N3 side-chain bond.

#### 2.6.5. Modification to include metal cross link

The modification of the LR model to include metal cross linking was proposed by Kise and Bowler (2002) based on experimental results in a short helical peptide model, Ac-AHAAAHA-NH<sub>2</sub>, where the histidines cross-linked to a Ru(III) ion. Because the cross-link changes the peptide conformation from essentially structureless to 37% helix, the Ala3, Ala4, and Ala 5 between His2 and His6 are all constrained. This increases the initiation parameter  $v$  for these intervening residues by the same amount and they are assigned a new parameter of  $v_{35}$ . The initiation parameter for His residues is assumed a value of  $v_{26}$  which is the geometric mean of  $v_{35}$  and the standard initiation parameter for an alanine-based helix,  $v$  ( $v^2 = 0.0013$  (Rohl *et al.*, 1996)), giving  $(v_{26})^2 = v_{35} \times v$ .

The propagation parameter  $w$  for the three central Ala residues relates to the initiation parameter  $v_{35}$  by:

$$w_{35} = w_{Ala} * (v_{35} / v), \text{ where } w_{Ala} = 1.7 \text{ (Rohl } et al., 1996).$$

The propagation parameter for the His cross-linked residues relates to the initiation parameter by:

$$w_{26} = w_{His} * (v_{26} / v) = w_{His} * (v_{35} / v)^{1/2}, \text{ where } w_{His} = 0.36 \text{ (Rohl et al., 1996).}$$

Since Ala1 and Ala7 are outside the crosslink, their conformational distributions are assumed to be unaffected and thus the standard  $v$  and  $w$  parameters are used for these residues. All these assumptions are supported by similar  $^3J(\text{C}\alpha\text{H}-\text{NH})$  coupling constants in an  $^1\text{H}$ -NMR experiment for the interacting and intervening residues within the cross-link.

## 2.7. AGADIR

AGADIR is a prediction algorithm based on helix-coil transition theory developed by the Serrano group. The original algorithm predicts the helical behaviour of monomeric peptides using the single-sequence approximation. The treatment of the helix-coil equilibrium differs from the ZB and LR models in a number of respects, and these have been discussed in detail by Munoz and Serrano (1997). In brief, AGADIR is similar to LR in terms of the numbering and coding of the residues in the polypeptide sequence. The basic conformational unit in both theories is the residue that is centred on the  $\text{C}\alpha$  atom. AGADIR is related to ZB in that both theories require a minimal number of four consecutive residues with helical dihedral angles to constitute a helical conformation. However, the helix conformation in AGADIR is considered to be flanked by a N-cap and C-cap rather than a coil residue in ZB. The minimal helix length in AGADIR is thus four residues in helical angles (five peptide bonds) plus the two caps.

The original model (Muñoz and Serrano, 1994) only considered empirical parameters for helix propensities, backbone hydrogen bond enthalpy, side-chain interactions and capping. The backbone hydrogen bonds attributed to conformational entropy were excluded from this original model. When tested using helicity from protein fragments and peptide models, the accuracy of the prediction was very good (correlation coefficient  $r = 0.97$ ). The model was later developed to include effects of the helix macrodipole (Muñoz and Serrano, 1995a), temperature and pH (Muñoz and Serrano, 1995b), called AGADIR1s. The latest model (AGADIR1s-2) includes terms for electrostatics, ionic strength, local motifs (the capping box, hydrophobic staple,

Schellman motif and Pro-capping motif) (Lacroix *et al.*, 1998), and N1, N2 and N3 preferences (Petukhov *et al.*, 1998)

AGADIR is accurate for peptide with less than 30 residues. However, with long peptides and helix contents higher than 50%, the prediction gives inaccuracies because the peptide tends to be fragmented. (Muñoz and Serrano, 1997) developed AGADIRms for helix prediction based on multiple sequence approximation. The prediction power of AGADIRms is, however, very similar to AGADIR1s. When compared to the LR model, AGADIR gives virtually identical results when an equivalent set of parameters is used, despite the differences on treating helix/coil cooperativity due to the difference in the residual numbering system. The authors thus suggested that the helix/coil transition is a solid theory independent of the particularities of the model for  $\alpha$ -helix formation.

The latest model, AGADIR1s-2, using the single sequence approximation (Lacroix *et al.*, 1998), calculates the free energy of a helical segment as  $\Delta G_{\text{helical-segment}} = \Delta G_{\text{int}} + \Delta G_{\text{Hbond}} + \Delta G_{\text{SD}} + \Delta G_{\text{dipole}} + \Delta G_{\text{nonH}} + \Delta G_{\text{electostat}}$ .  $\Delta G_{\text{int}}$  is the energy required to fix a residue in helical angles at internal position including N1, N2 and N3 positions and considered purely entropic.  $\Delta G_{\text{Hbond}}$  is the sum of the enthalpic contributions from the formation of main-chain - main-chain  $i, i+4$  hydrogen bonds.  $\Delta G_{\text{SD}}$  is the free energy arising from  $i, i+3$  or  $i, i+4$  side-chain interactions, excluding those between charged groups, capping and helix dipole interactions.  $\Delta G_{\text{dipole}}$  represents the interaction of charged groups with the helix macrodipole.  $\Delta G_{\text{nonH}}$  is the sum of the net contributions of the capping residues.  $\Delta G_{\text{electostat}}$  includes all electrostatic interactions between two charged residues inside and outside the helical segment and are calculated with Coulomb's equation.

AGADIR is at present the only model that can give a prediction of helix content for any peptide sequence. It can also predict NMR chemical shifts and coupling constants. However, the energetic database used for prediction is not complete. For example only few of the 400 possible  $i, i+4$  side-chain interactions have been accurately measured. Further determination of energetic contributions to helix stability is therefore still needed.

## 2.8. Helix-coil programs on the web

Some Lifson-Roig based helix-coil models are available at <http://www.bi.umist.ac.uk/users/mjfajdg/HC.htm>, namely the Caphelix, Scint, N1N2N3, 310helix, Pihelix and Mixedhelix programs. Caphelix includes the modification on N- and C-cap preferences. Scint incorporates the presence of side-chain interactions. The Caphelix and the Scint programs have been merged to incorporate N- and C-cap preferences and side-chain interactions simultaneously, into a program called Scint2. The N1N2N3 program can be used for peptides that take into account the N1, N2 or N3 preferences. The 310helix program is used to predict  $3_{10}$ -helix contents for helix interiors and capping positions without side-chain interactions. The Pihelix program is capable of predicting  $\pi$ -helix contents of peptides for helix interiors and capping positions without side-chain interactions. The Mixedhelix program is able to predict the  $\alpha$ - and  $3_{10}$ -helix contents of peptides without side-chain interactions. Unfortunately, only databases on amino acid preferences in  $\alpha$ -helix are available to date. The 310helix, Pihelix, Mixedhelix programs are, therefore, only for theoretical interest until databases of preferences for helix interiors and capping positions of other helices are known. AGADIR can be found at <http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html>. The Andersen version of a Lifson-Roig based helix-coil is at <http://faculty.washington.edu/nielshan/helix/>.

## CHAPTER 3

### MATERIALS AND METHODS

#### 3.1. Peptide design

In this study, four series of peptides were used for the experimental work, namely (1) N3 peptides, to study the effect of the N3 residues on  $\alpha$ -helix stability, (2) the CAAC peptide, to study the CXXC motif in an isolated  $\alpha$ -helix peptide, (3) RFM peptides, to study cooperativity in the  $\alpha$ -helix and (4) EKE peptides, to study anti-cooperativity in the  $\alpha$ -helix. All peptides were designed using the Scint2 program. It is a computer program based on a Lifson-Roig helix-coil theory written in Fortran. Scint2 is available from <http://www.bi.umist.ac.uk/users/mjfajdg/Scint.htm>. The implementation is based on the code originally written by Qian and Schellman (1992). Modifications to the original program have been made to incorporate capping (Doig *et al.*, 1994) and side-chain interactions (Stapley *et al.*, 1995). This program is capable of predicting helix contents of peptides that have  $i, i+3$  or  $i, i+4$  side-chain interactions within a helix, provided the side-chain interaction energies are known. The main reason why Scint2 was used for designing the peptides is that the parameter values for this program are readily available.

The program evaluates the probability that each residue has helix propagation ( $w$ ), helix initiation ( $v$ ), N-capping ( $n$ ), C-capping ( $c$ ),  $i, i+4$  interaction ( $p$ ), or  $i, i+3$  interaction ( $q$ ) weightings from inputs of a peptide sequence and tables of  $w$ ,  $v$ ,  $n$ ,  $c$ ,  $p$ , or  $q$  values. The program gives the partition function,  $Z$ , as the outputs calculated for the peptide. Properties for individual residues are also calculated. Average properties of the helix are calculated as the final output in the form of percentage of helix (see Chapter 2).

In general, three files were used to run the program, namely 'scint', 'helix.lib', and 'sample.seq'. 'scint' is the actual SGI executable program. 'helix.lib' is a library file which contains the intrinsic helix parameters  $w$ ,  $v$ ,  $n$  and  $c$  for all the amino acids (Table 3.1). 'sample.seq' is an example of a sequence file that uses the one-letter

code for the corresponding residue. The code must correspond to the code in 'helix.lib'.

In addition, the libraries 'pvalue.lib' and 'qvalue.lib' were used for sequences with side-chain interactions. 'pvalue.lib' and 'qvalue.lib' are library files containing  $i, i+4$  and  $i, i+3$  interaction parameters, respectively. Not all interaction parameter values are currently available. Unknown values were thus set to 1.0. Whenever available, the parameter in the 'pvalue.lib' and 'qvalue.lib' library files was replaced to include a particular side-chain interaction. The  $p$  or  $q$  values were calculated from previously reported  $\Delta G$  of interactions using  $\Delta G = -RT \ln p$  or  $-RT \ln q$ , as appropriate.

Before running the Scint2 program, the 'sample.seq' was inputted according to the sequence of peptide intended. The first line in the 'sample.seq' file was left as a comment. The second line was changed according to the number of residues in the tested sequence. The residues were inputted one per line as one-letter code, for example:

```

Comment
18
1      X
2      A
3      A
4      E
5      A
6      A
7      A
8      A
9      K
10     A
11     A
12     A
13     A
14     K
15     A
16     G
17     Y
18     Z

```

The characters representing the sequence were always in the sixth column following the residue number. Typing 'Scint2' and providing the requested sequence file and libraries information gave an output of percentage of stabilised helix.

Table 3.1 summarises the intrinsic helix parameters for helix-coil programs used in this study (see also section 3.7). The intrinsic helix parameters for  $w$ ,  $v^2$ ,  $n$ ,  $c$

and  $c$  of all amino acids were taken from Rohl *et al.* (1996). The  $n1$  and  $n2$  values were taken from Cochran *et al.* (2001) and Cochran and Doig (2001), respectively.

**Table 3.1.** Intrinsic helix parameters of all amino acids for helix-coil programs.

Amino acid		Value*						
		<i>w</i>	<i>v</i> <sup>2</sup>	<i>n</i>	<i>c</i>	<i>n1</i>	<i>n2</i>	<i>c1</i>
A	Ala	1.7	0.0013		1.0	0.27	1.60	0.036
B	Glu <sup>o</sup>	0.7	0.0013	1.0	1.0	0.15	2.10	0.036
	Cys <sup>o</sup>	0.32	0.0013		1.0			0.036
C	Cys <sup>-</sup>		0.0013	5.4	1.0	0.04	1.70	0.036
D	Asp <sup>-</sup>	0.38	0.0013	6.6	1.0	0.28	9.00	0.036
E	Glu <sup>-</sup>	0.54	0.0013	2.06	1.0	0.22	10.30	0.036
F	Phe	0.27	0.0013	2.06	1.0	0.02	1.80	0.036
G	Gly	0.048	0.0013	3.9	1.0	0.04	0.00	0.036
H	His <sup>+</sup>	0.22	0.0013	1.0	1.0	0.00	0.00	0.036
I	Ile	0.46	0.0013	1.57	1.0	0.011	2.10	0.036
J	His <sup>o</sup>	0.36	0.0013	2.12	1.0	0.08	1.80	0.036
K	Lys <sup>+</sup>	1.0	0.0013	0.72	1.0	0.08	0.50	0.036
L	Leu	0.87	0.0013	2.06	1.0	0.013	1.30	0.036
M	Met	0.65	0.0013	1.31	1.0	0.01	1.20	0.036
N	Asn	0.29	0.0013	6.8	1.0	0.00	0.40	0.036
O	Asp <sup>o</sup>	0.4	0.0013	1.0	1.0	0.12	1.80	0.036
P	Pro	0.001	0.0013	1.35	1.0	0.09	0.00	0.036
Q	Gln	0.62	0.0013	0.12	1.0	0.11	1.70	0.036
R	Arg <sup>+</sup>	1.14	0.0013	1.0	1.0	0.08	0.60	0.036
S	Ser	0.40	0.0013	3.9	1.0	0.13	2.10	0.036
T	Thr	0.18	0.0013	2.23	1.0	0.11	5.60	0.036
U	Lys <sup>o</sup>	0.0	0.0013	0.0	1.0	0.036	0.00	0.036
V	Val	0.25	0.0013	0.96	1.0	0.10	4.50	0.036
W	Trp	0.29	0.013	3.6	1.0	0.13	2.10	0.036
X	Acetyl		0.013	5.9	1.0	0.00	0.00	0.036
Y	Tyr	0.48	0.013	4.9	1.0	0.00	0.00	0.036
Z	Amide		0.013		1.0	0.00	0.00	0.036

\*Values are for peptides prediction in water at 273 K.

Some criteria were taken into consideration in designing the peptides, as follows:

1. The helix content of the most stabilising peptide in a series is preferably about 50%, although in certain cases this number is flexible. This is to ensure that the CD signal is sensitive to a change in free energy of  $\alpha$ -helix formation.
2. The peptide should be reasonably short for easy synthesis as well as to be cost effective. A peptide sequence of about 14-20 residues is preferable.
3. The peptide should be soluble otherwise the CD measurement is meaningless due to aggregation. This is normally achieved by placing charged residues of Lys or Gln in the sequence, although Lys will give higher helix content. The longer the sequence the more charged residues are required. Two or three Lys or Gln in a 20 residue poly-Ala peptide is normally adequate to ensure good solubility.
4. Intended side-chain interactions, where applicable, is achieved by placing the interacting residues  $i, i+4$  apart so they are in the same face of the helix. Conversely, unwanted side-chain interactions are avoided by placing the possible interacting residues normally  $i, i+5$  apart.
5. A spectrometric concentration measurement of pure peptide can be performed by placing a Tyr residue most commonly at the C-terminus end of a sequence. A Gly residue is placed immediately before Tyr to stop the helix elongation. This is to avoid perturbation of the CD signal by the Tyr residue.

There are some other criteria used which are later explained throughout the Results and Discussions chapters, based on the peptides needed.

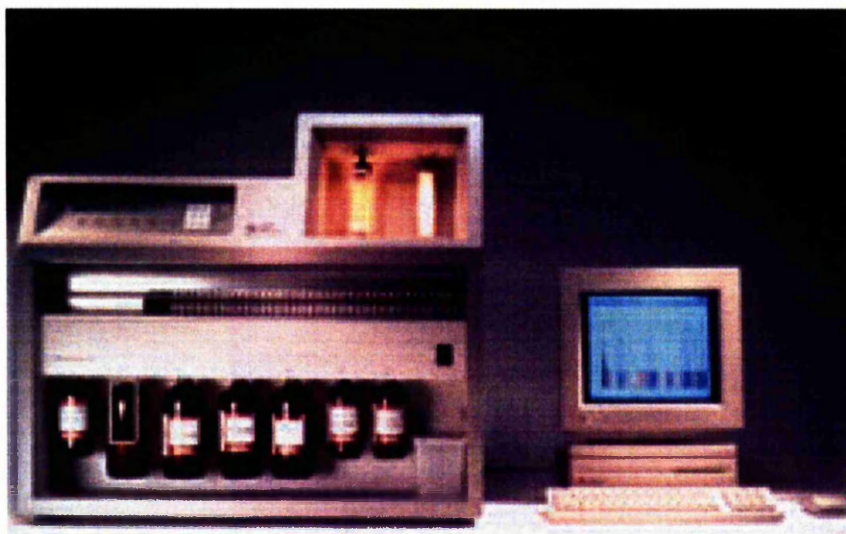


### 3.2. Peptide synthesis and purification

Peptides were made based on a solid phase peptide synthesis method that was firstly described by Merrifield (1963). It depends on (1) the attachment of the C-terminal amino acid residue to a solid resin, (2) the stepwise addition of protected amino acids to a growing peptide chain covalently bound to the resin and (3) the cleavage of the completed peptide from the solid support. The process is cyclical and repeated until the chain assembly process is completed.

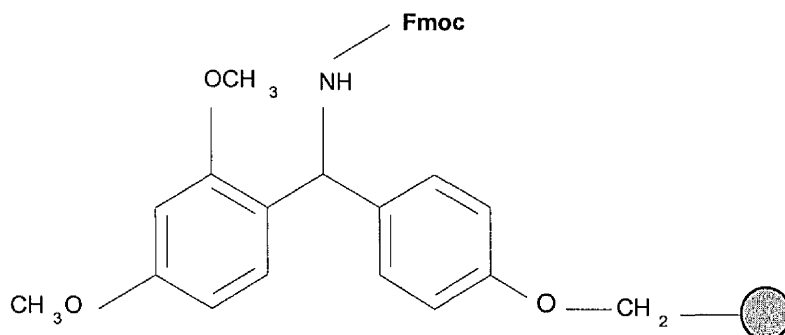
#### 3.2.1. Peptide synthesis

Automatic peptide synthesis was carried out on an Applied Biosystem 433A peptide synthesizer (Figure 3.1) on a 0.1mmol scale using 9-Fluorenylmethoxycarbonyl (Fmoc) solid phase chemistry in N-methyl-2-pyrrolidinone (NMP) solvent. Fmoc-rink amide resin 100-200 mesh (NovaBiochem®) - a modified benzyl hydrlamine resin containing ortho- and para-electron donating metoxy groups - was chosen to provide a final product with an amide group at C-terminus (Figure 3.2). This resin easily absorbs NMP to provide a large surface area for the removal of Fmoc protecting groups. For rink amide resin with a loading of 0.67mmol/g, 0.15g resin was used for 0.1mmol scale synthesis ( $=0.1/0.67$ ).



**Figure 3.1.** Applied Biosystem 433A peptide synthesizer.

The Fmoc group (Carpino and Han, 1970; 1972) is the  $\alpha$ -amino protecting group, which is widely used in solid-state peptide synthesis for its adaptability to linkers, resin and cleavage chemistries. Another example of an amino-protecting group is N- $\epsilon$ -t-butyloxycarbonyl (Boc). Boc-solid phase chemistry, however, requires more hazardous chemicals (for example HF) during the synthesis as well as in the down processing steps (White and Chan, 2000), so it is not considered here.



**Figure 3.2.** Structure of Rink amide resin, 4-(2,4'-dimethoxyphenyl-Fmoc-aminomethyl)-phenoxyresin.

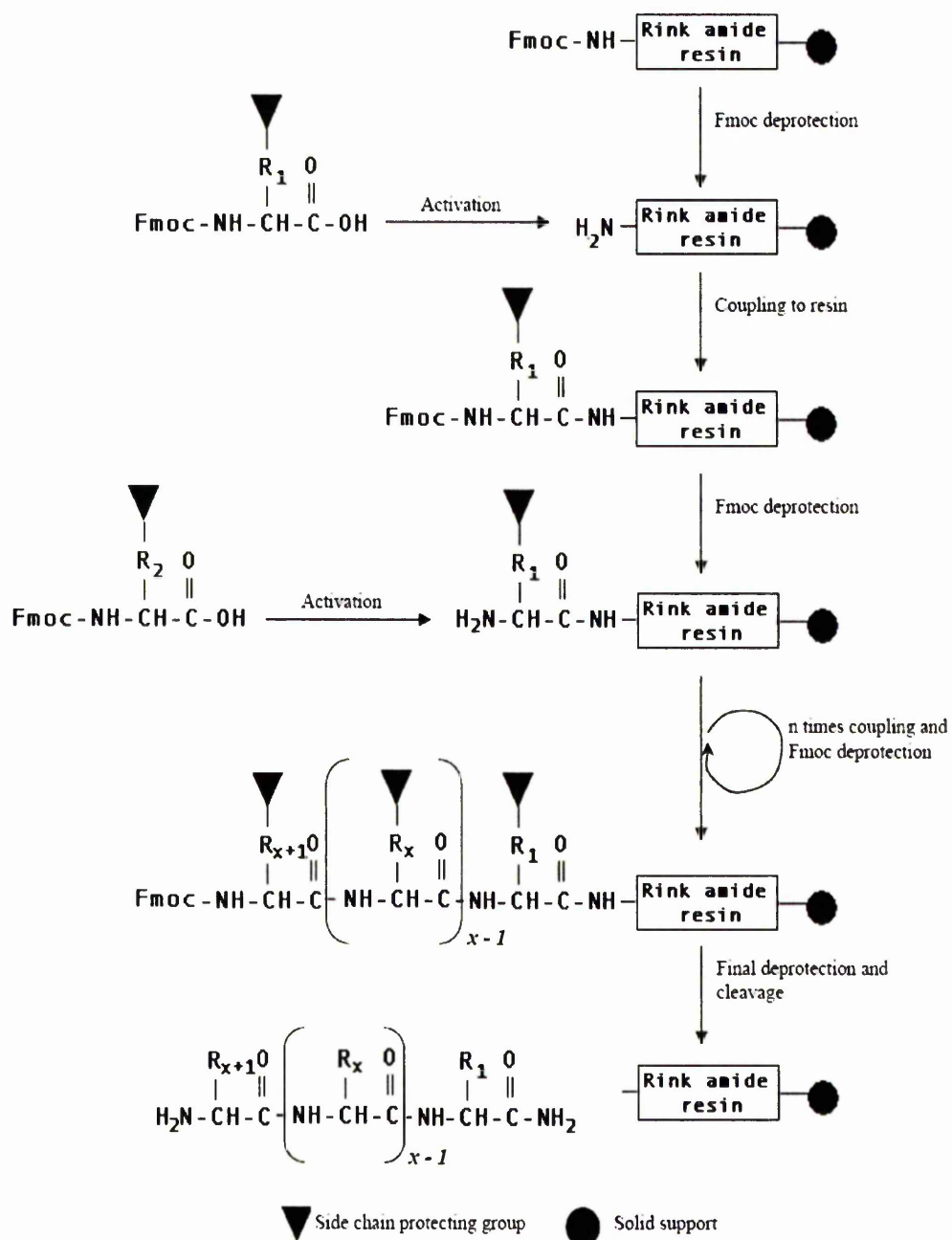
Amino acid derivatives (NovaBiochem®) used in the experiments were purchased from CN Biosciences UK (Table 3.2). All other chemicals for peptide synthesis were purchased from Applied Biosystem, unless stated otherwise.

**Table 3.2.** Fmoc-protected amino acids used for peptides synthesis.

Amino Acid	Abbreviation	Amino acid derivatives	Molecular weight (g/mol)
Alanine	Ala, A	Fmoc-Ala-OH	311.3
Arginine	Arg, R	Fmoc-Arg(Pbf)-OH	648.8
Asparagine	Asn, N	Fmoc-Asn(Trt)-OH	596.7
Aspartic acid	Asp, D	Fmoc-Asp(OtBu)-OH	411.5
Cysteine	Cys, C	Fmoc-Cys(Trt)-OH	585.7
Glutamic acid	Glu, E	Fmoc-Glu(OtBu)-OH	425.5
Glutamine	Gln, Q	Fmoc-Gln(Trt)-OH	610.7
Glycine	Gly, G	Fmoc-Gly-OH	297.3
Histidine	His, H	Fmoc-His(Trt)-OH	619.7
Isoleucine	Ile, I	Fmoc-Ile-OH	353.4
Leucine	Leu, L	Fmoc-Leu-OH	353.4
Lysine	Lys, K	Fmoc-Lys(Boc)-OH	468.5
Methionine	Met, M	Fmoc-Met-OH	371.5
Phenylalanine	Phe, F	Fmoc-Phe-OH	387.4
Proline	Pro, P	Fmoc-Pro-OH	337.4
Serine	Ser, S	Fmoc-Ser(tBu)-OH	383.4
Threonine	Thr, T	Fmoc-Thr(tBu)-OH	397.5
Tryptophan	Trp, W	Fmoc-Trp(Boc)-OH	526.6
Tyrosine	Tyr, Y	Fmoc-Tyr(tBu)-OH	459.6
Valine	Val, V	Fmoc-Val-OH	339.4

Boc = N- $\epsilon$ -t-Boc, tBu = O-t butyl, OtBu =  $\alpha$ -t-butyl ester, Pbf = N<sup>G</sup>-2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl, Trt = N- $\beta$ -trityl.

Figure 3.2 shows a summary of the peptide synthesis process where the cycle is repeated until the last residue is left with a free N-terminus. The peptide is assembled from the C-terminus towards the N-terminus with the  $\alpha$ -carboxyl group of the amino acid attached to the resin.



**Figure 3.3.** Peptide chain assembly for Fmoc solid phase peptide synthesis using Rink amide resin.

The first step in chain assembly was deprotection of the base labile  $\alpha$ -amino Fmoc group of carboxyamidated C-terminus rink amide resin with 20% piperidine (Sigma) in Dimethylformamide (DMF). Unprotected resin was then washed with NMP several times. After deprotection, the next Fmoc protected amino acids were sequentially coupled to the unprotected amino end of the growing peptide using 0.45M 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluoro phosphate (HBTU) dissolved in a solution of 1-Hydroxy-1H-benzotriazole (HOBt) in DMF and 1ml of 2M N,N-Diisopropylethylamine (DIEA). These reagents also served to activate the carboxyl group of Fmoc protected amino acid. Cyclical processes of Fmoc deprotection and coupling continued automatically in the peptide synthesiser. Finally, the peptide-resin was washed with Dichloromethane (DCM).

In the automatic Applied Biosystem 433A peptide synthesiser on a 0.1mM/0.25mM scale, the amount of reagents in table 3.3 was used per cycle.

**Table 3.3.** Chemicals used in FastMoc 0.1mM and FastMoc 0.25mM programs.

Reagents	Operations	Fastmoc 0.1mM (ml/cycle)	Fastmoc 0.25mM (ml/cycle)
20 % Piperidine in DMF	Deprotection	2.0	5.0
0.45 M HBTU/HOBt in DMF	Coupling	2.0	2.0
2 M DIEA in NMP	Amino acid activation/coupling	1.0	1.0
NMP	Wash	45	92

A sequential program (SynthAsist 2.0) to assemble a peptide was set on an Apple Macintosh platform computer connected to the synthesiser (listed in Table 3.4). The program was then sent to the synthesiser and executed to run the synthesis. The program is applicable for all peptides used in the experiment, provided the correct sequence was set.

**Table 3.4.** Pre-program Fastmoc cycles on SynthAsist 2.0 run for the synthesis of N3 peptides.

Step	Amino acid	Modules	Remarks
1	Resin	cD	A = Dissolving amino acid with NMP (2.1g) and 0.9mmol of HBTU/HOBT in DMF (2g) B = Piperidine deprotection, 20% piperidine in NMP D = NMP washes E = Addition of DIEA to cartridge and transfer to reaction vessel F = Washing the cartridge with NMP which later used to wash resin in the reaction vessel I = Wait c = DCM wash
2	Y	BADEF	
3	G	BADEF	
4	A	BADEF	
5	K	BADEF	
6	A	BADEF	
7	A	BADEF	
8	A	BADEF	
9	A	BADEF	
10	K	BADEF	
11	A	BADEF	
12	A	BADEF	
13	A	BADEF	
14	A	BADEF	
15	X	BADEF	
16	A	BADEF	
17	A	BADEF	
18	Final deprotection	BIDc	

### 3.2.2. Acetylation and cleavage

Having completed the chain assembly, the resin containing the peptide was placed in a sintered glass funnel and washed three times, each with about 10ml of DMF and DCM and further washed with methanol. Some suction using a water-vacuum pump was applied to the funnel to dry the peptide resin. Next, the free N-terminus of the peptide was acetylated by adding an acetylation solution consisting of 2.3ml DMF, 0.1ml acetic anhydride (Fluka) and 0.12ml pyridine (Sigma) to the peptide resin and left for 30 minutes. This process was, however, skipped in the synthesis of the CAAC peptide (see Chapter 5). The acetylation solution was then removed from the peptide resin. The resin was washed three times, each with about 10ml of DMF and DCM which was later dried by applying some suction before further treatment.

Cleavage of the peptide from the resin and removal of amino acids containing side-chain protecting groups was accomplished by suspending the peptide resin in 5ml cleavage solution containing 95% trifluoroacetic acid (TFA-Sigma), 2.5% H<sub>2</sub>O and 2.5% triisopropylxylan (TIS-Sigma). The latter served as scavengers, which react with side-chain protecting groups of amino acids released during TFA cleavage preventing unwanted side reactions.

After a minimum of three hours of cleavage, the peptide resin was placed into a sintered glass funnel and some suction was applied. The resin was then washed twice with 10ml of DCM. The volume of the filtrate containing the desired peptide was reduced to about 0.5-1.0ml by blowing compressed air onto the filtrate.

The concentrated filtrate was then dropped into about 30ml of dry ice-cold diethyl ether (Sigma). The solution was stored overnight at -20° C for the precipitate to develop. The resulting precipitate was then centrifuged at 10,000 rpm for 10 minutes at 4° C in a Dupont instrument Sorvall RC-5B Refrigerated Centrifuge using SS-34 rotor. The pellet was then further washed with 30ml of fresh dry ice-cold diethyl ether and further centrifuged at 10,000 rpm for 10 minutes. The pellet obtained was dried by placing the centrifuge tube in warm water for the ether to evaporate. Dry pellet containing crude peptide was dissolved in 5ml of water, vortexed and lyophilised in a freeze drier overnight. Crude peptide was dissolved in 1-2 ml of H<sub>2</sub>O before being purified by HPLC.

### 3.2.3. Peptide purification

Peptides were purified by High Performance Liquid Chromatography using a Hewlett Packard Series 1100 with analytical C18 reverse phase column (Phenomenex). Peptides were eluted by gradient elution using a solvent mixture of acetonitrile (Sigma) and H<sub>2</sub>O in the presence of 0.1% TFA. Table 3.5 describes the gradient elution program mainly used in the experiment.

Before performing the elution, the column was washed thoroughly with 95% acetonitrile for at least 20 minutes to clean dirtiness from previous runs. It was then equilibrated with 5% acetonitrile for 10 minutes before the purification run. Fractions

of the desired peptide were pooled, freeze-dried and lyophilised overnight in a freeze-dryer. The dry peptide was finally dissolved in 1 ml of H<sub>2</sub>O and stored at -20°C.

**Table 3.5.** HPLC gradient elution program mainly used for peptide purification at a flow rate of 3.5 ml/minute.

Time (minutes)	% Acetonitrile
0	5%
5	5%
50	60%

### 3.3. Verification of peptide mass

To ensure that the correct peptide was synthesized, the peptide mass was verified by either Electrospray Ionization (ESI) or Matrix Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) mass spectrometry at the Michael Barber Centre for Mass Spectrometry at UMIST.

For small molecules (< 2000 Daltons) ESI typically generates singly or doubly charged gaseous ions, while for large molecules (> 2000 Daltons) the ESI process typically gives rise to a series of multiply-charged species. Because mass spectrometers measure the mass-to-charge ( $m/z$ ) ratio, the resultant ESI mass spectrum contains multiple peaks corresponding to the different charged states. MALDI is considered more sensitive than ESI as it generates singly charged gaseous ions and is better used for peptides with higher mass. The  $[M+H]^+$ ,  $[M+2H]^+$  and  $[M+3H]^+$  for all peptides used in the experiment are listed in Table 3.6.

To clarify the mass spectra interpretation, the following example is given. By assuming a N3 peptide with Ala substitution has a molecular weight of 1388.8, fragmented  $[M+2H]^+$  and  $[M+3H]^+$  ions have theoretical  $m/z$  of 695.39 and 463.93, respectively. The following calculation is an example given for the  $[M+2H]^+$  for the N3 peptide with substituting Ala at N3 position.

$$[M+2H]^+ = \frac{(M + 2H)}{2} = \frac{(1388.8 + 2)}{2} = 695.4$$



The  $[M+2H]^+$  ion is present in the trace represented by a  $m/z$  of 694.7 confirming that the correct identity of the peptide (Appendix A.1). Mass spectrometry results for all other peptides can also be found in Appendix A.

**Table 3.6.** The mass to charge ratio ( $m/z$ ) of fragmented  $[M+H]^+$ ,  $[M+2H]^+$  and  $[M+3H]^+$  ions of the peptides used in the experiments.

Peptide <sup>a</sup>	Average mass (g/mol)	$m/z$		
		$[M+H]^+$	$[M+2H]^+$	$[M+3H]^+$
N3(Ala)	1388.8	1389.8	695.39	463.93
N3(Arg)	1473.9	1474.9	737.93	492.29
N3(Asn)	1431.8	1432.8	716.89	478.26
N3(Asp)	1432.8	1433.8	717.38	478.59
N3(Cys)	1420.8	1421.8	711.41	474.61
N3(Glu)	1446.8	1447.8	724.40	483.26
N3(Gln)	1445.8	1446.8	723.91	482.94
N3(Gly)	1374.7	1375.7	688.37	459.24
N3(His)	1454.8	1455.8	728.41	485.94
N3(Ile)	1430.9	1431.9	716.44	477.96
N3(Leu)	1430.9	1431.9	716.44	477.96
N3(Lys)	1445.9	1446.9	723.93	482.95
N3(Met)	1448.9	1449.9	725.44	483.96
N3(Phe)	1464.9	1465.9	733.43	489.29
N3(Pro)	1414.8	1415.8	708.40	472.60
N3(Ser)	1404.8	1405.8	703.38	469.25
N3(Thr)	1418.8	1419.8	710.39	473.93
N3(Trp)	1503.9	1504.9	752.95	502.30
N3(Tyr)	1480.9	1481.9	741.43	494.62
N3(Val)	1416.8	1417.8	709.41	473.27
CAAC <sup>b</sup>	1410.7	1411.7	706.35	471.24
RFM <sup>b</sup>	1681.0	1682.0	841.50	561.34
EKE <sup>b</sup>	1646.8	1647.8	824.42	549.95

<sup>a</sup> The full sequence of the peptides can be seen in the corresponding Results and Discussions chapters.

<sup>b</sup> Also represents isomers of peptide with different side-chain spacings.

### 3.4. Peptide concentration determination

Peptide concentrations were determined by measuring the tyrosine UV absorbance at 275 nm of diluted aliquots of stock solution in water or in 6M guanidine hydrochloride on a UVIKON 922 Spectrophotometer. Tyrosine has a characteristic to optimally absorb UV light at a wavelength range of 275 - 280 nm as the result of the conjugated  $\pi$  electrons in its aromatic ring. The absorbance value can be transformed into Lambert-Beer Law equation to calculate the concentration of the peptide.

$$A = \epsilon.b.C \quad (3.1)$$

Where  $A$  = absorbance,  $\epsilon$  = Molar extinction coefficient of Tyr ( $=1390 \text{ M}^{-1}\text{cm}^{-1}$  and  $1450 \text{ M}^{-1}\text{cm}^{-1}$  in water and 6M guanidine-HCl, respectively, at 275 nm),  $C$  = concentration (Molar),  $b$  = cell path length (cm).

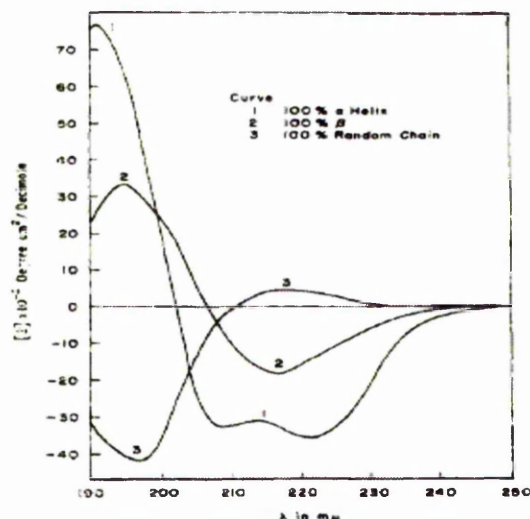
### 3.5. Circular Dichroism to determine helix content

Circular dichroism (CD) spectroscopy has been extensively applied to the structural characterization of peptides, including the estimation of secondary structural content. CD is displayed when an optically active substance absorbs left or right handed circularly polarized light preferentially.

Asymmetric molecules, such as L- and D- amino acids and right and left hand helices, have a preference for absorbing either left or right circularly polarised light. A left and right circularly polarised beam interacts differently with a right-handed  $\alpha$ -helix. The difference in absorption is called circular dichroism and is defined as  $\Delta A = \frac{A_L - A_R}{A}$ , where  $A_L$  is the absorbance for left circularly polarised light,  $A_R$  is the absorbance for right polarised light and  $A$  is the absorbance for unpolarised light.

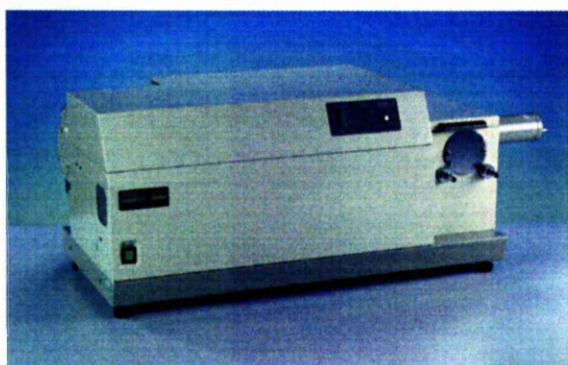
Peptides with different secondary structure conformations give different CD spectra as shown in Figure 3.4. The helical peptides used in the experiment have at least one of two types of chromophores in their structure, namely the peptide bond and the amino acid side-chains. The lowest energy transition in the peptide

chromophore is an  $n \rightarrow \pi^*$  transition observed in the  $\alpha$ -helical form of the peptide as a small negative shoulder near 222 nm on the tail of a much stronger absorption band centered at 192 nm. This intense band, responsible for the majority of the peptide bond absorbance, is a  $\pi \rightarrow \pi^*$  transition. Exciton splitting of the  $\pi \rightarrow \pi^*$  transition also results in the negative band at 208 nm.



**Figure 3.4.** Circular dichroism spectra of "pure" secondary structures of poly-L-lysine. Reproduced from (Greenfield and Fasman, 1969).

Figure 3.5 shows the Jasco J810 spectropolarimeter used to acquire the CD ellipticity data. The peptides were normally measured in 5mM sodium phosphate buffer in the presence of 10mM NaCl at 273K, unless stated otherwise. NaCl at low concentration was added for charge screening. The light source (xenon lamp) was flushed with clean dry nitrogen at 3.5ml/minute prior and during measurement to avoid the generation of ozone in the instrument.



**Figure 3.5.** Jasco J810 spectropolarimeter.

The following parameters were used for the CD measurement:

Scan wavelength: 190 - 250 nm (interval scan measurement)

222 nm (time course measurement)

Sensitivity : 100 mdeg

Bandwidth : 2 nm

Scan speed : 50nm/minute

Response time : 1 second

Data pitch : 0.2 nm

Accumulation : 4

Ellipticity is the unit of CD measurements and can be converted to mean residual ellipticity using:

$$[\theta]_{222} = \frac{\theta_{obs}}{Concentration * N * \ell} \text{ (deg.cm}^2\text{.dmol}^{-1}\text{)} \quad (3.2)$$

Where  $\theta_{obs}$  is the experimentally observed ellipticity,  $N$  is the number of residues and  $\ell$  is the cell pathlength.

The mean helix content was then calculated as:

$$\%helix = \frac{[\theta]_{222}}{[\theta]_{222(max)}} \quad (3.3)$$

$$\text{Where } [\theta]_{222(max)} = \theta_H * \left(1 - \frac{2.5}{N}\right) \quad (3.4)$$

$\theta_H$  is the ellipticity of a complete helix of infinite length at 0°C given by a number of -40,000 deg.cm<sup>2</sup>.dmol<sup>-1</sup>. The term  $1 - \frac{2.5}{N}$  is a correction to  $\theta_H$  for end

effects and  $N$  is the number amino acids in the peptide (Chakrabartty *et al.*, 1991).

### 3.6. pH titration

pKa values were evaluated for peptides containing the ionisable side-chains. These are Asp, Glu, Cys and His in the N3 series, Cys in the CAAC peptides and Glu in EKE peptides. The helicities were titrated as a function of pH between approximately pH 2 and pH 10. Solutions of peptide at 20  $\mu$ M in a 3mL, 1.0cm quartz cell were equilibrated at 273K in 10mM NaCl, 1mM sodium phosphate, 1mM sodium borate, and 1mM sodium citrate. The CD signals at 222 nm were monitored after pH adjustments that were made by the addition of either NaOH or HCl. The peptide concentrations were finally corrected by taking account the change in the total volume.

Depending on the number of the apparent titrable groups, the ellipticity data from the titrations were fitted to a Henderson-Hasselbach equation using the software program KaleidaGraph 3.6 to determine the pKa. For one apparent pKa,  $[\theta]_{222}$  as a function of pH is given by:

$$[\theta]_{222} = [\theta]_{222, \text{high\_pH}} * \left(1 - \frac{1}{1 + 10^{pH - pKa}}\right) + [\theta]_{222, \text{low\_pH}} * \left(\frac{1}{1 + 10^{pH - pKa}}\right) \quad (3.5)$$

where  $[\theta]_{222, \text{high\_pH}}$  and  $[\theta]_{222, \text{low\_pH}}$  are the molar ellipticities measured at 222 nm at the titration end points at high and low pH.

The equation for two apparent pKas is given by:

$$[\theta]_{222} = [\theta]_{222, \text{mid\_pH}} * \left(1 - \frac{1}{1 + 10^{pH - pKa_1}}\right) + [\theta]_{222, \text{low\_pH}} * \left(\frac{1}{1 + 10^{pH - pKa_1}}\right) + [\theta]_{222, (\text{mid\_pH} - \text{high\_pH})} * \left(1 - \frac{1}{1 + 10^{pH - pKa_2}}\right) \quad (3.6)$$

where pKa1 and pKa2 are the pKas measured for the acid- base equilibrium at low and high pH, respectively.  $[\theta]_{222, \text{high\_pH}}$ ,  $[\theta]_{222, \text{mid\_pH}}$  and  $[\theta]_{222, \text{low\_pH}}$  are the molar ellipticities measured at 222 nm at the titration end points at high, mid, and low pH.  $[\theta]_{222, \text{mid\_pH} - \text{high\_pH}}$  is the change in molar ellipticity associated with pKa2.

The equation for three apparent pKas is given by:

$$\begin{aligned}
 [\theta]_{222} = & [\theta]_{222, \text{mid}^1_{\text{pH}}} * \left( 1 - \frac{1}{1 + 10^{pH - pK_{a1}}} \right) + [\theta]_{222, \text{low}_{\text{pH}}} * \left( \frac{1}{1 + 10^{pH - pK_{a1}}} \right) \\
 & + [\theta]_{222, \text{mid}^2_{\text{pH}}} * \left( 1 - \frac{1}{1 + 10^{pH - pK_{a2}}} \right) \\
 & + [\theta]_{222, (\text{mid}^2_{\text{pH}} - \text{high}_{\text{pH}})} * \left( 1 - \frac{1}{1 + 10^{pH - pK_{a3}}} \right) \quad (3.7)
 \end{aligned}$$

where, pKa1, pKa2, pKa3 are the pKas measured for the acid- base equilibrium at low, mid and high pH, respectively.  $[\theta]_{222, \text{high}_{\text{pH}}}$ ,  $[\theta]_{222, \text{mid}^1_{\text{pH}}}$ ,  $[\theta]_{222, \text{mid}^2_{\text{pH}}}$  and  $[\theta]_{222, \text{low}_{\text{pH}}}$  are the molar ellipticities measured at 222 nm at the titration end points at high, mid<sup>1</sup>, mid<sup>2</sup> and low pH.  $[\theta]_{222, \text{mid}^2_{\text{pH}} - \text{high}_{\text{pH}}}$  is the change in molar ellipticity associated with pKa3.

### 3.7. DTT titration of CAAC peptide

The redox potential of the CAAC peptide (Chapter 5) was determined by titrating a 20μM peptide solution with DTT of known concentrations. The final concentration of DTT in the solution was calculated after correcting for the volume of the solution. Peptide was dissolved in CD buffer containing 5mM sodium phosphate and 10mM NaCl pH 8.1. Prior to use, the buffer was thoroughly degassed and flushed with nitrogen. Peptide solution with added DTT was allowed to equilibrate for 15 minutes at room temperature before taking CD measurement.

The equilibrium constant,  $K_{eq}$ , of thiol-disulphide pair in the equilibrium system was calculated according to the following equation (Inaba and Ito, 2002):

$$f = \frac{[CAAC^{SH}]^2 / [CAAC^{SS}]}{K_{eq} + [CAAC^{SH}]^2 / [CAAC^{SS}]} \quad (3.8)$$

Where  $f$  is the fraction of the reduced CAAC peptide at equilibrium, which is equal to the fractional change of  $\theta_{222}$  measured at every point after DTT addition.  $CAAC^{SH}$  and  $CAAC^{SS}$  are peptide in partially reduced and oxidised forms,

respectively. The standard redox potential was calculated from the Nernst equation, using a value of -0.366V for the DTT at pH 8.1 at room temperature (Cleland, 1964).

$$E'_{oCAAC} = E'_{oCysteine/Cystine} - (RT/2F) \ln K_{eq} \quad (3.9)$$

Where  $F = 9.65 \times 10^4 \text{ C.mol}^{-1}$  and  $R = 1.987 \text{ cal.K}^{-1}.\text{mol}^{-1}$  are the Faraday and the gas constants, respectively.

### 3.8. Zinc titration of CAAC peptide

Zn titration of 20 $\mu$ M CAAC peptide in 5mM Phosphate buffer containing 10mM NaCl and 400mM DTT at pH 8.1 was followed with CD measurement. Aliquots of fresh ZnCl<sub>2</sub> solution were added into the peptide solution. The solution was allowed to equilibrate for 15 minutes at room temperature before measurement. The equation below was used to calculate the binding constant and the number of binding sites. The fractional saturation,  $Y$ , is defined as the fraction of peptide molecules that are saturated with Zn.

$$Y = \frac{B_{max}[Zn]}{K_d + [Zn]} \quad (3.10)$$

$Y$  is equivalent to the fractional change of  $\theta_{222}$  measured at every point after Zn addition.  $B_{max}$  and  $K_d$  are the maximum number of binding sites and the dissociation constant, respectively.

### 3.9. Analysis using helix-coil theory

Energetic analysis for the N3 residue substitution used the N1N2N3 program, available from <http://www.bi.umist.ac.uk/users/mjfajdg/n1n2n3.htm>. This program implements a Lifson-Roig based helix-coil model using the N1N2N3 helix coil algorithm (Sun *et al.*, 2000) which includes preferences for N-cap, N1, N2, N3, C-cap, C1 and helix interior (see Chapter 2). Three files were used, namely 'n1n2n3prog', 'helix\_H2O.lib' and 'short.seq'. The 'n1n2n3prog' is the actual SGI executable program. 'helix\_H2O.lib' is a library file which contains the helix parameters for all the amino acids in water. 'short.seq' is an example of a sequence file.

Before running the N1N2N3 program, the 'sample.seq' was changed according to the peptide sequence intended. The first line in the 'sample.seq' was read as a comment. The second line was the peptide sequence in single letter amino acid code. The letters were all in capitals and terminated with a '1'. The one-letter codes corresponded to the one-letter codes used in the 'helix\_H2O.lib' file, for example:

```
Sequence of N3(Glu) peptide
XAAEAAAAKAAAKAGY1
```

Typing 'n1n2n3prog' and providing the requested sequence file and libraries information gave an output file that was written to a file 'hcontent.rep'. This file was run to create an output file. The output file contained information about the input, the total partition function,  $Z$ , and the information on a per residue basis. Finally average properties of the helix were calculated as percentage of helix.

The  $n3$  values were calculated from helicity measurements using parameter values  $w$ ,  $n$ ,  $n1$ ,  $n2$ ,  $c$  and  $c1$  as listed in Table 3.1. In the N3 peptide sequences the only unknown parameter is the N3 preference ( $n3$ ). Initially,  $n3$  for the four sequences with an X residue matching residues of the generic peptide (Ala, Gly, Lys, and Tyr) were solved simultaneously by fitting the equilibrium equation to the observed helicities using the N1N2N3 program. Then,  $n3$  for other residues were found by varying them until the calculated helix content agreed with experiment.

The statistical weight for the N3 coil to helix transition is  $n3.w$ , the product of  $w$ , the intrinsic interior preference value, and  $n3$ , the intrinsic N3 preference value (Sun *et al.*, 2000). Free energies for transfer from the coil to N3 were thus calculated as  $\Delta G = -RT \ln(n3.w)$  and the free energy change relative to Ala is given by  $\Delta\Delta G = \Delta G_{Xaa} - \Delta G_{Ala}$ .

Energetic analysis of the CAAC peptide (Chapter 5) and the peptides to study pairwise coupling (Chapters 6 and 7) were conducted using the Scint2 program, as it was the side-chain interactions that were to be analysed. Details about this program have been previously described in sub-chapter 3.1. The side-chain interaction parameter values were calculated from helicity measurements. Parameters in appropriate 'pvalue.lib' and 'qvalue.lib' were varied until they agreed with the experimental results on essentially unchanging values. The program calculates the



partition function,  $Z$ , for the peptide followed by properties calculated for individual residues. Finally, average properties of the helix are calculated in the form of percentage helix.

### 3.10. Protein crystals search

A survey was conducted on a dataset obtained from protein crystal structures in the ASTRAL database (SCOP 1.63 Sequence Resources) (Brenner *et al.*, 2000), which is based on domains rather than the entire structures. Each domain in the data set represents one superfamily with the best SPACI score. The SPACI score is a measure of the reliability and precision of a crystallographically determined structure in a PDB file. The data set contained 1135 chains with a total of 195875 amino acid residues and there was less than 20% sequence identity between any pair. These were processed to give secondary structure information using the DSSP algorithm as implemented in the program SSTRUC (Smith, 1989). The complete list of the protein data set can be found in Appendix B.

#### 3.10.1. Probing N3 rotamer distribution and energy calculation

The motif  $xhhHhhh$  was used to select the N termini region (N-cap-N1-N2-N3-N4/C3-N5/C2-N6/C1).  $h$  is the residue in  $\alpha$ -helical conformation and  $x$  is not. The N-cap ( $x$ ) is defined as the first residue without  $\phi, \psi$  helical dihedral angles immediately precedes a helical stretch, hydrogen-bonded to a residue in the helical sequence. The fourth residue ( $H$ ) in the motif is the N3. Firstly, the data set was probed for the helices of which there were 4778 found. A script written using the Perl program was used for this purpose and can be seen in Appendix C. The distribution of amino acid side-chain angles  $\chi_1$  and  $\chi_2$  were determined for this position to calculate the rotamer energy. Three amino acids were excluded, namely Ala and Gly and Pro since they have no  $\chi$  angles. Definitions for  $\chi_1$  and  $\chi_2$  rotamer angles for  $sp^3$ - $sp^3$  bonds are as follows:

$$\text{Gauche}^+ (g^+) \quad \chi_1/\chi_2 = -60 (\pm 60)^\circ$$

$$\text{Trans} (t) \quad \chi_1/\chi_2 = 180 (\pm 60)^\circ$$

$$\text{Gauche}^- (g^-) \quad \chi_1/\chi_2 = 60 (\pm 60)^\circ$$

In the case of Val,  $g^+$  is defined as  $\chi_1 = 180 (\pm 60)^\circ$ ,  $t$  as  $\chi_1 = 60 (\pm 60)^\circ$  and  $g^-$  -  $60 (\pm 60)^\circ$ . Similar definitions as in  $\chi_1$  are used for the  $\chi_2$  angle of Arg, Gln, Glu, Ile, Leu, Lys and Met since they also have  $sp^3$ - $sp^3$  bonds. The  $\chi_2$  angle of Ser, Thr and Cys is usually not defined in X-ray crystal structures because they involve the position of a hydrogen atom. The  $\chi_2$  angle of Asn, Asp, His, Phe, Trp and Tyr is defined by a  $sp^3$ - $sp^2$  bond. The rotamer distribution is not discrete, so no rotamer definition is good enough to allow reliable distribution calculations for these amino acids. The amount of data used for rotamer energies calculation was smaller than expected because the side-chain angles were sometimes not defined in some structures, particularly for the  $\chi_1$   $\chi_2$  rotamer combinations.

The fractional population of amino acids adopting rotamer  $i$  and errors were calculated according to the following equations (Penel *et al.*, 1999):

$$p_i = n_i / N \quad (3.11)$$

where  $n_i$  is the number of amino acids adopting  $i$  rotamer in  $N$  observations. In this case,  $N$  is the total number of amino acids observed at N3 position.

$$\sigma p_i = (p_i(1 - p_i) / N)^{1/2} \quad (3.12)$$

where  $\sigma p_i$  is the standard deviation of  $p_i$ .

The rotamer energies were then calculated from the rotamer populations (see below). Details of the calculation have been previously described by Stapley and Doig (1997). In brief, the calculation of coil energies used rotamer populations of 'All residues' minus those of 'Helix'. Observed  $\chi_1$  rotamer populations in the coil state can be converted to the relative energies of each conformation by assuming that the populations are in a Boltzmann distribution. The free energy change for restricting the side-chain to rotamer  $i$  from equilibrium state in coil where all rotamer are populated is given by:

$$E_i = -RT \ln p_i \quad (3.13)$$

where  $p_i$  is the fraction population of rotamer  $i$ .

When the residue of interest is transferred from coil to the N3 position, the free energy for restricting the side-chain to rotamer  $i$  from equilibrium state in helix is given by:

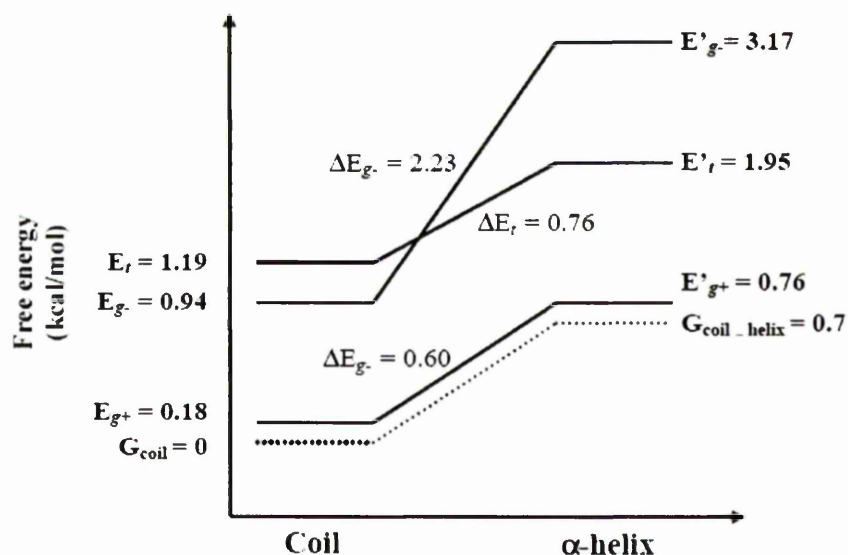
$$E'_i = -RT \ln p'_i + \Delta G_{coil \rightarrow N3} \quad (3.14)$$

where the prime indicates a rotamer in the folded N3 position and  $\Delta G_{coil \rightarrow N3}$  is the free energy for transferring a residue from coil to helix. The physical meaning of  $\Delta G_{coil \rightarrow N3}$  is the energy required for transferring residue from coil to helix. The values are thus added to the rotamer energy of each rotamer to account for the backbone conformational changes.

The free energy of the coil state, where all three rotamers are populated, is the reference state. The difference in rotamer  $i$  energy of a residue in helix and coil is thus given by:

$$\Delta E = E'_i - E_i \quad (3.15)$$

Figures 3.6 clarifies how the method is applied to the rotamer energies of amino acid  $X$  in the coil and helix. Let's assume that the rotamer fractional populations for  $X$  in the coil ( $p_i$ ) are 0.71 for  $g^+$ , 0.11 for  $t$  and 0.18 for  $g^-$ , giving rotamer energies in the coil ( $E_i$ ) of 0.18kcal/mol, 1.19kcal/mol and 0.94kcal/mol, respectively, by applying equation 3.13. In the helix, the rotamer populations ( $p'_i$ ) are assumed as 0.89, 0.10 and 0.01 for  $g^+$ ,  $t$  and  $g^-$ , respectively, giving relative energies within the helix of 0.06kcal/mol, 1.25kcal/mol and 2.47kcal/mol. The free energy change for transfer of amino acid  $X$  from the coil to helix is assumed as 0.7kcal/mol, so this is added to each rotamer energy in the helix by applying equation 3.14, giving free energies 0.76kcal/mol, 1.95kcal/mol and 3.17kcal/mol for  $g^+$ ,  $t$  and  $g^-$ , respectively. All rotamer energies increase when transferred from coil to helix. The  $g^+$ ,  $t$  and  $g^-$  free energies ( $\Delta E$ ) are therefore 0.6kcal/mol, 0.76kcal/mol and 2.23kcal/mol, respectively, using equation 3.15.



**Figure 3.6.** Illustration of rotamer energies in coil and helix for amino acid X at N3. Continuous lines are rotamer energies, while the broken lines show the free energy of transfer residue X from coil to helix. The value of  $\Delta G_{\text{coil-helix}}$  is already included in the  $E'_i$  term. The energy scale is only for illustrative purposes.

As the free energies for transferring an amino acid from coil state to N3 was empirically measured in this study (Chapter 4), the free energies of each rotamer of amino acids at N3 position were then calculated using equations 3.13, 3.14 and 3.15. Throughout this work, a temperature of 273K was used for the calculations, as the experimental measurements were made at this temperature.

### 3.10.2. Probing the CXXC motif

The CXXC motif was probed using similar methods as described above. The frequencies of all 400 possible amino acid pairs at N-cap and N3 in the 4778 helices were obtained as the output, along with the frequencies of each amino acid at each position of N-cap and N3. The pair containing Cys residues at N-cap and N3 was exclusively examined. The fractional population of amino acids and errors were calculated using equations 3.11 and 3.12. In this case, the N observations term was represented by the total number of helices in the dataset (4778). Local propensity of the pairs was calculated according to:

$$P = \frac{(O/N)}{p_i \cdot p_j} \quad (3.16)$$

where  $O$  is the number of pair occurrences in  $N$  observations,  $p_i$  and  $p_j$  are fractional population of an amino acid at N-cap and N3, respectively.

$$\sigma P = P \left( \left( \sigma p_i / p_i \right) + \left( \sigma p_j / p_j \right) \right) \quad (3.17)$$

where  $\sigma P$  is the error in propensity.

### 3.10.3. Probing the RFM motif

Firstly, the data set was probed for the RF and FM pairs spaced  $i, i+4$  apart (5 consecutive residues). The pairs were categorised as pairs in helix when they are in *hhhhh* conformation. Otherwise, they were categorised as pairs in coil. Triplets of Arg-Phe-Met were searched for on at least nine consecutive helical residues for the observation of  $i, i+4, i+8$  interactions. The rotamer  $\chi_1$  preference for Arg, Phe and Met residues were then determined. Definitions of  $\chi_1$  rotamer angle are given in section 3.10.1. The fractional population and errors were further calculated according to equations 3.11 and 3.12.

### 3.10.4. Probing the EKE motif

The EK and KE  $i, i+4$  pairs were searched from the dataset similarly as described for the RFM motif above. In addition to the  $\chi_1$  rotamer, the rotamer  $\chi_2$  preference for Glu and Lys residues were also searched. Definitions for both  $\chi_1$  and  $\chi_2$  rotamer angles for  $sp^3$ - $sp^3$  bond are similar to above. The fractional population and errors were then calculated. Examples of the Perl scripts used to search the EKE motifs can be seen in Appendix D.

## CHAPTER 4

### EFFECT OF THE N3 RESIDUE ON THE STABILITY OF THE $\alpha$ -HELIX

#### 4.1. Introduction

N1, N2 and N3 are the first three amino acids with helical  $\phi$ ,  $\psi$  backbone dihedral angles at the helix N-terminus. They differ from interior positions as their amide NH groups do not participate in backbone-backbone  $i, i+4$  hydrogen bonds within the helix. The presence of these otherwise unsatisfied hydrogen bond donors has profound structural effects, most often satisfied by hydrogen bonds to side-chains local in sequence, such as to preceding N-cap side-chains.

In native proteins, Ala, Asp, Gln and Glu have the highest propensity for the N3 position. A thorough study of the structures adopted by side-chains at the N-cap, N1, N2 and N3 positions in proteins has been made (Doig *et al.*, 1997; Penel *et al.*, 1999). Asn, Asp, Gln, Glu and Thr occasionally form  $i, i$  hydrogen bonds to the N3 backbone NH group, though these bonds may be weak as they are non-linear, particularly for the shorter side-chains (Penel *et al.*, 1999; Wan and Milner-White, 1999). A second important feature of the N3 position is the capping box, where the N3 side-chain accepts a hydrogen bond from the N-cap backbone NH group (Harper and Rose, 1993). This is most important for Gln or Glu at N3. The unique structural trends for N3 imply that the free energies for substituting amino acids at this position differ from all other positions in the helix. In this chapter, these free energies were measured using isolated helical peptides. Helix termini are highly solvent exposed and tertiary interactions are rare, so the environment of peptide and protein helices is very similar.

A study of the energetics of the N1 and N2 helix position for all 20 amino acid residues have been made (Cochran and Doig, 2001; Cochran *et al.*, 2001). Previous studies have measured position dependent effects on helix content of several amino acid residues, for example Asp (Huyghues-Despointes *et al.*, 1993a), Glu (Scholtz *et al.*, 1993), His (Armstrong and Baldwin, 1993), Val, Leu, Met, Ile, Gly, Ser, Thr,

Asn, Gln, (Petukhov *et al.*, 1998; Petukhov *et al.*, 1999). A complete survey of all 20 residues in the same peptide is needed, however.

$\alpha$ -Helices adopt a large number of structures in aqueous solution, with fully helical, fully coil and partly helical conformations all populated. For a complete understanding of helix formation and stability, all the factors contributing to this equilibrium need to be assessed thoroughly. These include the helix-forming tendencies of constituent amino acids, capping preferences at the carboxyl and amino termini and preferences for the N1, N2 and N3 positions. The helix-coil model N1N2N3 of (Sun *et al.*, 2000) distinguishes between each of these by assigning them unique weights. Factors such as side-chain-main chain hydrogen bonds, electrostatic interactions with the helix dipole, solvent exposure and conformational entropy are implicitly included in the positional preferences. Side-chain interactions between residues spaced  $i, i+3$  and  $i, i+4$  are also of considerable importance to helix stability, though are not included in the N1N2N3 helix/coil model or the peptides studied here.

The N-termini of naturally occurring helices in proteins are very often solvent exposed (Doig *et al.*, 1997). Thus the peptide is a good model for the study of factors which stabilise the  $\alpha$ -helix in naturally occurring proteins. The energetic contribution of the N3 residue to helix stability in the controlled environment of a synthetic peptide is of fundamental interest with practical implications in the development of secondary structure prediction algorithms, peptide and protein design and development of site directed mutagenesis strategies.

#### 4.2. Design of N3 peptide

N3 preferences can be determined by substituting the third residue in an acetylated peptide and measuring the change in helix content. In this study an intrinsically helical peptide was used to measure the energetic contribution (the  $n3$  parameter) of 19 amino acids in the N3 position in the same way that the  $n1$  and  $n2$  parameters have been investigated (Cochran and Doig, 2001; Cochran *et al.*, 2001). Here an AK based peptide with a sequence of Ac-AAXAAAKAAAKAGY-NH<sub>2</sub> was used to study the energetics of the N3 residue in relative isolation from neighbouring charges. This peptide is intrinsically helical because of its high Ala

content. X is the variable residue. The N-terminus is acetylated and the positively charged lysine residues included for solubility are spaced  $i$ ,  $i+5$  to each other and the substitution site so they do not interact. The Tyr residue is present to give the peptide a UV absorption for determination of concentration. The penultimate Gly ensures that the  $\alpha$ -helix almost always terminates at, or before, this position so that problems arising from a helical Tyr affecting the CD signal are minimized (Chakrabartty *et al.*, 1993b).

The acetyl group is a strong N-cap so helix formation will often initiate at the acetyl group. Ala is a poor N-cap and a relatively good N1 and N2, so conformations that initiate with the first two Ala residues at N-cap, and hence X at N1 or N2, will have low populations. The third residue (X) will therefore have a high probability of being at the N3 position, increasing the sensitivity of the helix content to the N3 preference.

The stability of the substituted peptides is readily evaluated by measurements of CD intensity at 222 nm, which is proportional to the average helicity of the peptide in solution. The N3 preference can thus be found by finding the value of  $n_3$  that gives predicted helix content in agreement with experiment. No side-chain interactions are present in this sequence. The energy scales of the interior and N-cap of Rohl *et al.* (1996) as well as the energy scales of N1 (Cochran *et al.*, 2001) and N2 (Cochran and Doig, 2001) are used in the calculations, leaving the N3 preference ( $n_3$  in helix-coil theory) values as the only unknown parameter.

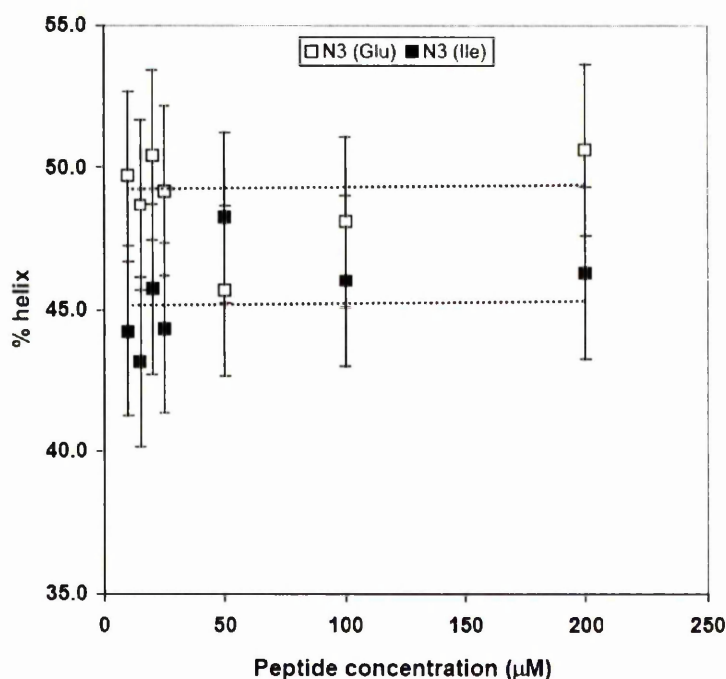
#### 4.3. Aggregation tests of N3 peptides

The helix content of each peptide was measured by CD at 222 nm at a concentration of 20  $\mu$ M in 10mM NaCl, 5mM sodium phosphate pH 7.0 at 273K. Low temperature measurements are made because the helix-coil equilibrium favours the helical conformation at lower temperature so signal intensities are consequently larger. Molar ellipticities for these types of simple peptides are known to be concentration independent (Padmanabhan *et al.*, 1990; Stapley and Doig, 1997b) indicating that aggregation does not occur. This was confirmed by checking that the



helix contents of N3(Glu) and N3(Ile) were relatively invariant between 10-200  $\mu$ M (Figure 4.1).

The peptide containing Val at the N3 position unexpectedly changes conformations at different concentrations, as shown by CD (Figure 4.2). At concentration below 5 $\mu$ M the peptide adopts a random coil conformation due to aggregation. At concentration 10 $\mu$ M, the minimum at 216 nm and zero ellipticity at 208 nm indicate  $\beta$ -sheet structure and aggregation was observed visually. There is no isodichroic point observed, indicating the presence of a mixture of secondary structures in the solution. Structures at concentrations higher than 20 $\mu$ M were thus not pursued as the determination of  $n_3$  would give uncertainty due to a mixed conformation.



**Figure 4.1.** Aggregation tests of N3(Glu) and N3(Ile) peptides. Helix contents were converted from CD ellipticity of peptides measured in 5mM phosphate buffer containing 10mM NaCl pH 7.0 at 273K. The horizontal dashed lines are to relate helicity at different concentrations and have no physical meaning.

As the N3(Val) peptide did not form monomeric helix, the helicity data could not be obtained. It is remarkable that the peptide with Ile at N3 is monomeric and helical up to 200 $\mu$ M, while the substitution of this Ile for the smaller and less

hydrophobic Val causes it to aggregate into  $\beta$ -sheet at only 20  $\mu$ M (Figure 4.2). We suspect that the  $\beta$ -sheet structure is the most stable form for many, if not all, of the peptides and it is only a very high activation energy that allows the peptides to form monomeric helices, in agreement with the ideas of Dobson (Dobson, 1999).

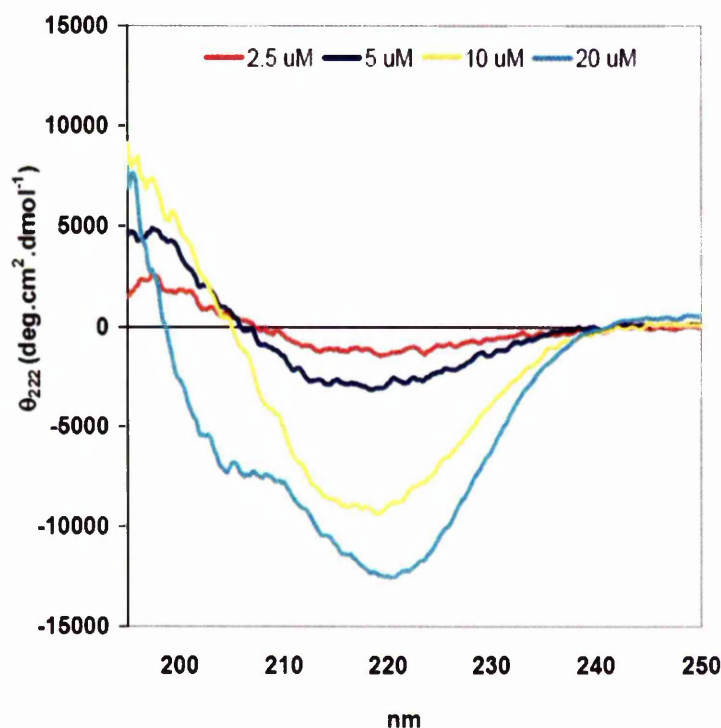


Figure 4.2. Aggregation tests of N3(Val) peptide. CD spectra were measured between 195-250 nm in 5mM phosphate buffer containing 10mM NaCl pH 7.0 at 273K.

#### 4.4. Determination of N3 values ( $n_3$ ) and $\Delta G$ from helix contents

The percentage helicities for the N3 peptides are shown in Table 4.1 and range widely from 59% for Ala to 16% for Pro. N3(Cys) was measured in the presence of 1 mM dithiotreitol to keep it reduced. This large range indicates that the N3 residue plays a significant role in the overall stability of the peptide. However, the equilibrium that exists between the multitudes of partially helical states must be analysed using helix-coil theory. This makes it possible to separate all the possible effects that can arise from making the substitution, since X could be at the N-cap, N1, N2 or N3 positions or in a coil state in different conformations. The contribution

of the helix dipole is dealt implicitly by helix coil theory as its effect is subsumed within the  $n3$  parameter.

The residue  $n3$  values are presented in rank order in Table 4.1. The order is different when compared with the order of helicity values. This confirms that analysis using helix-coil theory is essential for understanding the equilibrium. For example, a high helix content in one of these peptides may arise from a strong preference for N-cap, N1, N2 or N3, leading to high populations for that residue to be at each of these positions.

The  $n3$  values were calculated from helicity measurements using parameter values as described previously for the N1 and N2 positions (Cochran and Doig, 2001; Cochran *et al.*, 2001) using the N1N2N3 helix coil algorithm (Sun *et al.*, 2000). In these sequences the only unknown parameter is the N3 preference ( $n3$ ). Initially,  $n3$  for the four sequences with an X residue matching residues of the generic peptide (Ala, Gly, Lys, and Tyr) were solved simultaneously by fitting the equilibrium equation to the observed helicities using the N1N2N3 program (see Chapter 3 section 3.9). Then,  $n3$  for other residues were found by varying them until the calculated helix content agreed with experiment.

The statistical weight for the N3 coil to helix transition is  $n3.w$ , the product of  $w$ , the intrinsic interior preference value, and  $n3$ , the intrinsic N3 preference value (Sun *et al.*, 2000). Free energies for transfer from the coil to N3 are thus calculated as  $\Delta G = -RT \ln(n3.w)$  and the free energy change relative to Ala is given by  $\Delta\Delta G = \Delta G_{Xaa} - \Delta G_{Ala}$ .

$n3$  is an adjustment to the intrinsic helix preference ( $w$ ) and thus shows how the helix preference changes on moving from the helix interior to N3. Table 4.1 presents the free energies of each amino acid in rank order of decreasing stability. The  $n3$  values for several N3 peptides (C<sup>-</sup>, C<sup>o</sup>, G, H<sup>+</sup>, H<sup>o</sup>, N, P and R) could not be evaluated as positive numbers and have been assigned a value of zero. It is not clear why these data could not be fitted. Possibilities include errors in previously determined parameters, interactions in non-helical conformations and interactions involving these amino acids that extend beyond the N3 position.

Table 4.1. Helix contents of N3 peptides and energetic preferences for the N3 position.

N3 residue	$[\theta]_{222}^a$	% helix <sup>b</sup>	% helix AGADIR	n3 value <sup>c</sup> (range $\pm$ 3% helicity)	n3.w value	$\Delta G$ for coil to N3 transition <sup>d</sup>	$\Delta\Delta G^e$
A	-19970	59.2	59	0.59 (0.42 to 0.91)	1.00 (0.72 to 1.55)		0
E <sup>f</sup>	-17019	50.4	58	0.64 (0.51 to 0.85)	0.35 (0.27 to 0.46)	0.57 (0.70 to 0.42)	
M	-15381	45.6	46	0.42 (0.35 to 0.54)	0.28 (0.22 to 0.35)	0.70 (0.81 to 0.57)	0.36
I	-15428	45.7	48	0.60 (0.48 to 0.76)	0.28 (0.22 to 0.35)	0.70 (0.81 to 0.57)	0.54
L	-15231	45.1	51	0.28 (0.23 to 0.35)	0.24 (0.20 to 0.30)	0.77 (0.88 to 0.65)	0.36
K	-13348	40.0	46	0.18 (0.15 to 0.21)	0.18 (0.15 to 0.21)	0.93 (1.03 to 0.84)	
E <sup>g</sup>	-13478	39.9		<sup>g</sup>			
S	-13723	40.7	44	0.31 (0.27 to 0.35)	0.13 (0.11 to 0.14)	1.13 (1.21 to 1.07)	0.65
Q	-14749	43.7	53	0.41 (0.34 to 0.50)	0.12 (0.10 to 0.14)	1.16 (1.26 to 1.05)	0.25
T	-12800	37.9	43	0.64 (0.56 to 0.67)	0.12 (0.10 to 0.12)	1.17 (1.24 to 1.15)	0.65
Y	-13434	39.8	44	0.23 (0.19 to 0.26)	0.11 (0.09 to 0.13)	1.20 (1.29 to 1.12)	
D <sup>o</sup>	-11820	35.0		<sup>g</sup>			
F	-11595	34.4	39	0.31 (0.28 to 0.32)	0.08 (0.07 to 0.09)	1.35 (1.41 to 1.34)	
D <sup>r</sup>	-14594	43.2	52	0.20 (0.18 to 0.19)	0.08 (0.07 to 0.08)	1.15 (1.21 to 1.15)	
W	-10835	32.1	44	0.002 (0 to 0.039)	0.001 (0 to 0.001)	4.04 (2.43 to <sup>f</sup> )	
H <sup>o</sup>	-9645	28.6	36	0.023 (0.042 to <sup>f</sup> )	0.008 (0 to 0.015)	2.60 (2.27 to <sup>f</sup> )	
C <sup>r</sup>	-10077	29.9		0 <sup>f</sup>	0	-	
G	-9309	27.6	39	0 <sup>f</sup>	0	-	0.76
N	-9133	27.1	43	0 <sup>f</sup>	0	-	0.7
R	-7103	21.1	52	0 <sup>f</sup>	0	-	
H	-5636	16.7		<sup>g</sup>		-	
P	-5486	16.3	16	0 <sup>f</sup>	0	-	

<sup>a</sup>The peptide sequence is CH<sub>3</sub>CO-AAXA<sub>4</sub>KAKAGY-NH<sub>2</sub>.  $[\theta]_{222}$  units are deg·cm<sup>2</sup>·dmol<sup>-1</sup>. <sup>b</sup>Calculated as  $([\theta]_{222} \times 100)/(-40000 \times (1-2.5/N))$ . <sup>c</sup>From fitting with N1N2N3 program. <sup>d</sup>-RT ln( $n3w$ ), in kcal mol<sup>-1</sup>. <sup>e</sup>From Petukhov et al., (1998) & Petukhov et al., (1999). <sup>f</sup>Not possible to fit a positive value for  $n3$ . <sup>g</sup>Not fitted.

The  $n_3$  values for W, Y and F are unreliable since aromatic groups perturb the CD signal (Chakrabartty *et al.*, 1993b) though this is a minor effect for Phe and Tyr (Andrew *et al.*, 2002a). This aromatic effect may well differ at N3 from the helix centre, due to the differing conformations between these sites. The N-cap energy of Asp<sup>o</sup>, Cys<sup>o</sup>, Glu<sup>o</sup> and His<sup>+</sup> has not been measured previously. The  $n_3$  values of these protonated amino acids were therefore omitted, since this also requires knowing their N-cap preferences.

All amino acids in N3 peptides destabilise the helix relative to Ala. The free energies range from 0kcal/mol for N3(Ala) to 4kcal/mol for N3(Trp). Helix breaking residues at the third position of the peptide have a larger destabilising effect than at N1 as they are moving toward the centre of the sequence. For instance,  $\Delta\Delta G$  for N1(Pro) is 0.6kcal/mol and  $\Delta\Delta G$  N1(Gly) is 1.06kcal/mol. In this study, the  $\Delta\Delta G$  for these residues could not be measured as their  $\Delta G$  values are very high.

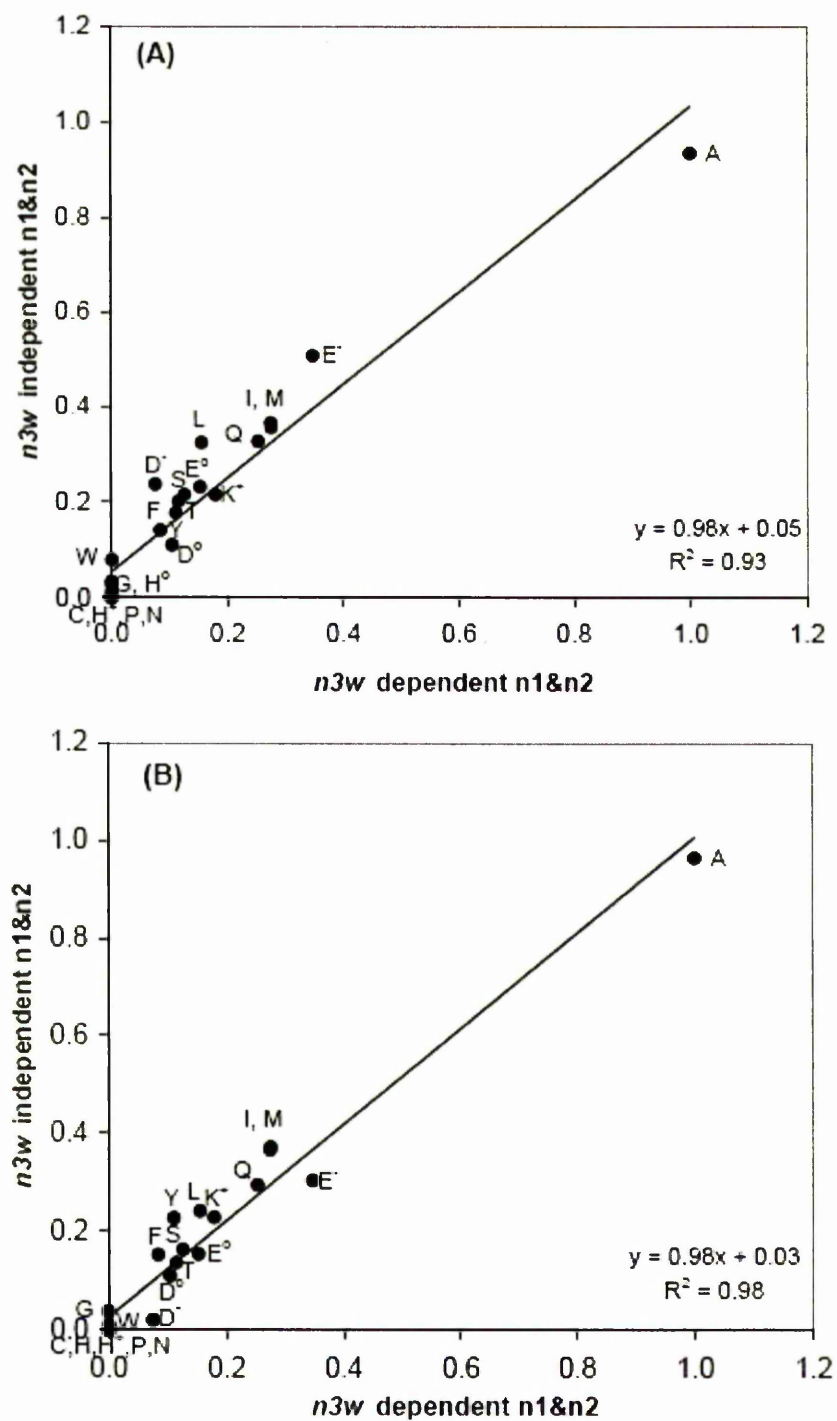
#### 4.5. Determination of N3 values ( $n_3$ ) from helix contents with different parameters

If we assume that peptide with a sequence of Ac-AAXAAAKAAAKAGY-NH<sub>2</sub> is the most populated, Ala will be the N1 and N2 residues and X will be the N3 residue. In the original helix-coil theory, N1 has a statistical weight of  $v$  which is the helix initiation constant with a value of 0.048 in the absence of N-capping interactions (Rohl *et al.*, 1996). The modified helix-coil theory, however, has  $n_1$  and  $n_2$  of Ala significantly higher at 0.27 (Cochran *et al.*, 2001) and 1.6 (Cochran and Doig, 2001), respectively, so  $n_3$  needs to be low to compensate. This may explain why the  $n_3$  for C<sup>-</sup>, G, H<sup>+</sup>, H<sup>o</sup>, N, P and R could not be evaluated as positive numbers. To evaluate this further the helicity data were refitted in two different ways using parameter values of (i)  $n_1(\text{Ala}) = 0.048$  and  $n_2(\text{Ala}) = 1$ ;  $n_1$  and  $n_2$  for other residues are as previously reported. (ii)  $n_1(\text{X}) = 0.048$  and  $n_2(\text{X}) = 1$ , where X is all amino acids. The  $n_3$  values were then calculated as described previously.

**Table 4.2.** Comparison of  $n3.w$  values and energetic preferences for the N3 amino acids calculated using different N1 and N2 values.

N3 residue	$n1(X)=N1, n2(X)=N2^*$			$n1(X)=0.048, n2(X)=1.0$			$n1(Ala)=0.048, n2(Ala)=1.0$		
	n3w value	$\Delta G$	$\Delta\Delta G$	n3w value	$\Delta G$	$\Delta\Delta G$	n3w value	$\Delta G$	$\Delta\Delta G$
A	1.00	0.0	0	0.94	0.04	0	0.96	0.02	0
C <sup>-</sup>	0.00	high	high	0.01	2.4	2.4	0.00	high	high
D <sup>-</sup>	0.08	1.4	1.4	0.24	0.8	0.7	0.02	2.1	2.1
D <sup>0</sup>	0.10	1.2	1.2	0.16	1.0	0.9	0.11	1.2	1.2
E <sup>-</sup>	0.35	0.6	0.6	0.51	0.4	0.3	0.30	0.6	0.6
E <sup>0</sup>	0.15	1.0	1.0	0.23	0.8	0.8	0.15	1.0	1.0
F	0.08	1.3	1.3	0.14	1.1	1.0	0.15	1.0	1.0
G	0.00	high	high	0.03	1.9	1.9	0.04	1.8	1.8
H <sup>+</sup>	0.00	high	high	0.00	high	high	0.04	1.8	1.8
H <sup>0</sup>	0.00	high	high	0.03	1.8	1.8	0.00	high	high
I	0.27	0.7	0.7	0.36	0.5	0.5	0.37	0.5	0.5
K	0.18	0.9	0.9	0.22	0.8	0.8	0.23	0.8	0.8
L	0.24	0.8	0.8	0.32	0.6	0.6	0.04	1.8	1.8
M	0.28	0.7	0.7	0.35	0.6	0.5	0.37	0.5	0.5
N	0.00	high	high	0.00	high	high	0.00	high	high
P	0.00	high	high	0.00	high	high	0.00	high	high
Q	0.12	1.2	1.2	0.15	1.0	1.0	0.14	1.1	1.1
R	0.00	high	high	0.00	high	high	0.00	high	high
S	0.13	1.1	1.1	0.21	0.8	0.8	0.16	1.0	1.0
T	0.12	1.2	1.2	0.20	0.9	0.8	0.14	1.1	1.1
W	0.00	4.0	4.0	0.08	1.4	1.3	0.01	2.4	2.4
Y	0.11	1.2	1.2	0.18	0.9	0.9	0.23	0.8	0.8

\*Values are from Table 4.1.



**Figure 4.4.** Plots of  $n3.w$  values from different fittings. The abscise is the  $n3.w$  values fitted using previously reported  $n1$  (Cochran *et al.*, 2001) and  $n2$  (Penel *et al.*, 1999; Cochran and Doig, 2001) values. (A) The ordinate is the  $n3.w$  values fitted using  $n1(X) = v = 0.048$  and  $n2(X) = 1$ . (B) The ordinate is the  $n3.w$  values fitted using  $n1(\text{Ala}) = v = 0.048$  and  $n2(\text{Ala}) = 1$ .

There are good agreements of the  $n3w$  values when fitted differently. There is however variation on the  $n3w$  values observed, with D<sup>+</sup>, L and W show the most notable disparities (Table 4.2). For example, the  $n3w$  of D<sup>+</sup> calculated using  $n1(X) = 0.048$  and  $N2(X) = 1$  is higher ( $n3w = 0.24$ ) compared to that calculated using previously reported values of  $n1$  and  $n2$  ( $n3w = 0.08$ ). Similarly, the  $n3w$  of L<sup>-</sup> calculated using  $n1(\text{Ala}) = 0.048$  and  $n2(\text{Ala}) = 1$  is lower ( $n3w = 0.04$ ) compared to that calculated using previously reported values of  $n1$  and  $n2$  ( $n3w = 0.24$ ). The reason behind this is not totally clear. It could be because different fittings give variations in helix populations thus changing the equilibrium constant of those residues. The  $n3$  values for N, P and R still cannot be evaluated as positive values. However, the  $n3$  value for G H<sup>o</sup> and C<sup>-</sup> can now be obtained, but still give a high  $\Delta\Delta G$ . This implies that the destabilising effect of these residues when transferred from coil to N3 is dependent on different fittings.

Results from different fittings were compared by plotting the  $n3w$  values as can be seen in Figure 4.4. In general, the deviations from the best fit line are small. The fitting using previously reported values of  $n1$  and  $n2$  (Cochran and Doig, 2001; Cochran *et al.*, 2001) however is the most sound. Its advantage is that it gives the absolute values of  $n3$ . Fitting using parameter values such as  $n1(\text{Ala})$  and  $n2(\text{Ala})$  only gives energies relative to each other. The absolute  $n3$  values, i.e. values fitted using previously reported  $n1$  and  $n2$  values were thus used thereafter throughout this chapter.

#### 4.6. Comparison with other reported N3 values

Free energies for N3 for non-polar and non-charged polar residues for a series of AR peptides have been reported (Petukhov *et al.*, 1998; Petukhov *et al.*, 1999) and are included in Table 4.1 for comparison. No other  $\Delta\Delta G$  (N3) values have been reported. The free energies measured here are somewhat higher than those of Petukhov *et al.* This may be due to differences in their peptide sequence and computational methods of Petukhov *et al.* The peptides studied here are 16 mers, with a wider helicity range, from 59% for N3(Ala) to 16% for N3(Pro), while those of Petukhov *et al.* which range from 28% of N3(Thr) to 47% of N3(Ala). The data presented here may therefore give more accurate N3 energies.



#### 4.7. Comparison with N3 values calculated using AGADIR

The AGADIR program (Muñoz and Serrano, 1994) predicts the helical behavior of monomeric peptides in aqueous solution. Table 4.1 also includes predictions of helix contents for all the peptides studied here using the AGADIR program. Results calculated by AGADIR are all quite similar. Errors range from 0% (Ala and Met) to 31% (Arg). The most notable differences are found in C, G, N, R and W. The average deviation of the prediction from experimental values (obtained from CD measurements) using AGADIR is -2% of helical content, with a standard deviation of 6%. A peptide with 30% helix content would be predicted to have  $28 \pm 6\%$ .

#### 4.8. pH titrations of charged amino acids

The helicities of the N3(Asp), N3(Cys), N3(Glu) and N3(His) peptides were titrated as a function of pH. pKa values of the ionisable side-chain groups were evaluated between approximately pH 2-10 by curve fitting to the Henderson-Hasselbach equation (Figures 4.4). The upper range of accessible pH is given by Lys deprotonation which leads to peptide aggregation.

Titration of N3(Asp) and N3(Glu) demonstrate that the neutrally charged species are both destabilising compared to the negatively charged residues. As the pH rises above pH 6, helix content of N3(Asp) and N3(Glu) increases maximally by 8% and 10%, respectively, in comparison to their protonated counterparts. As previously described, the  $n_3$  values calculation of protonated Asp and Glu were omitted as their N-cap values are unknown. The equation to calculate the free energy for protonation, given by  $\Delta G = -RT \ln[n_3(\text{charged})/n_3(\text{uncharged})]$  could therefore not be calculated.

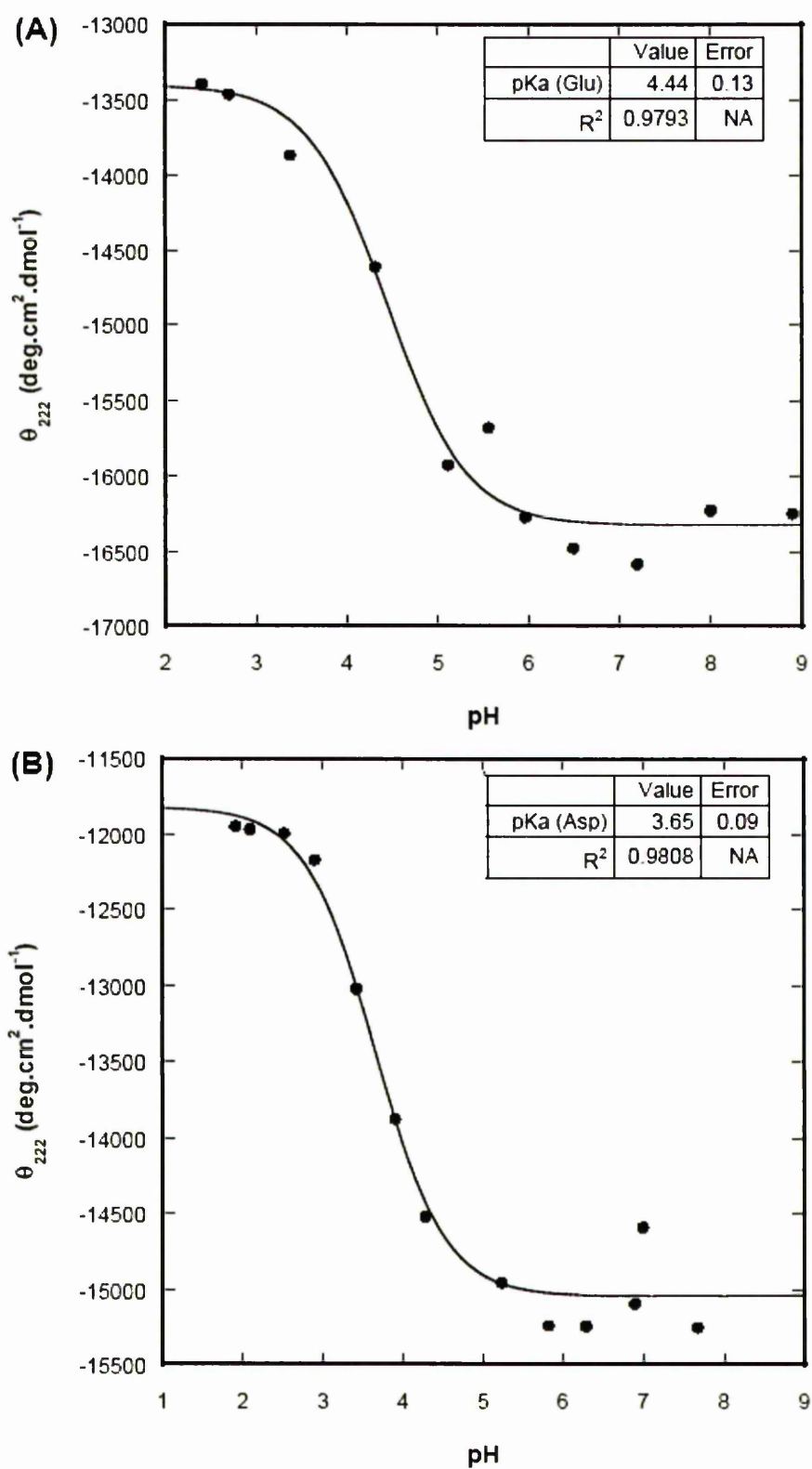
The pKa of N3(Glu) is  $4.48 \pm 0.11$  and this value is similar to model compounds (Glu in an N-acetyl acid amide has a pKa of 4.5 (Nozaki and Tanford, 1967)). The unperturbed pKa value for residue Glu at N3 indicates that the side-chain is not involved in any strong interactions, yet the helix content is strongly influenced by the ionisation state.

The pKa of N3(Asp) of  $3.64 \pm 0.09$  is below those of other reported compounds. In comparison with the pKa value of Asp in N-acetyl acid amide (pKa 4.1 (Nozaki and Tanford, 1967)), N3(Asp) is about 0.5 units lower. The pKa of Asp has also been measured in the coil state of  $\text{CH}_3\text{CO}-[\text{AADAA}]-\text{NH}_2$  and reported as  $3.91 \pm 0.02$  (Huyghues-Despointes *et al.*, 1993b). The significant difference between the perturbations of the pKas of Asp and Glu could be attributed to the shorter side-chain of Asp, interacting more strongly with the helix dipole.

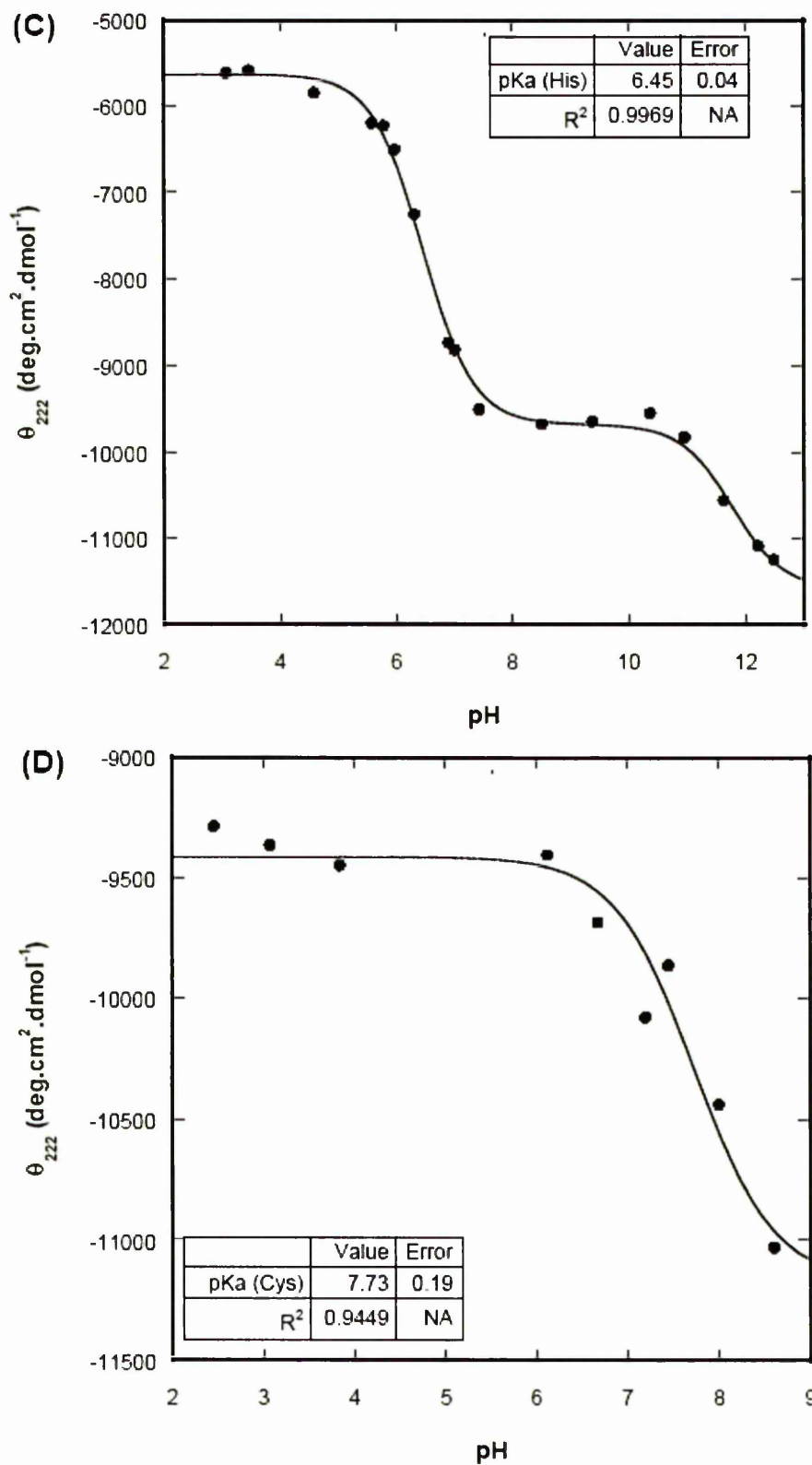
The pKa of N3(Cys) of  $7.73 \pm 0.12$  differs from other reported pKa values (Miranda, 2003). However, the pKa value of Cys in a helical peptide is significantly lower than that in a coil peptide (pKa Cys  $8.69 \pm 0.07$  (Kortemme and Creighton, 1995)). The increase in helix content of N3(Cys) at higher pH was not determined since titration beyond pH 8.5 in this peptide begins to deprotonate the side-chain of Lys residues leading to peptide aggregation.

In general, Asp, Glu, and Cys residues are helix stabilising at N3 when in the ionised form. The stabilising contribution might be due to the ability of these residues to form a hydrogen bond to one of unsatisfied amide groups at the N-terminus (Penel *et al.*, 1999). Another reasonable contribution to stabilisation is the electrostatic interactions of the charged side-chains with the positive end of the helix dipole. These two contributions cannot be separated experimentally, however.

The helicity of N3(His) is also pH dependent (Fig. 4.4d). The uncharged N3(His) at pH 9 is about 9% more helical than the fully protonated species at pH 3. The  $\text{His}^+$  side-chain destabilises the helix, possibly as a result of repulsion with the positive end of the helix macrodipole and free amide groups at the N-terminus. The pKa for His at N3 is  $6.45 \pm 0.04$ , which is comparable to values commonly found in the range of 6.5 and 7.1 (Kyte, 1995). His-dipole interactions have also been measured (Armstrong and Baldwin, 1993).  $\text{His}^+$  is destabilising to the helix and this effect is reduced in 1 M NaCl. The pKa of His increases as the His is moved further from the N-terminus. This was again interpreted as a charge-dipole effect. The titration above pH 11 is due to Tyr deprotonation.



**Figure 4.4.** pH titrations of N3 peptides. (A) N3(Glu), (B) N3(Asp), continued...



**Figure 4.4.** (from previous page). pH titrations for N3 peptides. (C) N3(His), (D) N3(Cys).

#### 4.9. Correlation of N3 preferences with N3 propensities

Figure 4.5 shows the correlation between the N3 energies and the N3 amino acid propensities found by surveying helices in proteins (Penel *et al.*, 1999). There is a weak correlation observed. Deviations from the line of best fit for Glu, Gln and Asp show that these are more favoured at N3 in proteins than in this peptide. This is due to these side-chains form capping box hydrogen bonds to the backbone NH of the N-cap (Harper and Rose, 1993).

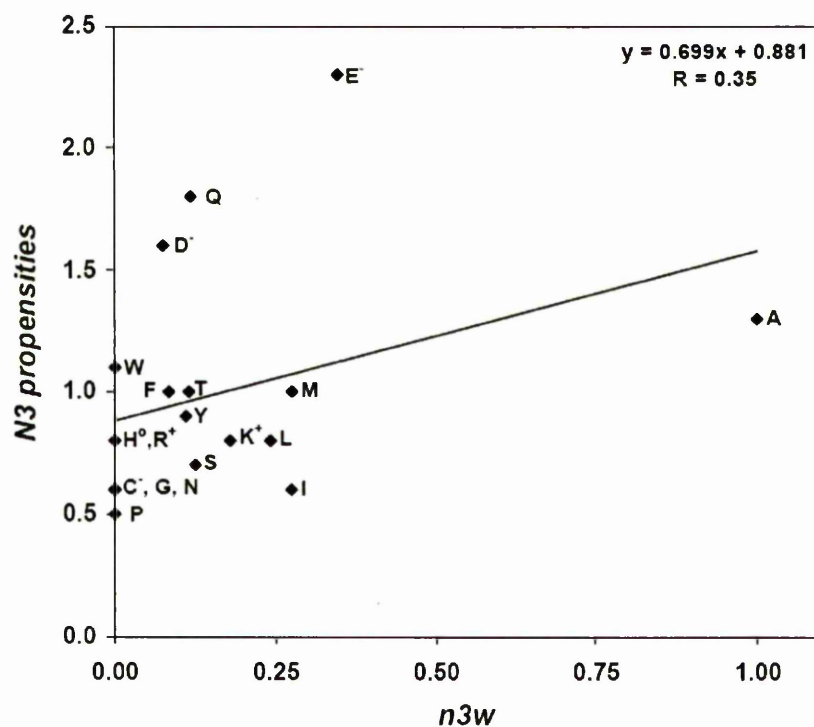
While propensities are measured relative to all other positions in a protein, N3 energies are measured relative to the coil state. The statistical propensities observed in proteins are thus not free energies, though they often correlate with them. Surveying protein structures is thus no substitute for experimental measurements of free energies from amino acid substitutions, whether by peptide synthesis or site-directed mutagenesis in proteins.

#### 4.10. Comparison of N3 preferences with N-cap preferences

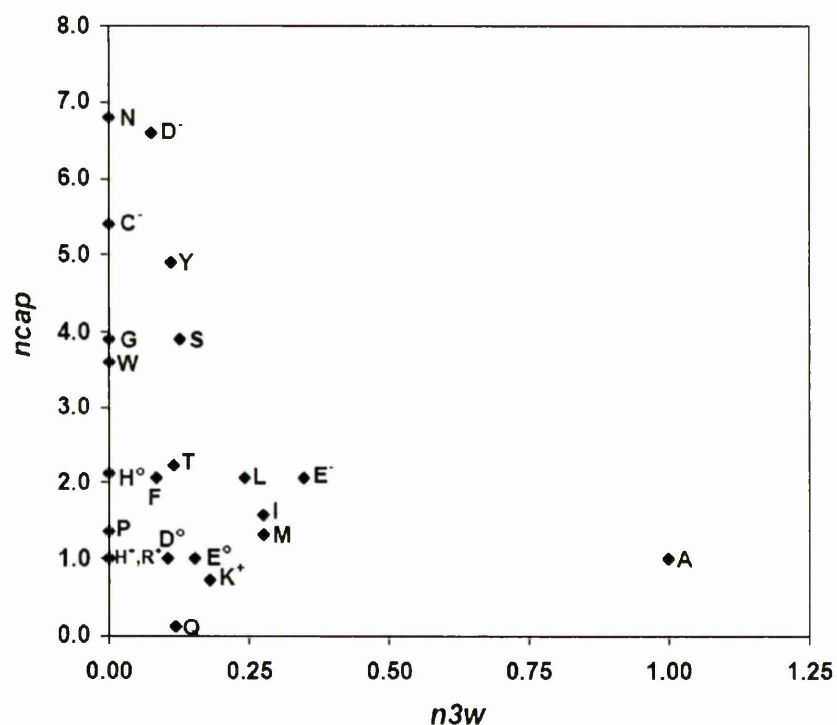
Figure 4.6 shows a plot of the N-cap preference (Rohl *et al.*, 1996) versus the N3 preference ( $n3.w$ ). There is no correlation, showing that the amino acid preferences for the N-cap position are completely different to N3. Although both are being at the helix N-terminus, they adopt different conformations as N-cap and N3 is in non-helical and helical conformations, respectively.

#### 4.11. Comparison of N3 preferences with N1 and N2 preferences

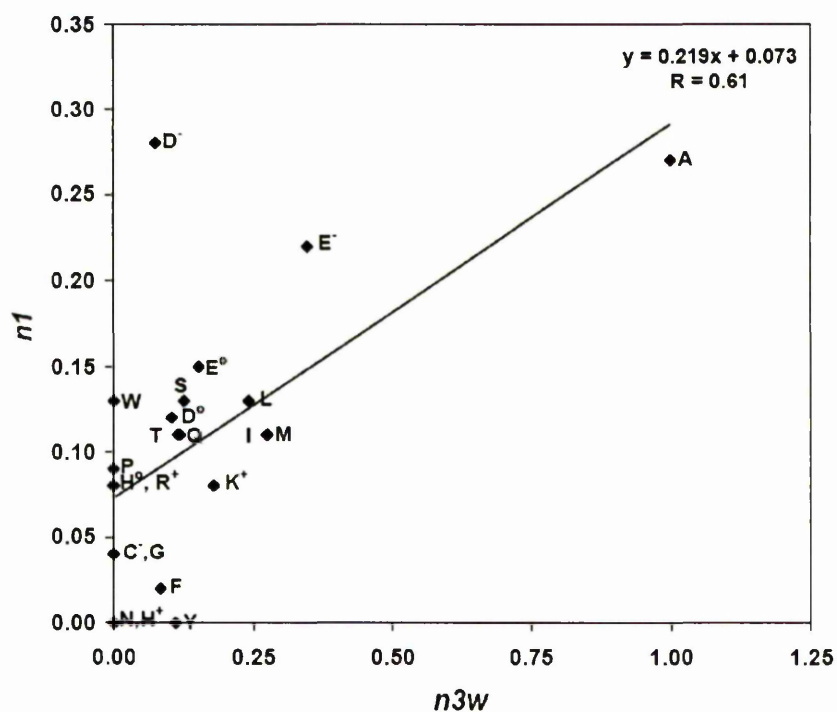
Figure 4.7 and 4.8 compare the N1 and N2 preferences (Cochran and Doig, 2001; Cochran *et al.*, 2001), respectively, to the N3 preference. There is a reasonable correlation, showing that the forces at both sites are similar. The notable deviations from the linear correlation are Asp<sup>-</sup> and Glu<sup>-</sup> in both comparisons of which have higher N1 and N2 preferences than N3, respectively. A possible reason for this is that Asp<sup>-</sup> and Glu<sup>-</sup> at N1 and N2 can form  $i,i+1$  hydrogen bonds with the adjacent backbone amide group (Penel *et al.*, 1999). This is not possible when they are at N3 because the amide group at N4 is occupied hydrogen bonding to the N-cap C=O.



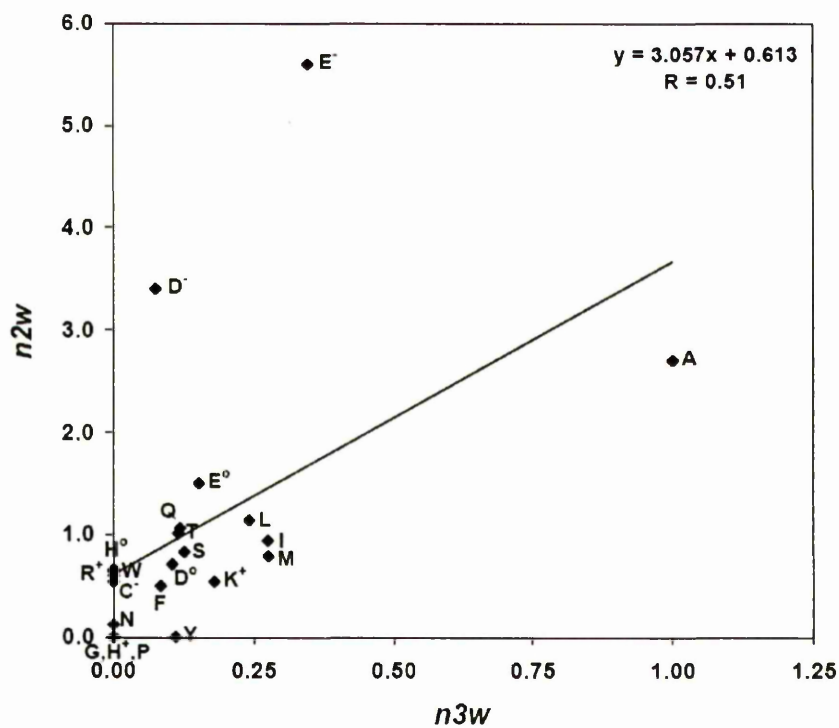
**Figure 4.5.** Correlation between N3 propensity (Pg N3) and N3 statistical weight (n3.w).



**Figure 4.6.** Correlation between N-cap statistical weight (n) and N3 statistical weight (n3.w).



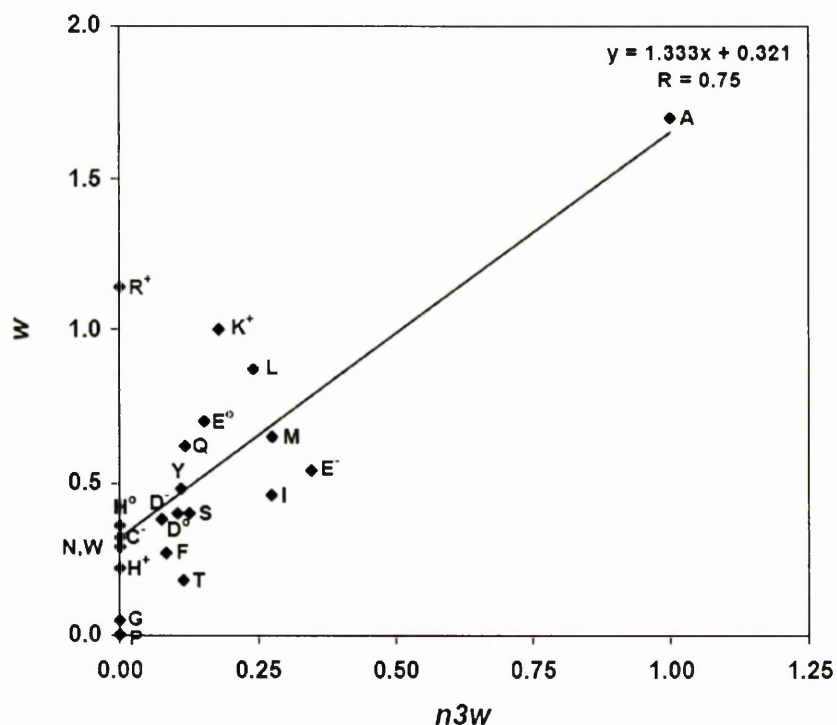
**Figure 4.7.** Correlation between N1 statistical weight ( $n1$ ) and N3 statistical weight ( $n3.w$ )



**Figure 4.8.** Correlation between N2 statistical weight ( $n2.w$ ) and N3 statistical weight ( $n3.w$ ).

#### 4.13. Comparison of N3 preferences with helix interior preferences

Figure 4.9 compares preferences at N3 and helix interior positions. There is a good correlation showing that the N3 and interior environments are similar. Deviations from a linear correlation shown by Arg<sup>+</sup> and Lys<sup>+</sup> can be attributed to charge effects where both residues are disfavoured at N3 compared to the helix interior. In general, residues with the ability to hydrogen bond with free amide groups at the helix N-terminus (for example Asp<sup>0</sup>, Glu, Ser and Thr) are more stable at N3 than at helix interior.



**Figure 4.9.** Correlation between interior statistical weight ( $w$ ) and N3 statistical weight ( $n3.w$ ).

#### 4.14. Conformational entropy of residues at N3 position

The major force opposing protein folding is the loss of conformational entropy. Although N3 residues are on the surface of proteins, thus exposed to solvent, they can be more restricted than in the unfolded state. Conformational entropy is related to the distribution of side-chain rotamers, calculated as  $\Delta S = -R \sum p_i \ln p_i$ , where  $p_i$  is the fractional population of each rotamer state  $i$ . Table 4.3 shows conformational



entropy calculation for residues at N3 based on  $\chi_1$  rotamer distribution, as  $\chi_1$  is the most restricted rotamer. The rotamer populations of N3 were searched from ASTRAL database (SCOP 1.63 Sequence Resources) (Brenner *et al.*, 2000). The data set contained 1135 chains of 4778 helices with a total of 195875 residues (see Material and Methods). It is assumed that the conformations adopted by side-chains in protein crystal structures apart from helices are representative of unfolded conformations (Pickett and Sternberg, 1993). Conformational entropy for residues at N3 is higher than at internal position (for review see Doig and Sternberg (1995)). This is simply because residues at N3 are less restricted.

#### 4.15. Free energies of amino acid side-chain rotamers at N3 position

Conformational entropies for transferring residues to N3 position do not consider the difference in side-chain rotamers adopted by N3 residues. For example, Glu at N3 has  $\chi_1$  67% gauche<sup>+</sup> ( $g^+$ ), 31% trans ( $t$ ) and 3% gauche<sup>-</sup> ( $g^-$ ). The  $\chi_1$   $g^+$  is thus more favoured than  $g^-$  and  $t$ , reflecting differences in energies between these conformations. This force, accompanying conformational entropy, has been suggested previously to oppose protein folding (Penel and Doig, 2001).

Observed rotamer populations in the coil or helix states can be converted to the relative energies by assuming that the populations are in a Boltzmann distribution,  $E_i = -R.T.\ln p_i$ .  $E_i$  is the free energy change for restricting the side-chain to rotamer  $i$  and  $p_i$  is the fraction of rotamer  $i$  where all rotamer are populated (see Chapter 3 section 3.10). The rotamer energies of residues at N-cap, helix interior and  $\beta$ -sheet have previously been determined by (Stapley and Doig, 1997a). Similar attempts for N3 were thus made here to broaden the energetics analysis of residues at N3. As  $\chi_1$  and  $\chi_2$  are not independent (McGregor *et al.*, 1987), all the possible  $\chi_1\chi_2$  pairs where  $\chi_2$  is an sp<sup>3</sup>-sp<sup>3</sup> bond were also analysed. The results are shown in tables 4.3 and 4.4.

When a residue is transferred from coil to N3, the folded state free energy ( $\Delta E = E_{\text{coil}} - E_{\text{N3}}$ ) can be calculated as the difference between the energy of rotamer  $i$  in helix and energy of rotamer  $i$  in coil. The  $E_{\text{N3}}$  consists of the free energy terms for rotamer energy and for transferring an amino acid from coil to N3 ( $\Delta G_{\text{coil-N3}}$ ). The

rotamer free energies for each rotamer were calculated using equations 3.13, 3.14 and 3.15 in the Materials and Methods chapter.

Although it has been previously mentioned in the Material and Method section, it is worth mentioning the method again to further clarify the calculation using the real data in Table 4.4. The following example is thus given. If the rotamer fractional population for Glu  $g^+g^+$  in coil ( $p_i$ ) is 0.15 (1063/7159), by applying  $E_i = -R.T.\ln p_i$ , rotamer energy in the coil ( $E_i$ ) is 1.04kcal/mol. Similarly, if the rotamer fractional population of Glu  $g^+g^+$  in helix ( $p_i'$ ) is 0.07 (50/685), the rotamer energy ( $E_i'$ ) is 1.42kcal/mol. The free energy change for transferring Glu from coil to N3  $g^+g^+$  ( $\Delta G_{\text{coil-N3}}$ ) is 0.57kcal/mol. This value is added to rotamer energy in helix giving energy for  $g^+g^+$   $1.42 + 0.57 = 1.99$ kcal/mol. The rotamer energies ( $\Delta E$ ) for transferring Glu from coil to N3 is thus  $1.99 - 1.04 = 0.95$ kcal/mol for the  $g^+g^+$  rotamer. The difference means that  $g^+g^+$  in coil is more favoured than in helix. Likewise, the free energies for other rotamers can then be determined. Calculations in table 4.3 and 4.4 used  $T = 273\text{K}$ , as most experimental measurements were made at this temperature. The  $\Delta G_{\text{coil-N3}}$  values were taken from Table 4.1.

The  $\Delta E$  values in table 4.3 and 4.4 are mainly positive, primarily because the restriction of a side-chain into any rotamer in a helix carries an entropic penalty (Creamer and Rose, 1992). The differences in  $\Delta G$  between rotamers are large, showing that differing rotamer energies are an important aspect of protein stability.

Table 4.3.  $\chi_1$  rotamer energies and conformational entropy for transferring a residue from coil to N3.

Amino acid	Rotamer energy in coil (kcal/mol)			$\Delta G_{\text{coil-N3}}^a$ (kcal/mol)	Rotamer energy in N3 (kcal/mol)			$\Delta E_{\text{coil-N3}}$ (kcal/mol)			-TAS (interior) (kcal/mol) <sup>f</sup>	-TAS (N3) (kcal/mol)	-TAS (interior) (kcal/mol) <sup>f</sup>
	$g^+$	$t$	$g^-$		$g^+$	$t$	$g^-$	$g^+$	$t$	$g^-$			
C	0.39 $\pm$ 0.01	0.67 $\pm$ 0.02	0.82 $\pm$ 0.02	na	0.2 $\pm$ 0.1 <sup>b,c</sup>	0.7 $\pm$ 0.1	2.0 $\pm$ 1.4	-	-	-	0.38	0.85	0.85
D	0.47 $\pm$ 0.01	0.54 $\pm$ 0.01	0.83 $\pm$ 0.01	1.15 $\pm$ 0.57	1.20 $\pm$ 0.04 <sup>e</sup>	2.5 $\pm$ 0.1	4.1 $\pm$ 0.5	0.72 $\pm$ 0.05 <sup>d</sup>	2.0 $\pm$ 0.1	3.2 $\pm$ 0.5	0.16	0.78	0.78
E	0.32 $\pm$ 0.01	0.65 $\pm$ 0.01	1.07 $\pm$ 0.02	0.57 $\pm$ 0.28	0.8 $\pm$ 0.2	1.2 $\pm$ 0.2	2.5 $\pm$ 0.3	0.5 $\pm$ 0.2	0.6 $\pm$ 0.2	1.5 $\pm$ 0.3	0.39	1.46	1.46
F	0.27 $\pm$ 0.01	0.80 $\pm$ 0.01	0.98 $\pm$ 0.02	1.35 $\pm$ 0.67	2.1 $\pm$ 0.1	1.5 $\pm$ 0.1	4.1 $\pm$ 1.8	1.8 $\pm$ 0.1	0.7 $\pm$ 0.1	3.2 $\pm$ 1.8	0.33	0.62	0.62
H	0.32 $\pm$ 0.01	0.70 $\pm$ 0.02	0.97 $\pm$ 0.02	2.60 <sup>e</sup>	2.9	3.1	<sup>f</sup>	2.6	2.4	-	0.36	0.95	0.95
I	0.18 $\pm$ 0.004	1.19 $\pm$ 0.02	0.94 $\pm$ 0.01	0.70 $\pm$ 0.34	0.8 $\pm$ 0.1	2.0 $\pm$ 0.2	3.2 $\pm$ 0.6	0.6 $\pm$ 0.1	0.8 $\pm$ 0.3	2.2 $\pm$ 0.6	0.21	0.76	0.76
K	0.27 $\pm$ 0.01	0.68 $\pm$ 0.01	1.18 $\pm$ 0.02	0.93 $\pm$ 0.46	1.3 $\pm$ 0.1	1.3 $\pm$ 0.1	3.1 $\pm$ 0.4	1.0 $\pm$ 0.1	0.6 $\pm$ 0.1	2.0 $\pm$ 0.4	0.42	1.89	1.89
L	0.20 $\pm$ 0.004	0.68 $\pm$ 0.01	1.98 $\pm$ 0.03	0.77 $\pm$ 0.37	1.2 $\pm$ 0.1	1.1 $\pm$ 0.1	3.6 $\pm$ 0.6	1.0 $\pm$ 0.1	0.4 $\pm$ 0.2	1.7 $\pm$ 0.6	0.38	0.71	0.71
M	0.30 $\pm$ 0.01	0.66 $\pm$ 0.02	1.13 $\pm$ 0.03	0.70 $\pm$ 0.33	1.2 $\pm$ 0.2	1.0 $\pm$ 0.2	2.8 $\pm$ 0.6	0.9 $\pm$ 0.2	0.4 $\pm$ 0.2	1.7 $\pm$ 0.6	0.42	1.46	1.46
N	0.38 $\pm$ 0.01	0.61 $\pm$ 0.01	0.94 $\pm$ 0.02	too high	<sup>g</sup>	<sup>g</sup>	<sup>g</sup>	-	-	-	0.26	1.03	1.03
Q	0.28 $\pm$ 0.01	0.69 $\pm$ 0.01	1.13 $\pm$ 0.02	1.16 $\pm$ 0.57	1.4 $\pm$ 0.1	1.9 $\pm$ 0.2	2.9 $\pm$ 0.3	1.1 $\pm$ 0.1	1.2 $\pm$ 0.2	1.8 $\pm$ 0.3	0.41	1.73	1.73
R	0.29 $\pm$ 0.01	0.68 $\pm$ 0.01	1.12 $\pm$ 0.02	too high	<sup>g</sup>	<sup>g</sup>	<sup>g</sup>	-	-	-	0.45	1.88	1.88
S	0.74 $\pm$ 0.01	0.77 $\pm$ 0.01	0.37 $\pm$ 0.01	1.13 $\pm$ 0.56	1.4 $\pm$ 0.1	2.1 $\pm$ 0.2	1.8 $\pm$ 0.1	0.7 $\pm$ 0.1	1.3 $\pm$ 0.2	1.5 $\pm$ 0.1	0.53	1.11	1.11
T	0.54 $\pm$ 0.01	1.24 $\pm$ 0.02	0.34 $\pm$ 0.01	1.17 $\pm$ 0.58	1.3 $\pm$ 0.1	3.9 $\pm$ 0.5	2.2 $\pm$ 0.1	0.7 $\pm$ 0.1	2.6 $\pm$ 0.5	1.9 $\pm$ 0.1	0.25	1.08	1.08
V	0.22 $\pm$ 0.004	1.25 $\pm$ 0.02	0.79 $\pm$ 0.01	na	0.07 $\pm$ 0.01 <sup>b</sup>	1.5 $\pm$ 0.1	1.6 $\pm$ 0.1	-	-	-	0.23	0.43	0.43
W	0.32 $\pm$ 0.01	0.75 $\pm$ 0.02	0.89 $\pm$ 0.03	4.04 <sup>e</sup>	4.5	4.3	<sup>f</sup>	4.2	3.6	-	0.36	0.99	0.99
Y	0.28 $\pm$ 0.01	0.77 $\pm$ 0.01	1.01 $\pm$ 0.02	too high	<sup>g</sup>	<sup>g</sup>	<sup>g</sup>	-	-	-	0.37	1.13	1.13
MEAN <sup>h</sup>	0.34	0.78	0.96	1.11	1.26	1.84	3.05	0.95	1.05	2.05	0.35	0.95	0.95

<sup>a</sup>Values from Table 4.1. <sup>b</sup>Calculations not include  $\Delta G_{\text{coil-N3}}$ . <sup>c</sup>Error is sum of sampling error and experimental error in  $\Delta G_{\text{coil-N3}}$ , estimated as half of the difference of high and low  $\Delta G$  values. <sup>d</sup>Error is sum of coil and helix energy errors. <sup>e</sup>Not possible to calculate error. <sup>f</sup>Not possible to calculate due to zero number of observation. <sup>g</sup>Energy is too large. <sup>h</sup>Rotamer energy values for C, H, N, V, w and Y were excluded. <sup>i</sup>From (Doig and Sternberg, 1995).

**Table 4.4.**  $\chi_1\chi_2$  rotamer energies for transferring a residue from coil to N3.

Amino acid	$\chi_1\chi_2$	Number of observations		Rotamer energy in coil (kcal/mol)	$\Delta G_{\text{coil-N3}}^a$ (kcal/mol)	Rotamer energy in N3 (kcal/mol)	$\Delta E_{\text{coil-N3}}^a$ (kcal/mol)
		Coil	N3				
E	$g^+g^+$	1063	50	$1.04 \pm 0.02$	$0.57 \pm 0.15$	$2.0 \pm 0.2^b$	$1.0 \pm 0.2^c$
	$g^+t$	2489	339	$0.57 \pm 0.01$		$1.0 \pm 0.2$	$0.4 \pm 0.2$
	$g^+g^-$	459	69	$1.49 \pm 0.05$		$1.8 \pm 0.2$	$0.3 \pm 0.3$
	$tg^+$	106	12	$2.29 \pm 0.05$		$2.8 \pm 0.3$	$0.5 \pm 0.4$
	tt	1587	146	$0.82 \pm 0.01$		$1.4 \pm 0.2$	$0.6 \pm 0.2$
	$tg^-$	467	51	$1.48 \pm 0.02$		$2.0 \pm 0.2$	$0.5 \pm 0.2$
	$g^-g^+$	247	3	$1.83 \pm 0.03$		$3.5 \pm 0.5$	$1.7 \pm 0.5$
	$g^-t$	702	15	$1.26 \pm 0.02$		$2.6 \pm 0.3$	$1.4 \pm 0.3$
	$g^-g^-$	49	0	$2.70 \pm 0.08$		<sup>d</sup>	-
Total		7169	685				
I	$g^+g^+$	1115	51	$1.00 \pm 0.02$	$0.70 \pm 0.12$	$1.4 \pm 0.2$	$0.4 \pm 0.2$
	$g^+t$	3727	114	$0.35 \pm 0.01$		$1.0 \pm 0.2$	$0.6 \pm 0.2$
	$g^+g^-$	215	5	$1.90 \pm 0.04$		$2.7 \pm 0.4$	$0.8 \pm 0.4$
	$tg^+$	10	0	$3.6 \pm 0.2$		<sup>d</sup>	-
	tt	608	7	$1.33 \pm 0.02$		$2.5 \pm 0.3$	$1.2 \pm 0.4$
	$tg^-$	177	12	$2.00 \pm 0.04$		$2.2 \pm 0.3$	$0.2 \pm 0.3$
	$g^-g^+$	20	0	$3.19 \pm 0.12$		<sup>d</sup>	-
	$g^-t$	1139	2	$0.99 \pm 0.02$		$3.2 \pm 0.6$	$2.2 \pm 0.6$
	$g^-g^-$	96	0	$2.34 \pm 0.06$		<sup>d</sup>	-
Total		7107	191				
K	$g^+g^+$	1141	26	$0.99 \pm 0.02$	$0.93 \pm 0.10$	$2.1 \pm 0.2$	$1.1 \pm 0.2$
	$g^+t$	2973	88	$0.47 \pm 0.01$		$1.5 \pm 0.1$	$1.0 \pm 0.2$
	$g^+g^-$	199	8	$1.94 \pm 0.04$		$2.8 \pm 0.3$	$0.8 \pm 0.3$
	$tg^+$	89	9	$2.38 \pm 0.06$		$2.7 \pm 0.3$	$0.3 \pm 0.3$
	tt	1593	66	$0.81 \pm 0.01$		$1.6 \pm 0.2$	$0.8 \pm 0.2$
	$tg^-$	326	40	$1.67 \pm 0.03$		$1.9 \pm 0.2$	$0.2 \pm 0.2$
	$g^-g^+$	40	0	$2.81 \pm 0.09$		<sup>d</sup>	-
	$g^-t$	732	4	$1.24 \pm 0.02$		$3.2 \pm 0.4$	$1.9 \pm 0.4$
	$g^-g^-$	41	0	$2.80 \pm 0.09$		<sup>d</sup>	-
Total		7134	241				
L	$g^+g^+$	192	6	$2.14 \pm 0.04$	$0.77 \pm 0.12$	$3.0 \pm 0.4$	$0.9 \pm 0.4$
	$g^+t$	5948	149	$0.27 \pm 0.004$		$1.3 \pm 0.2$	$1.0 \pm 0.2$
	$g^+g^-$	628	10	$1.49 \pm 0.02$		$2.8 \pm 0.3$	$1.3 \pm 0.3$
	$tg^+$	96	11	$2.51 \pm 0.06$		$2.7 \pm 0.3$	$0.2 \pm 0.3$
	tt	395	29	$1.74 \pm 0.03$		$2.2 \pm 0.2$	$0.4 \pm 0.2$
	$tg^-$	2331	188	$0.78 \pm 0.01$		$1.2 \pm 0.1$	$0.4 \pm 0.2$
	$g^-g^+$	11	0	$3.69 \pm 0.17$		<sup>d</sup>	-
	$g^-t$	110	0	$2.44 \pm 0.05$		<sup>d</sup>	-
	$g^-g^-$	141	2	$2.30 \pm 0.05$		$3.6 \pm 0.6$	$1.3 \pm 0.6$
Total		9852	395				

Continued next page...

**Table 4.4.** (continued from previous page)

Amino acid	$\chi_1\chi_2$	Number of observations		Rotamer energy in coil (kcal/mol)	$\Delta G_{\text{coil-N3}}^a$ (kcal/mol)	Rotamer energy in N3 (kcal/mol)	$\Delta E_{\text{coil-N3}}^a$ (kcal/mol)
		Coil	N3				
M	$g^+g^+$	465	17	$0.85 \pm 0.02$	$0.70 \pm 0.12$	$1.68 \pm 0.2$	$0.8 \pm 0.3$
	$g^+t$	782	26	$0.57 \pm 0.02$		$1.45 \pm 0.2$	$0.9 \pm 0.2$
	$g^+g^-$	38	1	$2.21 \pm 0.09$		$3.22 \pm 1.8$	$1.0 \pm 1.8$
	$tg^+$	41	3	$2.17 \pm 0.09$		$2.62 \pm 0.5$	$0.5 \pm 0.6$
	tt	451	23	$0.87 \pm 0.02$		$1.52 \pm 0.2$	$0.7 \pm 0.2$
	$tg^-$	166	32	$1.41 \pm 0.04$		$1.34 \pm 0.2$	$-0.1 \pm 0.2$
	$g^-g^+$	8	0	$3.05 \pm 0.20$		<sup>d</sup>	-
	$g^-t$	247	2	$1.19 \pm 0.03$		$2.84 \pm 0.6$	$1.7 \pm 0.6$
	$g^-g^-$	30	0	$2.3 \pm 0.1$		<sup>d</sup>	-
Total		2228	104				
Q	$g^+g^+$	861	19	$0.86 \pm 0.02$	$1.16 \pm 0.11$	$2.7 \pm 0.2$	$1.9 \pm 0.2$
	$g^+t$	1489	178	$0.56 \pm 0.01$		$1.5 \pm 0.1$	$0.9 \pm 0.1$
	$g^+g^-$	173	27	$1.73 \pm 0.04$		$2.5 \pm 0.6$	$0.8 \pm 0.6$
	$tg^+$	84	2	$2.12 \pm 0.06$		$3.9 \pm 0.6$	$1.8 \pm 0.6$
	tt	810	54	$0.89 \pm 0.02$		$2.1 \pm 0.2$	$1.3 \pm 0.2$
	$tg^-$	268	35	$1.49 \pm 0.03$		$2.4 \pm 0.2$	$0.9 \pm 0.2$
	$g^-g^+$	91	4	$2.08 \pm 0.06$		$3.6 \pm 0.4$	$1.5 \pm 0.5$
	$g^-t$	398	9	$1.28 \pm 0.03$		$3.1 \pm 0.3$	$1.8 \pm 0.3$
	$g^-g^-$	37	1	$2.57 \pm 0.09$		$4.3 \pm 2.1$	$1.7 \pm 2.1$
Total		4211	329				
R	$g^+g^+$	846	21	$1.05 \pm 0.02$	too high	<sup>e</sup>	-
	$g^+t$	2445	56	$0.48 \pm 0.01$		<sup>e</sup>	-
	$g^+g^-$	162	2	$1.95 \pm 0.04$		<sup>e</sup>	-
	$tg^+$	73	2	$2.38 \pm 0.06$		<sup>e</sup>	-
	tt	1341	61	$0.80 \pm 0.01$		<sup>e</sup>	-
	$tg^-$	256	33	$1.70 \pm 0.03$		<sup>e</sup>	-
	$g^-g^+$	39	0	$2.72 \pm 0.09$		<sup>e</sup>	-
	$g^-t$	683	7	$1.17 \pm 0.02$		<sup>e</sup>	-
	$g^-g^-$	37	0	$2.75 \pm 0.09$		<sup>e</sup>	-
Total		5882	182				

<sup>a</sup> Values from table 4.1. <sup>b</sup> Error is sum of sampling error and experimental error in  $\Delta G_{\text{coil-N3}}$ , estimated as half of the difference of high and low  $\Delta G$  values. <sup>c</sup> Error is sum of coil and helix energy errors. <sup>d</sup> Not possible to calculate due to zero number of observation. <sup>e</sup> Energy is too large.

#### 4.16. Discussion

CD measurement on N3 peptides has given helicities ranging widely from 59% for Ala to 16% for Pro. The order is different when compared with the order of  $n3$  values confirming the importance of the analysis using helix coil theory to explain the equilibrium.  $n3$  values for C<sup>-</sup>, C<sup>o</sup>, G, H<sup>+</sup>, H<sup>o</sup> N, P and R cannot be determined. In

general, all amino acids in N3 peptides destabilise the helix relative to Ala. The most stabilising effect for Ala could be attributed to similar environment between the N3 and helix interior.

The  $n$ ,  $n1$ ,  $n2$ , and  $n3$  parameters represent the probabilities that a particular residue will adopt the corresponding N-cap, N1, N2 or N3 conformations, respectively. These parameters are interdependent, so the  $n3$  values could be perturbed by errors in  $n$ ,  $n1$ , or  $n2$  values. This problem is most important when X in the sequence has a high N-cap preference, such as Asp, Asn and Cys (Doig and Baldwin, 1995), as conformations with these residues at the N-cap will be more populated. When the  $n$  value for an amino acid is large and of similar magnitude to the acetyl group ( $n = 5.9$ ), it will spend much of its time in a non-helical N-cap conformation. The large  $n$  values for Asp (6.6), Asn (6.8), Cys (5.4) and Tyr (4.9) tend to shift the equilibrium in favour for the residue X to be at N-cap. Thus, the helices will nucleate with X as the N-cap. In contrast, when X has a low N-cap preference, the acetyl will then be the N-cap, X will be at N3. The conformation when X is the N-cap will therefore be rarely populated and an error in  $n(X)$  will be small.

Residues of Glu, Gln and Asp that can form a hydrogen bond to the preceding N-cap seem to be more favoured at N3 in proteins than in the peptide. In the peptides the N-cap is an acetyl that lacks an NH group so this stabilising hydrogen bond as in the capping box cannot form. In principle, it is possible to design a series of peptides that could form this capping box bond to measure its contribution to helix formation. In practice, however, this is difficult. If the peptides had Gly at their N-terminus, they could be studied at high pH, where the N-terminus is  $H_2N-CH_2^-$ . These peptides would have low helix contents, particularly at their N-termini, as Gly is a poor N-cap, making N3 energies inaccurate. They would also need to be studied at high pH where the N-terminal amine is neutral. At high pH the peptides are likely to be insoluble as Lys side-chains would lose their positive charges. In addition, the titration of Tyr would interfere with helix content measurements.

In general, the negatively charged residues Asp, Glu, and Cys are helix stabilising when at N3. The stabilising contribution might be due to the ability of

these residues to form a hydrogen bond to one of unsatisfied amide groups at the N-terminus and the electrostatic interactions of the charged side-chains with the positively charged N-terminus. In contrast, the uncharged N3(His) is more stabilising in comparison to the fully protonated species.

The unperturbed pKa value for residue Glu at N3 indicates that the side-chain is not involved in any strong interactions. The Asp pKa is however shifted to a lower value in comparison to that in model compound indicating that Asp interacts more strongly with the helix dipole. This could be explained from the side-chain rotamer preference of Asp and Glu. The preferred rotamer for N3(Asp) is  $g^+$  (91%). N3(Glu) also favours  $g^+$  but the population is lower (74%) (Penel *et al.*, 1999). The high preference of  $g^+$  is simply because the  $C_\alpha$ - $C_\beta$  bond vector points toward the N-terminus. The  $\chi_1$  *t* is thus not favoured since there are steric clashes between non-hydrogen atoms on  $C_\gamma$  and the backbone CO group. The more constrained Asp side-chain could provide better interactions with the helix dipole as the  $\chi_2$  and  $\chi_3$  in Glu makes it more flexible, notably for interaction with solvent molecules. The Asp interaction with preceding N1 or N2 is, however, not as strong as Glu because of shorter side-chain (Penel *et al.*, 1999; Wan and Milner-White, 1999).

The lowered Cys pKa also suggests a similar stabilising interaction. Thiolates have been found to stabilise  $\alpha$ -helix by 0.3kcal/mol and 0.7kcal/mol when at the N1 and N2 positions, respectively (Miranda, 2003). The energetic of thiolate at N3 is predicted to be similarly stabilising. In this study, the energetic of having a thiolate in the N3 peptide was not calculated as the  $\Delta G$  cannot be measured. The stabilising effect is, however, qualitatively observed from pH titration curve.

Comparisons with *n*, *n1*, *n2* and *w* values show that N3 is unique. A similar environment, to some extent, gives similar preference of residues at N3 as in the interior. However, the forces involving residues at N3 are also similar to those at N1 and N2. Particular attention should be given to residues that can hydrogen-bond to the unsatisfied amide groups or give a charged-dipole effect (e.g. Glu, Asp or Lys) as these residues give deviations from general trend of N3 when compared to the helix interior environment.

Rotamer energy analysis for restricting N3 side-chains into different rotamers has given an additional important insight of protein folding. This is important as surface residues have their conformations restricted by the secondary structure to some extent, though can still retain considerable flexibility, particularly for long unbranched side-chains.

#### 4.17. Conclusions

The results on N3 preferences complete the work on studying the structural and energetic preferences of the amino acids for the N-terminal positions of the  $\alpha$ -helix. The substitution of amino acids at the N3 position in an AK peptide has shown the unique structural trends for N3 and yielded the range of intrinsic preferences for residues at this position in the  $\alpha$ -helix. Although the preferences at N3 correlate well to those at the helix interior, the N3 position favours negatively charged residues due to electrostatic interactions with helix dipole. A complementary rotamer energy analysis gives a deepened understanding to the structural uniqueness of the N3. The information can be used to rationally modify protein stability; particularly as helix N-termini are solvent-exposed and very rarely form tertiary interactions.



## CHAPTER 5

### THE CXXC MOTIF AT THE N-TERMINUS OF AN ISOLATED $\alpha$ -HELICAL PEPTIDE

#### 5.1. Introduction

An active site containing a CXXC motif is always found in the thiol-disulfide oxidoreductase superfamily. This superfamily includes thioltransferase, thioredoxin, glutaredoxin, and protein disulfide isomerase (PDI). In all of these proteins characterised structurally, the two Cys in the motif are at the N-cap and N3 positions of the helix N-terminus. Residues in these positions are in close proximity and can potentially interact with each other.

The CXXC motifs in various proteins exhibit differences in their physical behaviours. Depending on the functions of the proteins, the motif in the native folded proteins can be reduced or oxidised, i.e. forming a disulfide bond. Natively reduced CXXC in *E. coli* thioredoxin is highly reducing with  $E_o' = -0.270\text{V}$  (Moore *et al.*, 1964; Mossner *et al.*, 1999). In contrast, the motif acts as an oxidant when forming a disulfide bond. For example, DsbA with  $E_o' = -0.089\text{V}$ , is the most oxidizing protein known in the family (Wunderlich and Glockshuber, 1993). This difference in redox potentials can ensure the formation and transfer of disulfide bonds in newly produced proteins in the cell (Creighton *et al.*, 1995; Debarbieux and Beckwith, 1999). The  $\text{pK}_a$  values of the Cys residues, which are the determinant for the redox potential, also vary. The  $\text{pK}_a$  of the active site Cys residues can vary widely, depending upon the different X residues in the middle of the motif (Chivers *et al.*, 1997a; Mossner *et al.*, 2000) and on interactions with neighbouring amino acids close in space (Kortemme *et al.*, 1996; Chivers *et al.*, 1997b; Dillet *et al.*, 1998).

The regions in which the CXXC motif is found in the various members of the thiol-disulfide oxidoreductase family, known as the thioredoxin fold, exhibit similar structures (Martin, 1995). Despite the differences in the CXXC redox reactivity, there are minimal changes in structure between proteins upon breaking or formation

of the disulfide bond (Dyson *et al.*, 1990; Jeng *et al.*, 1994; Weichsel *et al.*, 1996). In general, the disulfide bonds stabilise folded proteins (for review see Wedemeyer *et al.* (2000)). Cleavage of the disulfide bond in the thioredoxin fold reduces the stability of native thioredoxin by 2.4kcal/mol (Kelley *et al.*, 1987). The disulfide bond of DsbA, however, destabilises the folded state of the protein by 4.5kcal/mol (Zapun *et al.*, 1993). These conflicting results show how the different CXXC motifs can have widely differing implications for the structural stability and catalytic properties of proteins containing it.

The CXXC sequence is also found within the Zn-finger motif which is important for protein-DNA recognition. Although the two cysteine ligands are not in helical segments, they stabilize the protein structure by coordinating a single zinc ion, together with another two His residues (reviewed by Pabo *et al.* (2001)). It is also part of the highly conserved motif within metallotheinins (Romero-Isart and Vasak, 2002) and in other metal-binding and transport proteins (Opella *et al.*, 2002; Romero-Isart and Vasak, 2002).

There have been very few studies on the CXXC motif in isolated peptides (Ookura *et al.*, 1995; Nguyen *et al.*, 2003) and no study has been made on the secondary structure effects of this motif in an isolated helical peptide. Nguyen & co-workers examined a 16-residue fragment (50-65) from macrophage migration inhibitory factor (MIF, a cytokine involved in the inflammation response). This fragment has a CXXC motif with redox potential  $E_o'$  of -0.258V and formed mixed disulfides with glutathione and cysteine. It also retains some biological functions of MIF. Ookura *et al.* (1995) showed that when a peptide (APWCGHCK - from the active site of protein disulfide isomerase) was present in a synthetic multimeric form attached on a peptide synthesis resin, it could mimic the parent enzyme activity. The activity diminished, however, when the peptide was in free monomeric form. A model  $\alpha$ -helical peptide having a disulfide bond between Cys residues  $i, i+7$  apart, also shows a stabilising effect (Jackson *et al.*, 1991).

In this study, a helical peptide containing the CXXC motif of the active site of thiol-disulfide oxidoreductases was designed and synthesised. The aim is to understand the secondary structural effects of the disulfide bond and metal binding,

independent of the tertiary structure and neighbouring amino acids. The energetics of disulfide bond formation and Zn-binding, were analysed using Lifson-Roig based helix-coil theory (Lifson, 1961).

## 5.2. Design of CXXC peptide

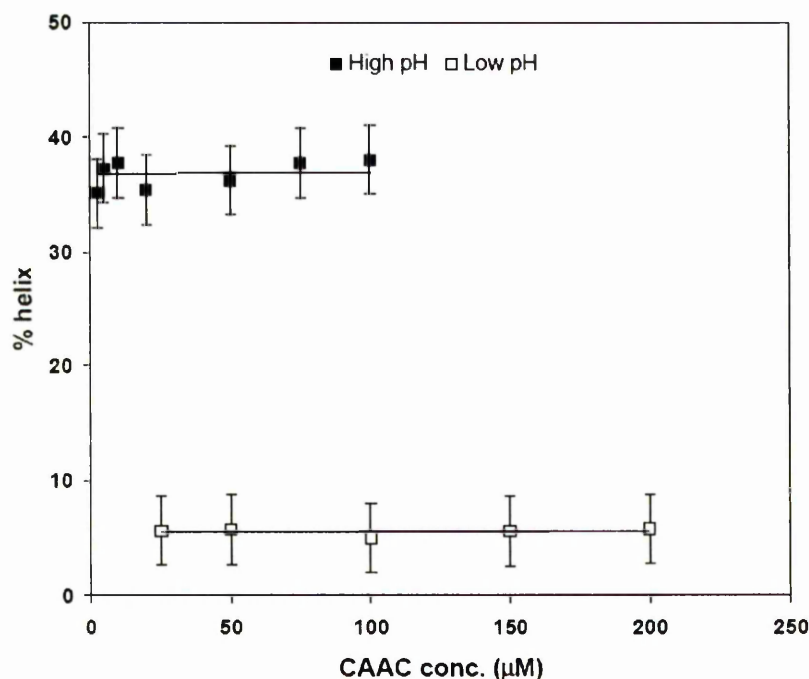
An intrinsically helical poly-Ala peptide with a sequence of CAACAAAAKAAAAGY-NH<sub>2</sub> was used to study the CXXC motif in an isolated  $\alpha$ -helix. This peptide is dubbed as CAAC peptide throughout this chapter. As the more N-terminal Cys is desired to act as the N-cap, the peptide N-terminus was not modified by acetylation. Cys is a favourable N-cap (Doig and Baldwin, 1995), whilst Ala is a poor N-cap, but relatively good at both N1 and N2 positions (Cochran and Doig, 2001; Cochran *et al.*, 2001). Cys, however, is not commonly found at either N1 or N2 in crystal structures (Doig *et al.*, 1997; Penel *et al.*, 1999). The  $\alpha$ -helix should therefore most often initiate with the first Cys as the N-cap, making the second Cys an N3 residue, although Cys is not heavily preferred in the N3 position in isolation (see results in Chapter 4). Helices that initiate with either of the first two Ala residues at N-cap, and hence the second Cys at N1 or N2, will be unfavourable and have low populations. The two Cys residues in the peptide are in close spatial proximity in this peptide allowing them to easily interact to each other.

The positively charged Lys residues, included for solubility, are spaced  $i, i+5$  to each other and to the second Cys so that no interactions between them can take place as they are on opposite faces of the helix. The Tyr residue is present to give the peptide a UV absorption for concentration determination. The penultimate Gly ensures that the  $\alpha$ -helix almost always terminates at, or before, this position so that problems arising from Tyr affecting the CD signal are minimized (Chakrabartty *et al.*, 1993b).

The N-termini of naturally occurring helices in proteins are very often solvent exposed (Doig *et al.*, 1997). The CXXC motif in thiol-disulfide oxidoreductases is also partially exposed to solvent (Chivers *et al.*, 1997a; Chivers *et al.*, 1997b; Dillet *et al.*, 1998). The CAAC peptide may therefore be a good model for the study of a CXXC motif that is isolated from neighbouring interactions.

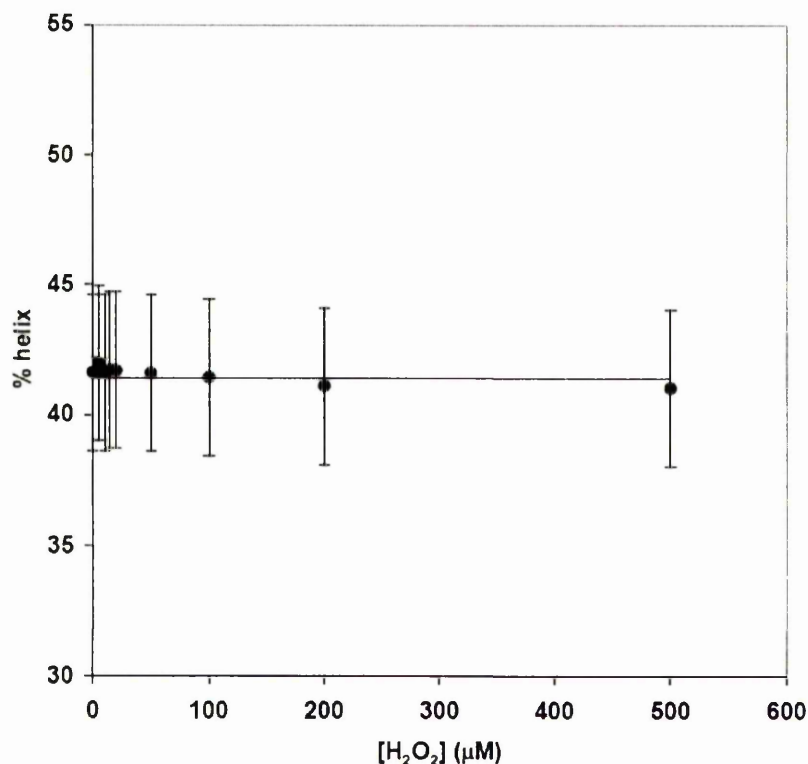
### 5.3. Aggregation and oxidations tests of the CAAC peptide

The CAAC peptide has a concentration independent CD signal at 222nm, showing no indication of aggregation (Figure 5.1). Aggregation tests were conducted on the CAAC peptide at high pH (5-100 $\mu$ M) and low pH (25-200 $\mu$ M).



**Figure 5.1.** Aggregation tests of the CAAC peptide. The CD ellipticities at 222 nm were measured at 273K in 5mM phosphate buffer containing 10mM NaCl. Filled and opened squares represent measurement at pH 7.2 and pH 4.0, respectively. Error is the CD measurement error of  $\pm 3\%$ . The horizontal lines are to relate helicity at different concentrations and have no physical meaning.

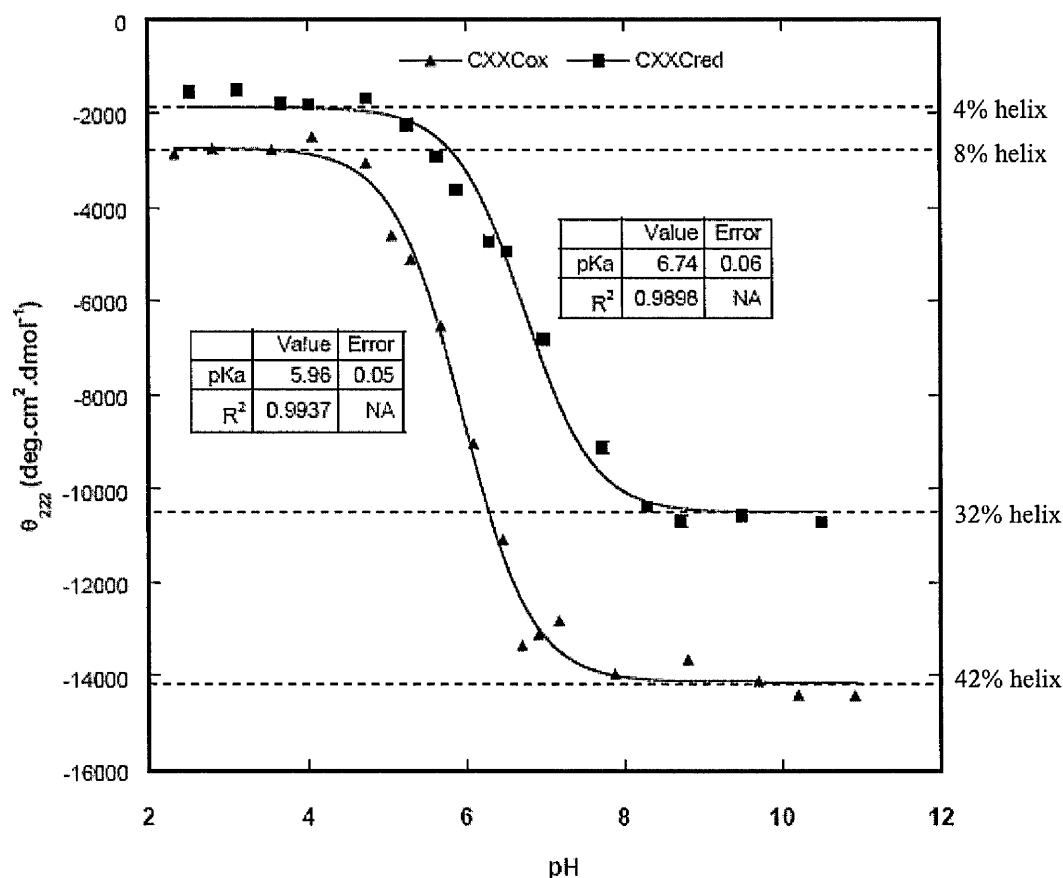
Oxygen-free buffers were always used in the experiments to control the formation of disulfide bonds. Prior to use, the buffer was purged with nitrogen for at least one hour. Data from an oxidation test using  $\text{H}_2\text{O}_2$  suggest that the disulfide bond from the two Cys residues is readily formed after dissolution of the pure, solid peptide. Addition of  $\text{H}_2\text{O}_2$  up to a concentration of 500 $\mu$ M to a 20 $\mu$ M peptide solution does not increase the helix content (Figure 5.2). This implies that the maximum level of disulfide bond formation is attained rapidly after solvation.



**Figure 5.2.** Oxidation tests on 20  $\mu$ M CAAC peptide by adding various concentrations of  $\text{H}_2\text{O}_2$ . The CD ellipticities at 222 nm were measured at 273K in 5mM phosphate buffer containing 10mM NaCl pH 8.0. Error is the CD measurement error of  $\pm 3\%$ . The horizontal lines is to relate helicity at different concentrations and has no physical meaning.

#### 5.4. pH titration of the CAAC peptide

The helicities of the the  $\text{CAAC}_{\text{ox}}$  and  $\text{CAAC}_{\text{red}}$  peptides were followed as a function of pH using far UV CD (Figure 5.3). The terms  $\text{CAAC}_{\text{ox}}$  and  $\text{CAAC}_{\text{red}}$  refer to the CAAC peptide with and without the disulfide bond, respectively.  $\text{pK}_a$  values of the N-terminus of the helix were evaluated between approximately pH 2-11 by curve fitting the CD data to the Henderson-Hasselbach equation for one apparent  $\text{pK}_a$  (see Chapter 3 equation 3.5). The upper range of accessible pH is given by Lys deprotonation, which leads to peptide aggregation.



**Figure 5.3.** pH titration of CAAC<sub>red</sub> (filled square) and CAAC<sub>ox</sub> (filled triangle). The CD ellipticities at 222 nm were measured at 273K in a buffer containing 1mM phosphate, 1mM borate, 1mM citrate and 10mM NaCl.

The titration curves demonstrate that the peptide with protonated Cys at low pH is less stabilising than when containing the negatively charged Cys at high pH. As the pH rises above the pK<sub>a</sub>, helix content increases maximally by 34% and 28% for CAAC<sub>ox</sub> and CAAC<sub>red</sub>, respectively. The helicity of CAAC<sub>ox</sub> increases by about 10% in comparison to CAAC<sub>red</sub> at high pH, suggesting that the disulfide bond formation stabilises the peptide.

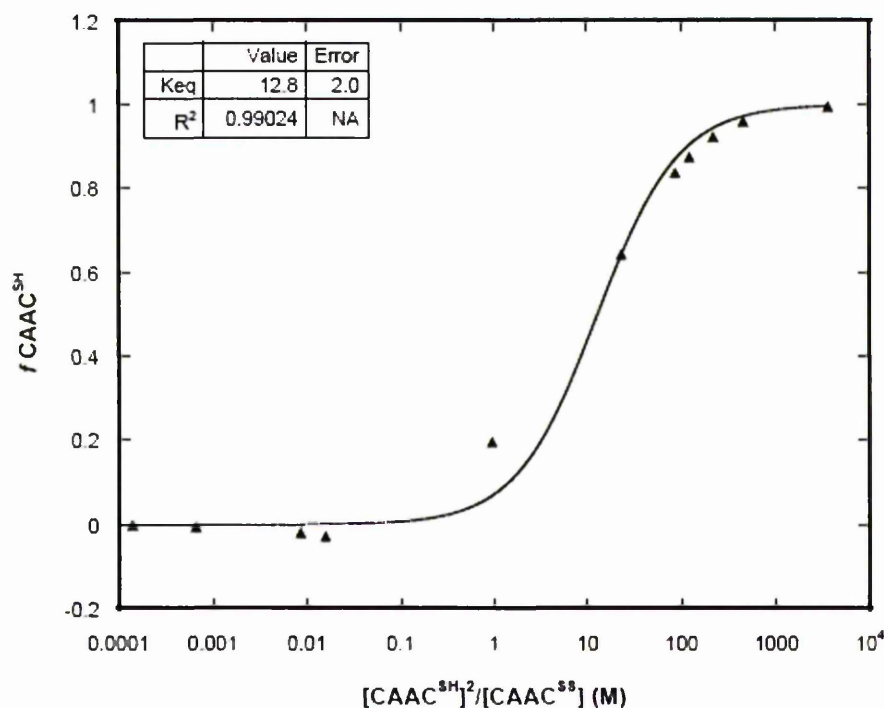
The apparent pK<sub>a</sub> values of CAAC<sub>ox</sub> and CAAC<sub>red</sub> are 5.96±0.05 and 6.74±0.06, respectively. The CAAC<sub>ox</sub> peptide has only the N-terminus amine to titrate, whereas the CAAC<sub>red</sub> peptide has the N-terminus amine and each of the Cys residues which are titrable. The pK<sub>a</sub> of the CAAC<sub>ox</sub> is thus due to the N-terminus amine, whereas the pK<sub>a</sub> for CAAC<sub>red</sub> has contributions from the N-terminus amine and each of the free Cys residues. The pK<sub>a</sub> of N-cap(Cys) amine in isolation has been previously

measured (Doig and Baldwin, 1995). The pK<sub>a</sub> is  $8.36 \pm 0.10$  which is much higher than that in CAAC<sub>ox</sub>. The reason for such difference is not clear. The absence of the adjacent CO<sub>2</sub><sup>-</sup> group that is present in the free amino acid, the nearby presence of free NH groups at N1, N2 and N3 and the positive end of the helix dipole may lower the pK<sub>a</sub>. The effect however seems profound for the N-cap(Cys) amine in CAAC<sub>ox</sub>.

The reported pK<sub>a</sub> values for the individual Cys side chain at N-cap and N3 in isolated  $\alpha$ -helical peptides are  $6.18 \pm 0.19$  (Doig and Baldwin, 1995) and  $7.73 \pm 0.12$  (Chapter 4 Figure 4.3D), respectively. These values differ significantly when compared to the pK<sub>a</sub> of Cys side chain in a coil peptide of  $8.69 \pm 0.07$  (Kortemme and Creighton, 1995) and are also higher than the value found for the CXXC<sub>red</sub> peptide. This is common in thioredoxins where the more N-terminal active-site Cys is generally a strong nucleophile with an abnormally low pK<sub>a</sub> value (Chivers *et al.*, 1997b). In contrast, the more C-terminal Cys is at least partially buried in the native proteins and little is known about its effective pK<sub>a</sub> during catalysis of disulfide exchange reactions (Chivers *et al.*, 1997b; Mossner *et al.*, 2000).

### 5.5. Redox properties of the CAAC peptide

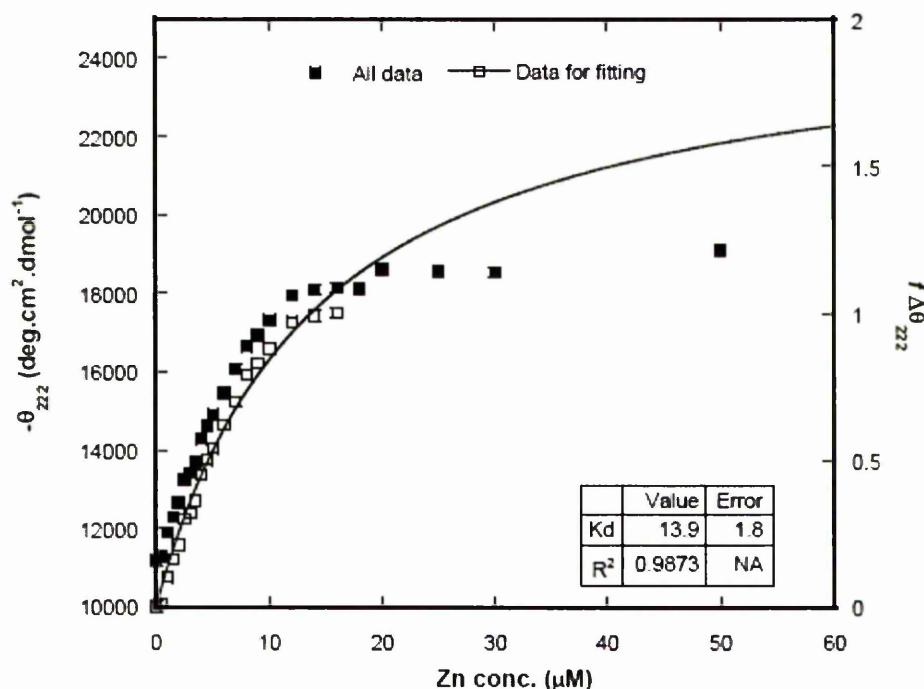
The redox potential of the CXXC motif was determined in the CAAC<sub>ox</sub> peptide by quantifying the steady state ratios of its reduced and oxidised forms in the presence of DTT. The redox potential of DTT is low and is capable of maintaining monothiols completely in the reduced state and of reducing disulfides quantitatively (Cleland, 1964). In addition, this compound is highly water-soluble, has little tendency to undergo air oxidation and gives negligible noise to CD spectra at low concentrations. The data in Figure 5.4 were analysed using non-linear least-squares analysis, as in equation 3.8 in Chapter 3. The term  $[\text{CAAC}^{\text{SH}}]^2/[\text{CAAC}^{\text{SS}}]$  relates to the presence of different states of CAAC as a result of DTT titration. The equilibrium constant ( $K_{\text{eq}}$ ) for the redox couple is  $12.8 \pm 2.0\text{M}$ . This is the point where the redox reaction between CAAC<sub>ox</sub> and DTT is in the equilibrium. Addition of DTT beyond this point will no longer increase the concentration of CAAC<sub>red</sub> in the solution.



**Figure 5.4.** Redox equilibrium between 20  $\mu$ M CAAC peptide and DTT at various concentrations. The CD ellipticities at 222 nm were measured at 273K in phosphate buffer containing 10mM NaCl pH 8.1. Fitting was done using equation 3.8 in Chapter 3.

The standard state redox potential  $E'_o$  of the CAAC peptide was calculated from the equilibrium constant of the redox reaction involving DTT using the Nernst equation (see Chapter 3 equation 3.9). The free energy of redox reaction was calculated as  $\Delta G = -RT \ln(12.8) = 6.1 \text{ kcal/mol}$ . The redox potential between  $\text{CAAC}_{\text{ox}}$  and DTT couple is related to the  $\Delta G$  as  $E'_{o_{\text{DTT}}} - (\Delta G/nF)$ , where  $E'_{o_{\text{DTT}}} = 0.366 \text{ V}$  at pH 8.1 (Cleland, 1964),  $F = 9.65 \times 10^4 \text{ C.mol}^{-1}$  and  $n = 2$ . This gives a value of  $E'_o = -235 \text{ mV}$ . The reported  $E'_o$  value of PDI-like thioredoxin is  $-235 \text{ mV}$  (Lundstrom and Holmgren, 1993) and that for human thioredoxin is  $-230 \text{ mV}$  (Watson *et al.*, 2003), suggesting that the CXXC motif in the CAAC peptide is reducing.





**Figure 5.5.** Zn titration of the CAAC peptide at 20  $\mu$ M. The CD ellipticities at 222 nm were measured at 273K in phosphate buffer containing 10mM NaCl and 400  $\mu$ M DTT pH 8.0.

### 5.6. Zinc titration on the CAAC peptide

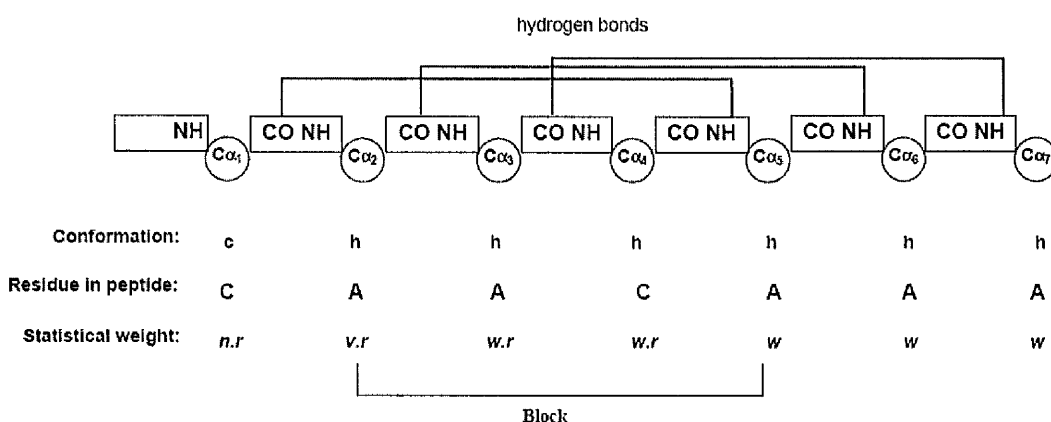
Equilibrium experiments of Zn binding to CAAC peptide were followed by the change in CD signal at 222 nm as a function of Zn concentration. The mean residual ellipticity of free peptide at 222 nm is approximately  $-11200 \text{ deg.cm}^2.\text{dmol}^{-1}$ . The addition of Zn caused a large, progressive, and saturable decrease in the ellipticity, to a value of approximately  $-18500 \text{ deg.cm}^2.\text{dmol}^{-1}$ . The numerical decrease in ellipticity shows that Zn binding is stabilising the helix in the CAAC peptide.

After the addition of Zn to 15  $\mu$ M, the change in  $\theta_{222}$  reaches a plateau (Figure 5.5. filled squares). This suggests that whatever interaction is occurring between the helical peptide and the zinc ion has no further effect on the helicity of the peptide above this point. This is supported by non-linear least-squares analysis of the data prior to this point (Figure 5.5. open squares), using equation 3.10 in Chapter 3. This yields an average binding constant ( $K_d$ ) of  $13.9 \pm 1.8 \mu\text{M}$ . Data above this point could

not be analysed in this way, as there is no further change in the observed signal. The Zn ion however may dimerise the peptide as well as stabilise the helix structure. This is due to the tetrahedral coordination sites of Cys residues can be satisfied by the Zn.

### 5.7 Lifson-Roig helix coil theory to calculate $\Delta G$ of interactions in CAAC peptide

In the CAAC peptide, the first Cys residue is not acetylated and thus not flanked by peptide units (CONH) on both sides. It is then considered as an N-cap residue and assigned a statistical weight  $n$ , if it assumed that the peptide is fully helical at the N-terminus. The following Ala residue is the nucleating residue and therefore has a statistical weight  $v$ . The remaining residues towards the C-terminus, consequently, are the propagating residues with statistical weights  $w$  (Figure 5.6).



**Figure 5.6.** Statistical weight assignments for residues in CAAC peptide.

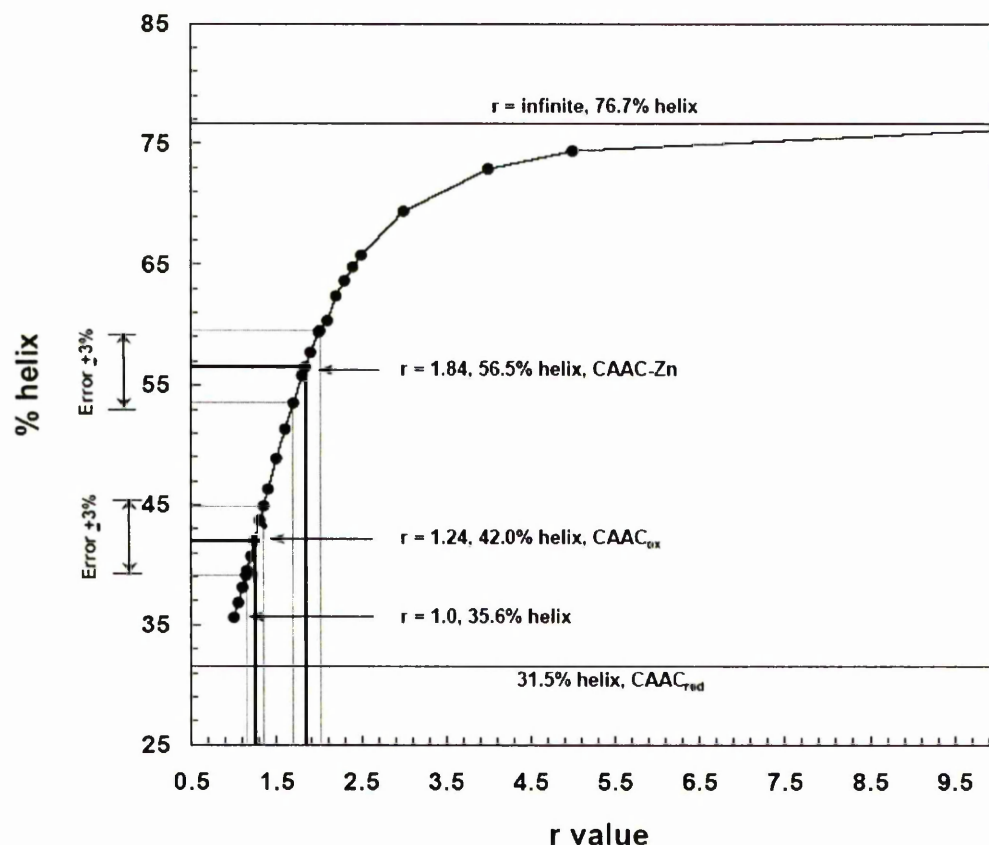
In modified Lifson-Roig-based models (Lifson, 1961) with side chain interactions (discussed in Chapter 2), a statistical weight  $q$  for an  $i, i+3$  interaction is assigned for the middle residue in the quintet  $hhhhc$ . Similarly a statistical weight  $p$  for an  $i, i+4$  interaction is assigned for the middle residue in the quintet  $hhhhh$ . The free energy of interactions are  $\Delta G_{(i,i+3)} = -RT \ln q$  and  $\Delta G_{(i,i+4)} = -RT \ln p$ , as  $q$  and  $p$  are the equilibrium constants of the corresponding interactions (Stapley *et al.*, 1995). In the CAAC peptide containing interactions involving the two Cys residues, the statistical weight  $q$  for the  $i, i+3$  is however not applicable. This is because the interacting Cys residues are not both in a helical conformation. When an interaction occurs involving the two Cys residues, the existing helix-coil theory that includes

side chain interactions thus needs additional considerations, although the basic algorithm remains unchanged.

The most significant change in the calculation is that the CAAC sequence is treated as a block. This approach has been utilised previously to calculate side chain interactions of residues in  $h$  conformations (Kise and Bowler, 2002). There are two reasons for doing this; firstly, this kind of interaction will involve both the interacting residues and those in between them, all of which will be affected by the interaction. Secondly, it has been already found that parameters of residues at the N-cap (Doig and Baldwin, 1995), N1 (Cochran *et al.*, 2001), N2 (Cochran and Doig, 2001) and N3 (Chapter 4) positions are interdependent. Parameter values, for example N3, could be perturbed by errors of parameters for N-cap, N1 or N2. Changing only one parameter, as a consequence of a side chain interaction, will shift the populations of helical conformations to start from one particular position. For example Cys has a high N-cap preference ( $n$  value) so if the  $n$  value is increased due to Cys-Cys interactions, the helices starting at either N-cap(Cys) or N3(Cys) will theoretically be roughly equally populated. This is because the  $n1$  and  $n2w$  values for Ala residues in the middle will be comparatively low and force the helix to initiate at N-cap(Cys). If this occurs, the error in determining the helix-coil parameters will be high. To overcome this, the other parameter values of N1(Ala) and N2(Ala) should also be increased by the same magnitude. In this way the short CAAC sequence can be regarded as a single block.

The  $i, i+3$  interaction between N-cap(Cys) and N3(Cys) in the block is assigned a statistical weight  $r$ . The consequence of treating the CAAC sequence as a block is that the parameter values for all residues in the block need to be multiplied by  $r$ . For example, N-cap(Cys), N1(Ala) and N2(Ala) and N3(Cys) will have conformational weights  $r.n_{cys}$ ,  $r.v_{ala}$ ,  $r.w_{ala}$  and  $r.w_{cys}$ , respectively. The interaction energies in the CAAC block can be calculated simply as  $\Delta G = -RT \ln K_{eq}$  where  $K_{eq} = r^4$ . The  $r^4$  term is present as there are four residues in the block, each being affected by the parameter  $r$ . The theoretical treatment of them as one block means that the  $r$  value determined for the block will be the product of  $r$  for each of the residues. The  $r$  values were found by varying them using the Scint2 program until the calculation agrees with the experimentally determined helix contents (Figure 5.7). The equation

$\Delta G = -RT \ln r^4$  applies both to calculate the stabilizing energy of disulfide bond formation and to dimerization during Zn binding (Table 5.1). The energy is arbitrarily relative to the reduced form of the helical CAAC. Please note that the helix propagation ( $w$ ) and N-cap ( $n$ ) values for Cys in all states (i.e. SH, S-S and S-Zn) used in the calculations are assumed to be those of Cys S-H as other values are not yet available ( $w = 0.32$  and  $n = 5.4$ ).



**Figure 5.7.** Fitting of  $r$  values of CAAC peptides of different states to agree with the experimental results. The errors in CD measurement are calculated as  $\pm 3\%$ .

When using  $r = 1$  in the calculation (i.e. there are no side-chain interactions), the predicted helicity is 35.6% for CAAC<sub>red</sub>. This result differs slightly from the experimental helicity of 31.5%, which could be due to experimental error in the CD measurement and/or concentration determination, which is usually of the order  $\pm 3\%$ . Another possibility is that the Scint2 program does not consider the destabilising effect of having Cys residue at N3 (Chapter 4). To test the maximum helicity that can

be achieved,  $r$  was set equal to infinity. This gives a predicted helicity of 76.7%. Calculating  $r$  for CAAC<sub>ox</sub> and CAAC<sub>Zn</sub> results in values of 1.24 and 1.84, respectively. These values give corrections to all parameter values of residues within the CAAC sequence. For example, the values of  $n(\text{Cys})$  at the N-cap increase from 5.40 to 6.70 and 9.94 in CAAC<sub>ox</sub> and CAAC<sub>Zn</sub>, respectively. Similarly, the  $w$  value of N1(Ala) changes from 1.70 to 2.11 and 3.13 in CAAC<sub>ox</sub> and CAAC<sub>Zn</sub>, respectively.

**Table 5.1.** Energetic calculations of CAAC peptides.

Peptide	% helix	$r$ value (range $\pm 3\%$ helicity) <sup>a</sup>	Free energy of interactions (kcal/mol)		
			$\Delta G_1^b$	$\Delta G_2^c$	$\Delta G_3^d$
CAAC <sub>red</sub>	35.6	1.00	1.30	0.00	1.30
CAAC <sub>ox</sub>	42.0	1.24 (1.14 to 1.35)		-0.47 (-0.28 to -0.65)	0.83 (1.01 to 0.64)
CAAC <sub>Zn</sub>	56.5	1.84 (1.70 - 2.01)		-1.32 (-1.15 to -1.52)	-0.03 (0.14 to -0.22)

<sup>a</sup> From fitting with Scint2 program. Parameter values of  $n_{\text{Cys}} = 5.40$ ,  $v_{\text{Ala}}^2 = 0.0013$ ,  $w_{\text{Ala}} = 1.70$  and  $w_{\text{Cys}} = 0.32$  were from Rohl *et al.* (1996).

<sup>b</sup> Calculated as  $\Delta G = -RT \ln [(n_{\text{Cys}} * v_{\text{Ala}} * w_{\text{Ala}} * w_{\text{Cys}}) / (1+v)^4]$ .

<sup>c</sup> Calculated as  $\Delta G = -RT \ln r^4$ .

<sup>d</sup> Calculated as  $\Delta G = -RT \ln r^4 * [(n_{\text{Cys}} * v_{\text{Ala}} * w_{\text{Ala}} * w_{\text{Cys}}) / (1+v)^4]$ .

$\Delta G_1$  is the free energy for transferring the CAAC block from coil to helix, where the Cys residues are in the reduced form. Multiplication of all equilibrium constants of residues in the block gives  $\Delta G$  of 1.30kcal/mol calculated as  $-RT \ln [(n_{\text{Cys}} * v_{\text{Ala}} * w_{\text{Ala}} * w_{\text{Cys}}) / (1+v)^4]$ . The term  $(1+v)$  is the approximation of weight of a random coil residue (Qian and Schellman, 1992). The equilibrium constant for transferring a residue to an existing helical segment is the ratio of the weight of the residue when it is part of the helix to that when it is part of the random coil. Although N-cap is considered as coil, it is unique structurally. Hence, the  $n$  value is also divided by  $(1+v)$ .

$\Delta G_2$  is the free energy of interaction, calculated as  $-RT \ln r^4$ . Using  $r = 1.24$ , the disulfide bond in CAAC<sub>ox</sub> stabilises the helix by 0.47kcal/mol ( $\Delta G$  -0.28 to

-0.65kcal/mol). A significant stabilising effect is shown when the CAAC peptide is saturated with Zn with  $r = 1.84$ , giving a stabilising energy of -1.32kcal/mol ( $\Delta G$  -1.15 to -1.52kcal/mol). However, the energy calculation of CAAC<sub>Zn</sub> might contain error as the peptide dimerises when saturated with Zn.  $\Delta G_3$  is calculated as  $\Delta G_1 - \Delta G_2$ , giving the overall free energy for transferring the CAAC block from coil to helix in the presence of the interactions.

### 5.8. The capping box and CXXC motif

The capping box is a well-known stabilising motif at the helix N-terminus involving a reciprocal hydrogen bond from the N-cap position (mostly Ser and Thr) with polar residues at the N3 position (mostly Glu) (Harper and Rose, 1993; Wan and Milner-White, 1999). The possibility that CXXC is a new variant of capping box was explored through a survey of protein crystal structures (Table 5.2).

**Table 5.2.** Some selected N-cap-N3 pair propensities in helix N-terminus

Pair (X <sub>N-cap</sub> -Y <sub>N3</sub> )	Number found*			Probability X at N-cap	Probability Y at N3	Expected N-cap-N3	Propensity N-cap-N3
	XY pair	X at N-cap	Y at N3				
CC	9	42	39	0.01±0.001	0.01±0.001	0.34±0.11	26.3±8.2
TE	160	527	689	0.11±0.005	0.14±0.005	76.0±5.8	2.1±0.2
SE	158	671	689	0.14±0.005	0.14±0.005	96.8±6.9	1.6±0.1
TD	82	527	426	0.11±0.005	0.09±0.004	47.0±4.1	1.7±0.2
SD	103	671	426	0.14±0.005	0.09±0.004	59.8±4.9	1.7±0.1

\*Data were from domain search of 1135 non-redundant proteins containing 4778 helices in ASTRAL database (SCOP 1.63 Sequence Resources).

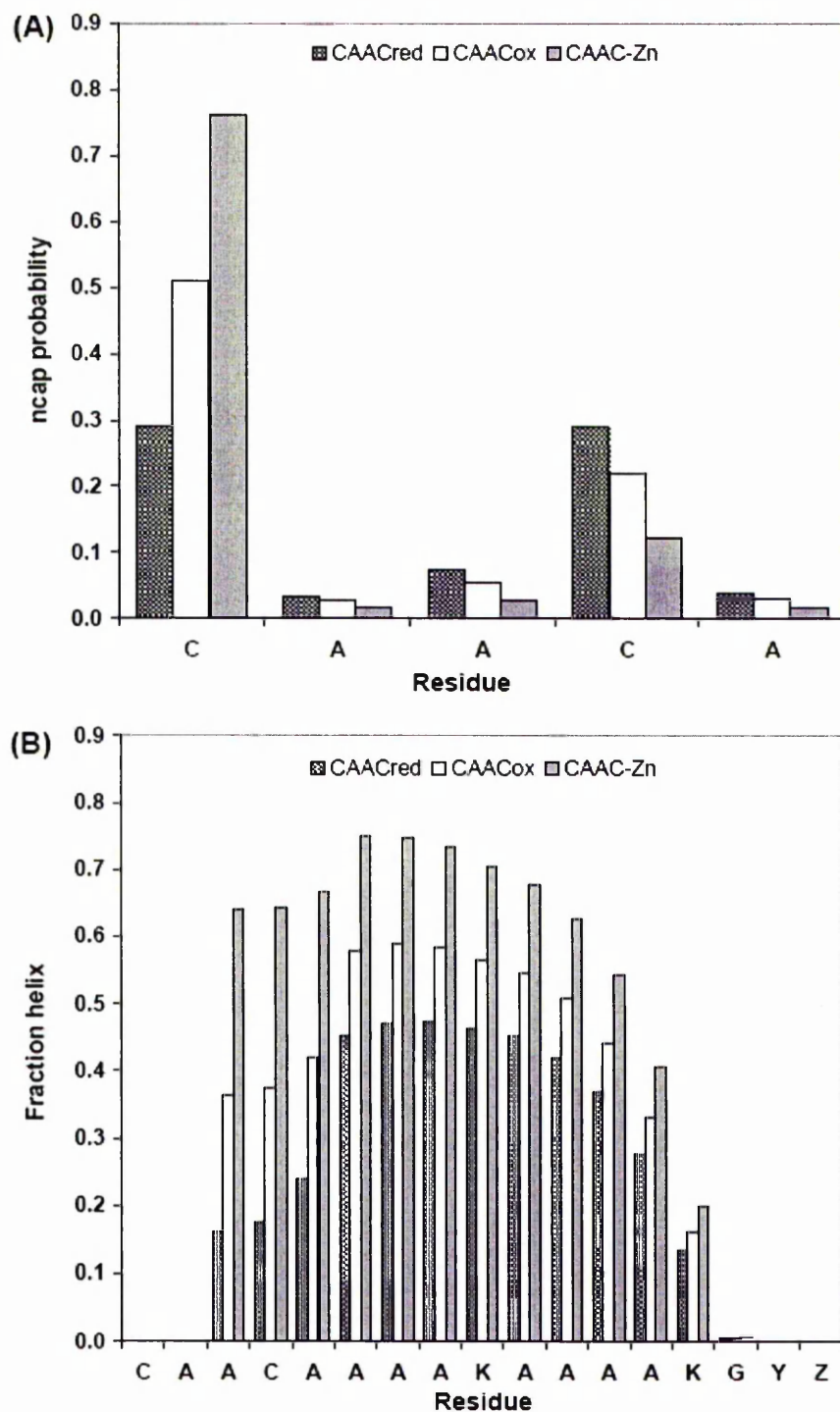
Propensity is used to normalise frequency so that a tendency of a particular amino acid to be at certain positions can be observed. The propensities showed here are 'local', that is the ratio between the amino acid frequency found in a defined position or pair of positions, i.e. N-cap and N3, and the amino acid frequency found

only in helices. Local propensities are used as the CXXC motif is found in the helix N-terminus. The global propensity is thus not necessary in this case. The global propensity would be the frequency at the defined positions in the helix compared to that in all protein secondary structures. If the frequency of an amino acid found in a certain position is different to that expected from a random distribution, the amino acid is then either preferred (propensity  $>1$ ) or not preferred (propensity  $<1$ ) in that position (Penel *et al.*, 1999).

Among the other 399 combinations (data not shown), CC pair at N-cap and N3 has a propensity of  $26.3 \pm 8.2$ , which is exceptionally high. Cys is one of the rarest residues found in proteins including in helices (Penel *et al.*, 1999). This lessens the chance of finding the CC pair in helices, making uncertainty in propensity calculations large. Interactions between Thr/Ser and Asp/Glu are also preferred although the propensities are not high and the errors are also significantly lower due to the increased frequencies of these residues. This intriguing result demonstrates that the CXXC motif could be an alternative sequence for a capping box since it is not only statistically preferred in proteins, but also it is helix stabilising experimentally.

### 5.9. N-cap and helix probability of CAAC peptide

The fraction of the helical conformation differs along the peptide, being greatest in the middle and less at the termini. This phenomenon, termed helix-fraying, has been proven experimentally (Chakrabartty *et al.*, 1991; Liff *et al.*, 1991). Cys residues at the end of the CAAC peptide where fraying is the greatest, will have considerable conformational flexibility. The presence of interactions involving the Cys residues would reduce fraying, thus increase helix stability. This effect can be predicted by using the Scint2 program by comparing the helix content on a residue-by-residue basis in the presence and absence of an interaction. The parameter values used for prediction were taken from Table 3.1 in Chapter 3. The corresponding parameter values for residues in the CAAC block in different states were adjusted using the  $r$  values in Table 5.1. The N-cap probabilities and helix profiles for CAAC<sub>red</sub>, CAAC<sub>ox</sub> and CAAC<sub>Zn</sub> are shown in Figure 5.8.



**Figure 5.8.** (A) N-cap probability (B) Predicted helicity as a function of as a function of positions of residues in CAAC peptide. Prediction used the Scint2 program using the  $r$  values in table 5.1. Other values ( $n$ ,  $w$  and  $v$ ) were taken from Table 3.1.

The probability of having shorter helices decreases when the interactions are present. The N-cap(Cys) probability shows a significant increase in CAAC<sub>ox</sub> and



CAAC<sub>Zn</sub> compared to CAAC<sub>red</sub>. The N-cap probability in CAAC<sub>ox</sub> shows an increase of about 40% compared to that in CAAC<sub>red</sub>. When Zinc is bound, the largest increase in N-cap probability is observed, being more than twice that in the CAAC<sub>red</sub> peptide (Figure 5.9A). At the same time, The N-cap probability of all other residues in the N-terminal region, i.e. N1(Ala), N2(Ala) and N3(Cys), show a decrease. A helix with the first Cys residue as N-cap(Cys) would therefore more highly populated as this would be the most favourable conformation. The presence of interactions involving the Cys residues at N-cap and N3 would also significantly reduce helix fraying at the N-terminus, relative to the predicted helix profile in the absence of an interaction.

A favourable side chain interaction stabilises the helical conformation not only on the interacting residues, but also the intervening residues. A favourable side-chain interaction enhances helix formation along the entire length of the peptide, although the stabilisation decreases with increasing distance from the site of interaction (figure 5.9B). The largest increase in helix content occurs at the peptide centre, with smaller effects at the helix termini. Similar mechanisms occur in the presence of capping boxes, since these reduce helix fraying thus decreasing the initiation and elongation energy and stabilizing the helices.

## 5.10. Discussion

The CXXC motif in an isolated  $\alpha$ -helical peptide is stabilising when the two Cys residues form a disulfide bond or bind to Zn. The formation of the intramolecular disulfide bond is rapid after solvation. Addition of an oxidising agent (H<sub>2</sub>O<sub>2</sub>) at up to 25 times peptide concentration does not increase the helicity indicating that the maximum level of disulfide bond is achieved immediately. The intermolecular disulfide bond, leading to dimerisation of the peptide, was also formed although the amount is negligible. This is proved by the mass spectra of pure CAAC peptide that give a small intensity at *m/z* 2808 (see Appendix A.21). As CD measurements were made in an equilibrium state, this should not have contributed to the results obtained.

The pK<sub>a</sub> of Cys in CAAC<sub>red</sub> is lower than that of free Cys. Possible reasons for perturbation include (i) interaction between thiolate and the positively charged helix macrodipole at the N-terminus (ii) hydrogen bonding between the thiolate anion and

unsatisfied amide groups at the N-terminus (iii) A reciprocal hydrogen bond between the two Cys residues. However, these three interactions could not be experimentally separated. Figure 5.3 shows only one observable transition across the pH range studied, implying that the  $pK_a$  values of the two Cys residues are similar. Previous results in CXXC motifs in proteins show that the  $pK_a$  of Cys at N-cap (7.45) is slightly lower than that at N3 (7.51) (Chivers *et al.*, 1997b; Mossner *et al.*, 2000). Mossner *et al.* (2000) show depression of the N-terminal Cys  $pK_a$  and elevation of the more C-terminal Cys  $pK_a$ . This complex behaviour might be due to (i) interaction of proximal residues to both Cys (ii) the N3(Cys) is buried and thus experiencing tertiary interactions. (iii) N-cap (Cys) is more easily made anionic at increased pH making deprotonation of N3(Cys) more difficult. (iv) a helix macrodipole interaction with N-cap(Cys) is stronger than that with N3(Cys). In contrast, although (iii) and (iv) continue to apply, the two Cys residues in the CAAC peptides are both solvent exposed and in isolation from other interactions. The titration behaviour is thus considerably less complex than in native proteins.

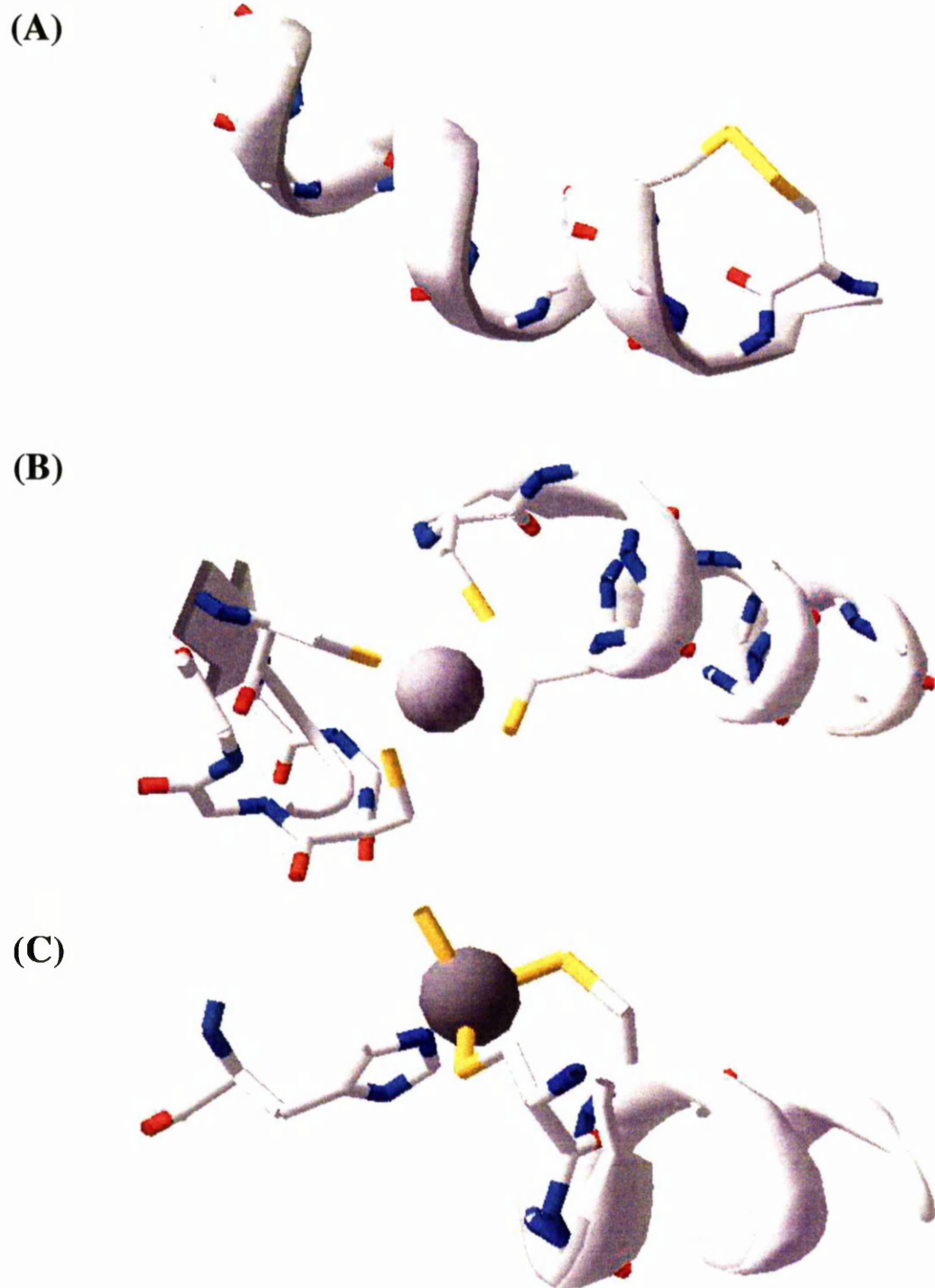
The standard state redox potential  $E'_o$  of the CAAC peptide calculated from the equilibrium constant of the redox reaction with DTT gives a value of  $E'_o = -235\text{mV}$ , indicating that the CXXC motif in the CAAC peptide is reducing. This value is very similar to the reported  $E'_o$  value of PDI-like thioredoxin of  $-235\text{mV}$  (Lundstrom and Holmgren, 1993) and that for human thioredoxin of  $-230\text{mV}$  (Watson *et al.*, 2003). The correspondence between these values suggests that the perturbation of the redox potential of CXXC in thioredoxin can be entirely attributed to the helix.

There is a correlation between the  $pK_a$  and the redox potential in that the lower the Cys  $pK_a$ , the more oxidising the motif (Grauschopf *et al.*, 1995). The reactive Cys in WT DsbA (CPHC) with  $E'_o = -120\text{mV}$  shows a  $pK_a$  value of 3.3. In highly reducing Trx-type DsbA (CGPC) with  $E'_o = -210\text{mV}$  the  $pK_a$  is 6.2 (Inaba and Ito, 2002). The X residues in CXXC are suggested to influence this, but not to be the sole determinant. Proximal charged amino acids are also a governing factor for the reactivity of the Cys residues (Kortemme *et al.*, 1996; Chivers *et al.*, 1997b; Dillet *et al.*, 1998). Direct comparisons between other reported  $E'_o$  values and that in the

CAAC peptide, however, are not possible because local electrostatic interactions with proximal amino acids are missing in the peptide. Fraying at the N-terminus of the helix may also affect  $E_o'$  values in the CAAC peptide.

The dimerisation of the CAAC peptide when titrated with  $\text{Zn}^{2+}$  is anticipated, as  $\text{Zn}^{2+}$  prefers a tetrahedral geometry (Dudev and Lim, 2000) and prefers 'soft' Cys ligands in protein binding (Dudev and Lim, 2003). The change in  $\theta_{222}$  reaches a plateau after the addition of more than 15  $\mu\text{M}$  Zn to 20  $\mu\text{M}$  peptide. Why the apparent saturation event occurs at this concentration is not clear, although it may be because the dimerisation is likely to proceed by successive bimolecular reactions. A single Zn molecule may bind to one molecule of CAAC in the first step, followed by binding of the second CAAC molecule in the second step. In the early stages, water molecules might occupy half of the tetrahedral coordination sites. When reaching saturation, there could be a change in the ligand-metal affinity. The second CAAC molecule could then replace the water molecules, leading to dimerisation. This would not change  $[\theta]_{222}$ , as there is no change in peptide helicity involved.

Helix coil analysis reveals that the presence of disulfide bond and Zn binding reduces the helix fraying at the N-terminus CAAC peptide. These favourable interactions at the helix N-terminus also enhance helix formation along the entire length of the peptide. The PDB was then searched to find examples of disulfide bond and Zn-binding in the CXXC motif in proteins. Figures 5.9A shows the presence of a disulfide bond in the CXXC motif. The bond is formed between N-cap(Cys<sub>210</sub>) and N3(Cys<sub>213</sub>) of PDB-1GAI. Figures 5.9B shows that N-cap(Cys<sub>222</sub>) and N3(Cys<sub>225</sub>) of PDB-1VFY provide half of the tetrahedral coordination for Zn binding. The other half is provided, interestingly, by Cys192 and Cys195 of a CRSC sequence in a bend in the protein. Figures 5.9C shows that N-cap(Cys<sub>87</sub>) and N3(Cys<sub>90</sub>) of PDB-1GPC provide half of the tetrahedral coordination. The other half could be provided by a water molecule and His64 which is in proximity to the Zn.



**Figure 5.9.** Structural examples of the CXXC motif. (A) Disulfide loop across the first turn of the helix in PDB-1GAI. (B) CC pair in helix pair serves partly as ligand for Zn binding in PDB-1VFY (C) CC pair in helix serves as ligand for Zn binding in PDB-1GPC.

In proteins, Zn binding sites vary. His, Cys, Asp, Glu and water molecule can occupy the Zn tetrahedral coordination (Alberts *et al.*, 1998). The occurrence of CAAC dimerisation is thus to be expected. At low concentration, half of the Zn tetrahedral coordination sites may attach to Cys while the other half are occupied by water molecules. At high concentration, all Zn coordinates may be occupied by the Cys side-chains, leading to dimerisation. It is, however, not clear whether Zn binding to the CXXC motif in the given protein examples is as a result of crystallisation procedures used for protein isolation. However, the structures show that the CXXC motif can form disulfide bonds as well as provide coordination for Zn binding.

### 5.11. Conclusions

An isolated helical peptide with a CXXC motif at the N-terminus has been used in this study as a model for the thioredoxin motif and metal binding motifs. The CXXC motif in the peptide stabilises the  $\alpha$ -helix and has similar values of redox potential to those found in natural thioredoxins. The  $pK_a$  values of the individual Cys residues are perturbed, showing the presence of strong interactions. Analysis of CD results with modified helix-coil theory indicates that stabilisation of the  $\alpha$ -helical structure is achieved by both disulfide bond formation ( $\Delta G$  -0.28 to -0.65 kcal/mol) and Zn binding ( $\Delta G$  -1.15 to -1.52 kcal/mol). The peptide binds with moderate affinity to zinc ions in solution, leading to dimerisation of the peptide, probably through the tetrahedral geometry of the zinc ion being fulfilled by four Cys side chains. This interaction also leads to stabilisation of the  $\alpha$ -helical structure. The peptide is a reasonable mimic of several motifs in which it can be found naturally, e.g. the thioredoxin motif and various metal-binding motifs. The results from this study may therefore have implications in the areas of peptide and protein design and as a probe to study the thiol-disulfide exchange mechanism.

## CHAPTER 6

### EFFECT OF PAIRWISE COUPLING IN AN ARG-PHE-MET TRIPLET IN AN ISOLATED $\alpha$ -HELICAL PEPTIDE

#### 6.1. Introduction

In proteins, side-chains interact within a complex network of multiple non-covalent bonds, which may reinforce or weaken each other. The simplest system to investigate these effects is to study triplets of side-chains. This has previously been studied as salt bridge interactions in  $\alpha$ -helical peptides. A triplet of charged Arg-Glu-Arg residues spaced  $i,i+4$  or  $i,i+3$  stabilises  $\alpha$ -helical peptides by more than the additive contribution of two single salt bridges (Olson *et al.*, 2001b). Other triplets have also been studied, for example Glu-Phe-Arg (Olson *et al.*, 2001a) and Glu-Phe-Glu (Shi *et al.*, 2002a), although they do not show significant effects on peptide stability.

The  $\Delta G$  of interaction between Arg and Phe spaced  $i,i+4$  in isolated  $\alpha$ -helix peptides has been previously measured. The  $\Delta G$  of Arg-Phe (RF) interaction is  $-0.1\text{kcal/mol}$  (Andrew *et al.*, 2002a). The  $\Delta G$  of Phe-Met (FM) in an  $i,i+4$  spacing has also been measured and it is stabilising as much as  $-0.75\text{kcal/mol}$  (Stapley *et al.*, 1995). The Serrano group has also measured the FM interaction, giving slightly lower  $\Delta G$  of  $-0.65\text{kcal/mol}$  (Viguera and Serrano, 1995b). It is thus interesting to study the effects of these interactions on helix stability when present simultaneously in and  $i,i+4,i+8$  spacing.

It is predicted that the triplet is cooperative via the shared Phe, as Phe in helix predominantly adopts  $\chi_1$  *t* conformation (Penel *et al.*, 1999). The Phe also adopts  $\chi_1$  *t* conformation when in the RF and FM pairs (see below). It can provide its planar ring as the interface for interactions. In this chapter, the coupling effects of pairwise interactions were investigated. The pairwise coupling is defined as cooperative when the free energy of the pairs present simultaneously is greater than the sum of the individual pairs. Conversely, when the free energy of the simultaneous pairs is less

than the sum of the individual parts, the interaction is called anti-cooperative. A series of peptides were used to study the stabilising effect of the pairwise coupling. As the results may be rationalised by shared rotamer preference, crystal structure surveys were conducted and  $\chi_1$  rotamers for the interacting residues were analysed.

## 6.2. Design of RFM peptides

A series of intrinsically helical peptides containing the interacting pairs were designed to measure the energetic contribution of the triplet. The full sequence of the peptides is listed in Table 6.1.

**Table 6.1.** Peptides used to study coupling effect in RFM triplet.

Peptide	Sequence	Predicted helicity <sup>a</sup>	Predicted helicity <sup>b</sup>	Exptl. Helicity <sup>c</sup>
R5F5M	Ac-AAKRAAAAFAAAAMKGY-NH <sub>2</sub>	30%	24%	24%
R4F5M	Ac-AAKARAAAFAAAAMKGY-NH <sub>2</sub>	33%	27%	32%
R5F4M	Ac-AAKRAAAAFAAAMAKGY-NH <sub>2</sub>	47%	40%	36%
R4F4M	Ac-AAKARAAAFAAAMAKGY-NH <sub>2</sub>	50%	44%	63%

<sup>a</sup>Predicted using the Scint2 program. Values used for side-chain  $i, i+4$  interactions ( $p$  values) for RF and FM were 1.2 (Andrew *et al.*, 2002a) and 4.0 (Stapley *et al.*, 1995), respectively. Other values of  $n$ ,  $v$ ,  $w$  and  $c$  were taken from (Rohl *et al.*, 1996). <sup>b</sup>Helicity calculated as in <sup>a</sup> but with  $w$  values multiplied by 0.967. <sup>c</sup>Calculated from CD ellipticities  $[\theta]_{222}$  as  $([\theta]_{222} \times 100)/(-40000 \times (1-2.5/N))$

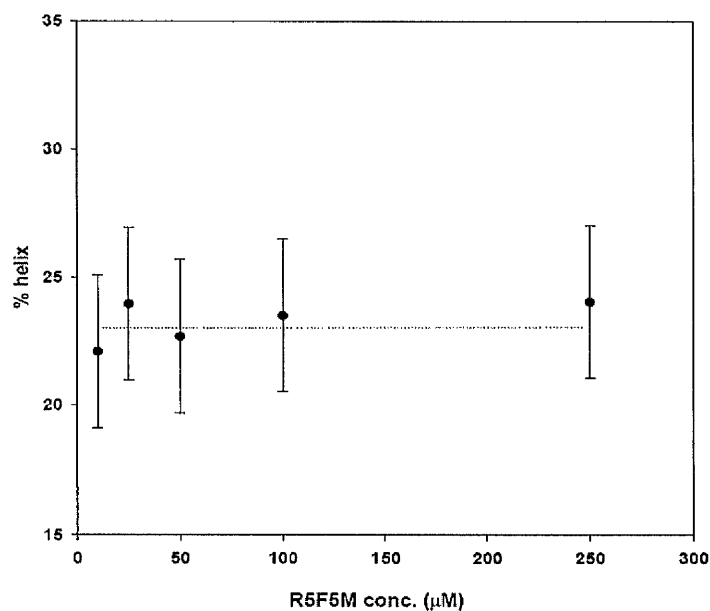
The analysis of cooperativity in RFM peptides is based on the control sequence Ac-AAKRAAAAFAAAAMKGY-NH<sub>2</sub>, utilising the high helix propensity of Ala. This control sequence is altered by swapping Arg at N4 with Ala at N5 to give an RF  $i, i+4$  interaction. Similarly, FM  $i, i+4$  interaction is achieved by interchanging Met at C5 with Ala at C4. The RFM interaction occurs when the three residues are in  $i, i+4, i+8$  spacing.

Two Lys residues are included for solubility and placed so that they cannot interact with residues of interest. Gly is the helix breaker. Tyr was introduced for concentration determination. As it may perturb the circular dichroism signal (Chakrabartty *et al.*, 1993b), Gly was thus placed between Tyr and the rest of the sequence. The N- and C-termini are blocked with an acetyl group and an amide group, respectively. This prevents unnecessary interactions of residues with the

charged termini and the helix dipole. This also decreases the number of free main chain amine and carboxyl groups, by creating an extra hydrogen bond at each terminus. Acetyl groups also substantially stabilize the helix (Doig and Baldwin, 1995).

### 6.3. Aggregation tests of R5F5M peptide

It is imperative to have peptides that are soluble in studying CD-derived helicity. Solubility is normally achieved by introducing several Lys (for a charged peptide) or Gln (for a neutral peptide) residues along the sequence. The 17-mer RF-Met peptides contain two Lys residues. Ala/Lys-based sequences are known to be monomeric in aqueous solution (Padmanabhan *et al.*, 1990). There is a possibility, however, that the peptides are insoluble as they contain Phe and Met with long non polar side-chains. Aggregation tests prove that this does not occur as the R5F5M peptide is monomeric up to 250 $\mu$ M (Figure 6.1). Other peptides should also be soluble because there are no charged residues introduced for solubilisation involve in the interactions.



**Figure 6.1.** Aggregation test of the R5F5M peptide. Helix content was converted from CD ellipticity of peptides measured at in 5mM phosphate buffer containing 10mM NaCl pH 7.4 at 273K. Error is CD measurement error of  $\pm 3\%$ . The dashed line is to relate helicity at different concentrations and have no physical meaning.



#### 6.4. Helicity of RFM peptides

The changes contributed from pairwise interactions are readily evaluated by measurements of CD intensity at 222 nm, which is proportional to the average helicity of the peptide in solution. Helicities were measured on 20  $\mu$ M peptides. The  $p$  value for RF  $i, i+4$  side-chain interaction (see Chapter 2) used for helicity prediction was taken from Andrew *et al.* (2002). The  $p$  value for FM side-chain interaction was taken from Stapley and Doig (1995). The value of Stapley and Doig is preferred in the initial prediction than the value of Viguera and Serrano (1995) for several reasons: (1) the side-chain interaction in Stapley and Doig's peptide is more isolated than that of Viguera and Serrano. The energetic of FM of Viguera and Serrano was obtained from peptide with a sequence of YGGSAAEAFKAMAR. There is a possibility that the Glu interacts with Lys  $i, i+4$  apart, giving inaccurate CD measurement. Moreover, the authors used peptides containing Phe-Ala and Ala-Met pairs in the FM position as the controls. This makes helicity dependent on the change in  $w$  value of Ala, Phe and Met; (2) the helicity range of Stapley and Doig's peptides is wider than that of Viguera and Serrano, relative to the control peptides.

The experimental helicity of the control R5F5M is about 24% (Table 6.1). Since the RF and FM interactions are stabilising, higher ellipticities for other peptides were expected. The highest helicity is observed on the peptide containing simultaneous RF and FM interactions (63% helicity). This wide range of ellipticity is thus sensitive to the pairwise coupling energy.

A helicity prediction of the control R5F5M using the original  $w$  values of Rohl *et al.* (1996) gives predicted helicity of about 6% higher than the experimental result. The discrepancy might be due to the fact that Scint2 does not take into account the preference of Lys at the N3 position. Lys at N3 is destabilising (see Chapter 4) due to an unfavourable interaction with the helix dipole at the N-terminus. Helicity disagreement in all peptides could also be due to changes in the ratio of  $\alpha$  to  $3_{10}$  conformations among Alanine-rich peptides that might affect CD-derived helicity. All  $w$  values of residues in the sequence (e.g. Ala, Lys, Arg, Phe, Met, Gly and Tyr) were then multiplied by a factor of 0.967 to correct the problem (Table 6.2). This correction also reduces predicted helicities for other peptides.

**Table 6.2.** Parameter values used for RFM peptides prediction using the Scint2 program. Original values were taken from Rohl *et al.* (1996).

Residue	$n^a$	$v^2{}^b$	$c^a$	$w^a$	$w*0.967$
A	1.0	0.0013	1.0	1.7	1.644
F	2.06	0.0013	1.0	0.27	0.261
G	3.9	0.0013	1.0	0.048	0.046
K	0.72	0.0013	1.0	1.0	0.967
M	1.31	0.0013	1.0	0.65	0.629
R	1.0	0.0013	1.0	1.14	1.102
Y	4.9	0.0013	1.0	0.48	0.464

<sup>a</sup> $n$ ,  $c$  and  $w$  are equilibrium constants for transferring a residue from coil to the corresponding N-cap, C-cap and internal positions. <sup>b</sup> $v^2$  is the nucleation constant. A complete list of parameter values for all amino acids are listed in table 3.1

There are disagreements observed between predicted and experimental helicity in R5F4M and R4F5M. It could be due to error in the  $p$  values used in the prediction (see below).

### 6.5. Refitting of $p$ values for RF and FM interactions

When an  $i,i+4$  interaction occurs between the residues at both ends of a helical stretch  $hhhhh$ , the residue in the middle of the quintet is mathematically assigned weight  $p$  which is simply the equilibrium constants of the interaction (see Chapter 2). It can then be used to calculate the free energy of interaction as  $\Delta G_{(i,i+4)} = -RT \ln p$  (Stapley *et al.*, 1995).

To agree with experimental helicity, the  $p$  values for RF and FM interaction were refitted. The  $p$  values for each pair were fitted individually by varying them until the calculated helix contents agreed with experiment. Refitting  $p$  values for RF and FM gives values of 1.70 and 2.94, respectively, equivalent to -0.29kcal/mol and -0.59kcal/mol. This gives an additive  $\Delta G$  of -0.88kcal/mol (Table 6.3).

The previous  $p$  value used to calculate helicity of the FM interaction was obtained from the peptides Ac-YGFAKAMAAKAAAKAA-NH<sub>2</sub>, Ac-YGAAKAAFAKAMAAKAA-NH<sub>2</sub> and Ac-YGAAKAAAKAAFAKAM-NH<sub>2</sub> (Stapley *et al.*, 1995). There may be additional favourable  $i,i+3$  side-chain interactions between Met and Lys residues in the first two peptides. There could also be an additional  $i,i+3$  side-chain interaction between Lys and Phe in the third

peptide. This would give an overestimate in the  $p$  value of the FM interaction. A sequence of Ac-AAAKARAAAFARAAAFKAGY-NH<sub>2</sub> was used to calculate the previous  $p$  value for the RF interaction. The peptide contains two RF pairs. It is not clear why the value is lower than that found in this study.

The predicted helicity for R4F4M is 44% using the refitted  $p$  values, which is much lower than the experimental result of 63% (Table 6.1). This large difference proves qualitatively that there is a large stabilizing effect as a result of pairwise coupling in the R4F4M triplet.

**Table 6.3.** Refitted  $p$  values and energetics of interaction in RFM peptides.

Interaction	Refitted $p$ value ( $\pm 3\%$ helicity error)	$\Delta G$ (kcal/mol) <sup>a</sup>	Multiplication factor (mf) ( $\pm 3\%$ helicity error)	Refitted $p$ values $\times$ (mf) in triplet ( $\pm 3\%$ helicity error)	$\Delta G$ when in triplet <sup>b</sup> (kcal/mol)	$\Delta G$ cooperativity $-RT\ln(mf)^2$ (kcal/mol)
RF	1.70 (1.39 to 2.06)	-0.29 (-0.18 to -0.39)		3.38 (2.96 to 3.93)	-0.66 (-0.59 to -0.74)	
FM	2.94 (2.40 to 3.56)	-0.59 (-0.47 to -0.69)		5.85 (5.12 to 6.79)	-0.98 (-0.89 to -1.04)	
RF-Met		-0.88 <sup>c</sup> (-0.65 to -1.08)	1.99 (1.74 to 2.31)		-1.62 <sup>c</sup> (-1.47 to -1.78)	-0.75 (0.60 to 0.91)

<sup>a</sup>Calculated using refitted  $p$  values.

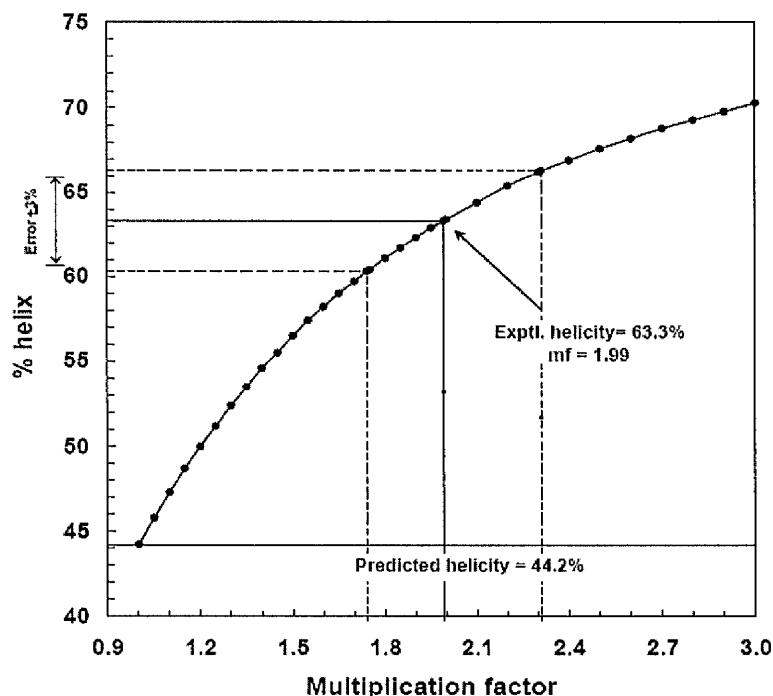
<sup>b</sup>Calculated using refitted  $p$  values $\times$  (mf).

<sup>c</sup>Additive  $\Delta G$  of the pairs (kcal/mol).

## 6.6. Adjustment of $p$ values for RF and FM interactions in R4F4M peptide

To get the predicted R4F4M helicity to agree with the experimental result, the  $p$  value of each pair was then increased by a multiplication factor (mf) of 1.99 (figure 6.2). This increases the statistical weight of all conformations with 9 consecutive helical residues from Arg to Met simultaneously by  $(mf)^2$ .  $p$  values for RF and FM in R4F4M peptide become 3.38 and 5.85, respectively, energetically equivalent to -0.96kcal/mol and -0.66kcal/mol. This calculation is based on the assumption that mf can be applied equally to RF and FM pairs. The error in  $p$  was calculated by

repeating the fitting procedure using experimental helicities increased or decreased by 3% to account for the experimental error in the measurement of helicity.



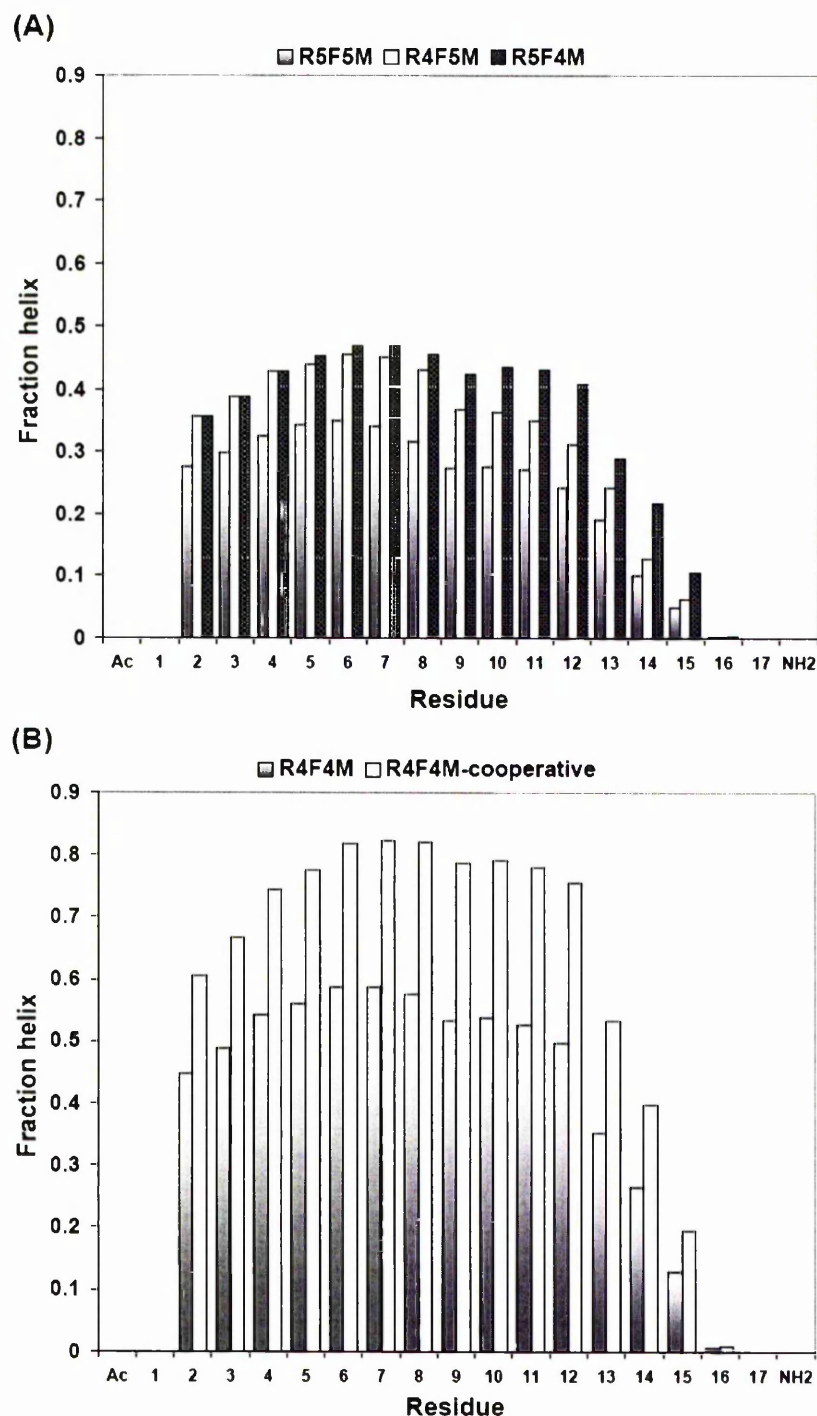
**Figures 6.2.** Adjustment of  $p$  values for the RF and FM pairs when present simultaneously in the R4F4M triplet to get predicted helicity to agree with experiment

$\Delta G$  for pairwise coupling was calculated as  $-RT \ln (1.99^2) = -0.75$  kcal/mol, reflecting a strong stabilising effect in R4F4M. The energy is nearly as much as the additive energy of the individual pairs ( $-0.87$  kcal/mol).

## 6.7. Helix probability analysis of RFM peptides

The Scint2 program is capable to predict helix contents of peptides that have  $i, i+3$  or  $i, i+4$  side-chain interactions within a helix, provided the experimental helicities are known. Helix probability in RFM peptides can then be used to analyse helix stability on a residue basis in the presence or absence of interactions. Helix-coil transition theory predicts the fraction of the helical conformation to differ along the peptide, being greatest in the middle and less at the termini, due to the effect of helix

fraying. The presence of side-chain interactions could reduce this effect when occurring at the helix termini.



**Figure 6.3.** Predicted helicity as a function of position in RFM peptides using the Scint2 program (A) In R5F5M, R4F5M and R5F4M using  $p_{RF}$  and  $p_{FM}$  of 1.70 and 2.94, respectively; (B) In R4F4M. Prediction of R4F4M-cooperative used refitted  $p$  values $\times$ (mf) with  $p_{RF}$  and  $p_{FM}$  of 3.38 and 5.85, respectively.

The presence of RF and FM interactions increase peptide helicity, relative to the E5K5E control. The effect of having RF and FM is similar in the helix N-terminus region. The FM interaction increases the helix content significantly in the C-terminal region, in where the FM interaction is positioned (Figure 6.3A). The presence of RF and FM simultaneously increases the helix probability dramatically along the entire length of peptide (Figure 6.3B).

### 6.8. Rotamer distribution of Arg, Phe, and Met residues

Table 6.4 shows the results of  $\chi_1$  rotamer analysis of Arg, Phe and Met in protein  $\alpha$ -helices. This PDB derived rotamer analysis, however, is based on the assumption that the rotamer distributions determined in the presence of tertiary contacts in proteins are applicable to peptides.

The  $\chi_1$  rotamer preference of Arg seems unchanged when in the RF pair. The population only varies slightly. Similar results also observed for the Met rotamer population when in the FM pair. However, the Met  $\chi_1 g^+$  population in the FM pair is higher (0.81) than the individual Met  $\chi_1 g^+$  in helices (0.67). Phe in the helix favours the  $t$  conformation (0.60). In both the RF and FM pairs, Phe also tends to be in the  $t$  conformation. This suggests that the two pairs incorporated in a RFM triplet could be stabilising. The Phe residue will be locked into the  $t$  conformation when in the middle of the triplet. RF and FM interactions will therefore not both have to pay the entropic cost of restricting the Phe to  $t$ . The stabilizing free energy resulting from the RFM triplet formation is indeed greater than the sum of the individual RF and FM energies as shown in Table 6.3.

**Table 6.4.**  $\chi_1$  Rotamer populations of Arg, Phe and Met in  $\alpha$ -helix.

Residue in helix	$\chi_1$ rotamer populations <sup>a</sup>			Total found in $\alpha$ -helices
	$g^+$	$t$	$g^-$	
Arg	$0.51 \pm 0.01^b$	$0.45 \pm 0.01$	$0.04 \pm 0.003$	5370
Phe	$0.38 \pm 0.01$	$0.60 \pm 0.01$	$0.02 \pm 0.003$	4953
Met	$0.67 \pm 0.01$	$0.31 \pm 0.01$	$0.02 \pm 0.01$	2166
Arg in RF	$0.46 \pm 0.06$	$0.44 \pm 0.06$	$0.10 \pm 0.04$	71
Phe in RF	$0.34 \pm 0.06$	$0.65 \pm 0.06$	$0.01 \pm 0.01$	71
Phe in FM	$0.16 \pm 0.04$	$0.81 \pm 0.05$	$0.03 \pm 0.02$	69
Met in FM	$0.81 \pm 0.05$	$0.19 \pm 0.05$	0	69

<sup>a</sup>Data from domain search of 1135 non-redundant proteins containing 4778 helices in ASTRAL database (SCOP 1.63 Sequence Resources). Rotamer definitions:  $-120^\circ < \chi_1 < 0^\circ = g^+$ ;  $0^\circ < \chi_1 < 120^\circ = g^-$ ;  $-120^\circ < \chi_1 < 240^\circ = t$ . <sup>b</sup>Errors were calculated as  $\sigma_i = [p_i(1-p_i)/N_i]^{1/2}$ , where  $p_i$  is the rotamer populations and  $N_i$  is the total number of pair found in helices.

## 6.8. Discussion

The differences in helicity of RFM peptides in Table 6.1 arise from the presence of side-chain interactions. Moving the Arg and the Met only gives little contribution to the overall stability and helix-coil theory takes account of this effect. For example, helicity prediction of a series of peptides below with Ala replacing Phe shows only a small helicity difference for all sequence (50-53% helicity).

Ac-AAKRAAAA-A-AAAAMKGY-NH<sub>2</sub> (helicity 52.6%)

Ac-AAKARAAA-A-AAAAMKGY-NH<sub>2</sub> (helicity 52.5%)

Ac-AAKRAAAA-A-AAAMAKGY-NH<sub>2</sub> (helicity 49.7%)

Ac-AAKARAAA-A-AAAMAKGY-NH<sub>2</sub> (helicity 49.6%)

One could argue that for a canonical helix with a periodicity of 3.6 residues per turn, two turns of the helix would more closely approximate  $i+7$  than  $i+8$ . So the  $i, i+4, i+7$  or  $i, i+3, i+7$  interactions is more suitable to measure the effect of pairwise coupling. However, the  $i, i+3$  energies are generally weaker. Interactions effect on such spacing would not be so clear, so were not measured.

The  $i, i+4$  RF and FM interactions are stabilising. When both pairs are incorporated in a triplet spaced  $i, i+4, i+8$ , the stabilising effect is greater than the sum of the free energy of the pairs, showing a cooperative effect. The observed cooperativity in R4F4M is suggested as a result of the Phe side-chain favouring the  $t$  conformation in both RF and FM pairs. Cooperativity thus occurs because the entropic cost of fixing the side-chain into a particular rotamer is paid only once, while the benefit of this conformation may be recovered twice, as it forms two side-chain interactions.

The conformational entropy cost of restricting a residue into a helix is given by  $\Delta S = -R \cdot \sum p_i (\ln p_i)$ , where  $p_i$  are the populations of the three  $\chi_1$  side-chain rotamers. From the dataset in table 6.4, the conformational entropies are 0.38, 0.60 and 0.02 for the  $g^+$ ,  $t$  and  $g^-$  rotamers respectively for Phe residue, giving  $-T\Delta S$  of 0.4kcal/mol. Conformational entropy can thus account for approximately half the free energy of cooperativity. The remainder of 0.35kcal/mol may be strain energy.

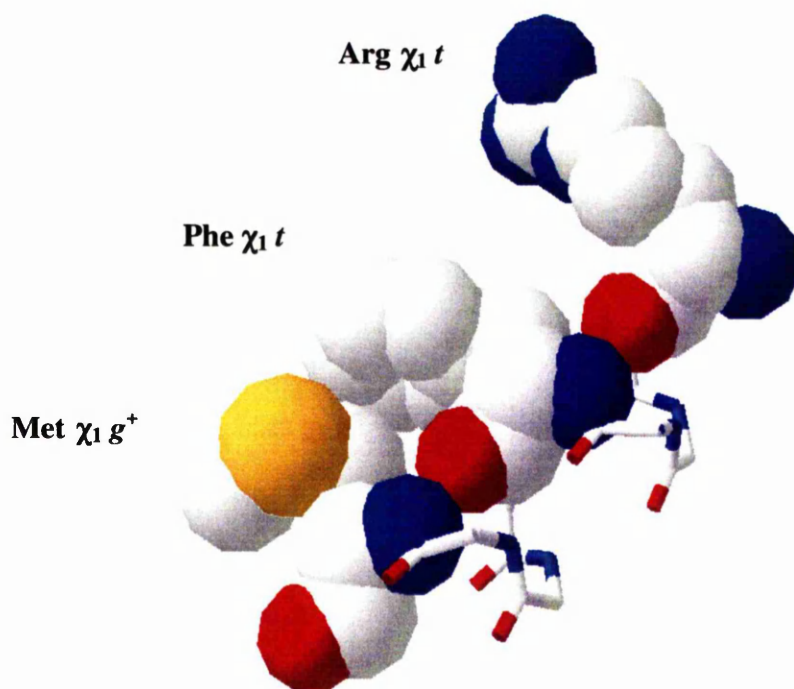
The rotameric strain energy for Phe  $\chi_1$  in an isolated poly-Ala helix AAAAAAXAAAAA has previously been determined computationally (Penel and Doig, 2001). Rotameric strain energy is the energy difference between a particular rotamer and the rotamer with the lowest energy. For example, Phe  $\chi_1$  favors the  $g^+$  conformation in isolated poly-Ala,  $g^+$  is thus the reference for  $t$  and  $g^-$ . This gives the Phe  $\chi_1$   $t$  strain energy of 0.61kcal/mol, relative to  $g^+$ . In proteins, the rotamer preferences are affected by many factors such as tertiary contacts. The most populated rotamer in proteins is thus not necessarily in the lowest energy state.

The rotamer energy is derived from rotamer populations grouped in a broad range of rotamer angles. For example, the rotamer energy of Phe  $t$  is calculated from the rotamer population in the range of  $120^\circ$  to  $240^\circ$ . It does not consider the free energy minimum of Phe  $\chi_1$   $t$  that is actually being adopted. The rotamer energy is thus also a function of  $\chi_1$  rotamer angles. A complete solution on this problem has been presented by Penel and Doig (2001), giving the mean  $\chi_1$  strain Phe  $t$  of 0.32kcal/mol. This also may contribute to the cooperativity effect in the RFM triplet. Loss of conformational entropy in  $\chi_2$  of Phe may also contribute to the coupling free



energy. The Phe  $\chi_2$  angles are defined by a sp<sup>3</sup>-sp<sup>2</sup> bond. This rotamer distribution is not discrete (Penel *et al.*, 1999). Propensity calculation for these angles is not possible because no rotamer definitions can be made.

The crystal structures from the PDB were searched for RFM triplets in helices utilising the Short Sequence Pattern facility in the PDB queries form. Only structures obtained with X-ray crystallography method of resolution 2.5Å or better with less than 50% homology were considered, giving 7 examples. The number observed is small probably because Phe and Met are rare in the helix interior (3.9% and 2.7% of the total residues, respectively). The Phe residue is in the  $\chi_1$  *t* conformation in all 7 structures. Structural analysis of the RFM triplet indicates that the dominant interaction is hydrophobic for both pairs (Figure 6.4), not cation- $\pi$  as previously proposed by Shi *et al.* (2002b), as it is the non-polar parts of the side-chains that are in contact. The example given in Figure 6.4, however, can not be used to predict the free energy of interactions.



**Figure 6.4.** Structural example of RFM triplet in  $\alpha$ -helix in pdb-1lw3.

## 6.9. Conclusions

Stabilising pairwise coupling is a result of the R4F4M peptide only needing to pay the energetic cost of restricting the Phe residue into the *t* conformation once in the triplet, rather than twice when the interactions are separate. The energetic benefit may consist of conformational entropy, rotameric strain energy and  $\chi_1$  strain energy. Rotamer and  $\chi$  strain is of similar magnitude to conformational entropy in opposing protein folding.

It is thus proposed that the shared rotamer preference effect that rationalises the results in this study is a general property of proteins, giving cooperativity in folding and increasing protein stability. Any bond which restricts a side-chain into a conformation that is favourable for forming additional interactions will show this effect. The effect is substantial, as fixing just a single side-chain into its preferred conformation is here worth 0.75kcal/mol.

## CHAPTER 7

### EFFECT OF PAIRWISE COUPLING IN A GLU-LYS-GLU TRIPLET IN AN ISOLATED $\alpha$ -HELICAL PEPTIDE

#### 7.1. Introduction

Salt bridge interactions are found in about 60% of protein structures. One-third of the charged amino acids participating in the salt-bridges are found as complex salt bridges (three or more amino acids) where triplet salt bridges are the most abundant. Approximately 20% of the total salt bridges found in proteins resides in  $\alpha$ -helix secondary structure in either  $i, i+3$  or  $i, i+4$  spacing (Musafia *et al.*, 1995).

The energetic role of salt bridges in protein structure stability has been found to vary. Solvent-exposed salt bridges contribute only marginally to protein stability (Horovitz *et al.*, 1990; Daopin *et al.*, 1991; Spek *et al.*, 1998). In contrast, a buried salt bridge stabilises the native state by up to 5kcal/mol (Anderson *et al.*, 1990; Tissot *et al.*, 1996). Theoretical studies of solvation suggest that internal or external salt bridges contribute only marginally to protein stability due to the high-energy penalty of desolvating ions within the interior or near the surface of a protein, thus diminishing the expected gain in strength of ionic interactions (Hendsch and Tidor, 1994).

Many attempts have been made to evaluate the strength of multiple salt bridges, particularly in proteins. A coupling interaction of an Asp8-Arg110-Asp12 triplet on the surface of barnase stabilises the triplet by 0.8kcal/mol relative to the sum of isolated pairs (Horovitz *et al.*, 1990). Salt-bridge-hydrogen bonds of Asp14-Ser77 and Arg17-Ser77 pairs in an Asp-Ser-Arg triplet in the  $\lambda$  repressor are stabilising by 1.5 and 0.8kcal/mol, respectively (Marqusee and Sauer, 1994). Kallenbach and co-workers reported that a GCN4 leucine zipper is stabilised by 1.7kcal/mol on introducing a triplet of Arg-Glu-Arg at the helix surface (Spek *et al.*, 1998). The presence of many additional charged side-chains close to the sites of interest, however, complicates the generalisation of the energetic role of salt bridge networks

in proteins. For this reason, the study of salt bridge networks in a more isolated environment, i.e. in short peptides, is becoming important.

The first evaluation of the strength of an engineered complex salt bridge in a peptide was reported by Mayne *et al.* (1998) after studying a multiple salt bridge involving Glu3, Asp4, and Arg7 in an 11-mer  $\alpha$ -helix. The triplet contributes substantially to the  $\alpha$ -helix with  $\Delta G = -1.2$  kcal/mol. A triplet of charged Arg-Glu-Arg residues spaced  $i, i+4, i+8$  or  $i, i+3, i+6$  also stabilises  $\alpha$ -helical peptides by -1.5 kcal/mol and -1.0 kcal/mol, respectively, which is more than the additive contribution of two single salt bridges (Olson *et al.*, 2001b). A similar stabilising effect in an Arg-Phe-Met triplet in  $i, i+4, i+8$  spacing is also observed, although the triplet is not a complex salt bridge (Chapter 6). Other non-salt bridge triplets in isolated helical peptides have also been reported, for example Glu-Phe-Arg (Shi *et al.*, 2002b) and Glu-Phe-Glu (Shi *et al.*, 2002a), although they do not show significant effects on peptide stability.

Most triplet salt-bridges studied so far were found to be cooperative, i.e. the free energy of the pairs present simultaneously is greater than the sum of the individual pairs. It is thus interesting to study a triplet salt-bridge in an isolated  $\alpha$ -helix having the opposite effect. In this study, we examine Glu-Lys-Glu (EKE) interactions when present simultaneously in an  $i, i+4, i+8$  spacing. The free energy of Glu-Lys (EK) and Lys-Glu (KE) interactions spaced  $i, i+4$  in isolated  $\alpha$ -helix peptides have been previously been found to be stabilising by about -0.4 kcal/mol in both orientations (Smith and Scholtz, 1998). In a EKE triplet, the central Lys side-chain cannot point towards both Glu side-chains simultaneously. The EKE triplet may therefore be less stabilising than the sum of the two individual interactions, giving rise to anti-cooperativity.

A series of helical peptide were thus synthesised to study the anti-cooperativity in the EKE triplet. As it has been demonstrated in Chapter 6 that cooperativity in RFM triplet is attributed to the shared Phe *trans* rotamer preference, the  $\chi_1$  rotamer preference for the interacting residues in EK-Glu were also searched. The rotamer preferences will be analysed to help explain the energetics of the EKE triplet.

## 7.2. Design of EKE peptides

A series of intrinsically helical alanine-based peptides with different spacing of the interacting residues was designed to allow the quantitative evaluation of the role of EKE triplets on helix stability (Table 7.1). E5K5E is the control peptide in the EKE series. The EK  $i,i+4$  interaction was achieved by swapping Ala at N5 with Glu at N4. Similarly, the KE  $i,i+4$  interaction was achieved by changing AEAK to EAKA. An Ala residue was placed in between Glu and Lys to avoid a close-range charge effect. The  $i,i+2$  spacing ensures the residues are on the opposite face of the helix, separating the charges. Shorter helices that could be made by moving Glu towards the N-terminus were avoided because Glu has unique preferences to be at N1 (Cochran *et al.*, 2001), N2 (Cochran and Doig, 2001) or N3 (Chapter 4). The helicity would be perturbed by Glu interacting with free amide NH groups at N1, N2 and N3, as well as with the helix macro dipole at the N-terminus.

**Table 7.1.** Peptides used to measure EKE triplet interactions.

Peptide	Sequence	Predicted helicity (%)		Experimental helicity (%)	
		Parameter values of E <sup>-</sup> &K <sup>+</sup> *	Parameter values of E <sup>o</sup> &K <sup>+</sup> *	pH 8.2	pH 2.9
E5K5E	Ac-AAAEAAAAKAAAAEAKGY-NH <sub>2</sub>	50	57	51	56
E5K4E	Ac-AAAEAAAAKAAAEAKAGY-NH <sub>2</sub>	56	59	67	65
E4K5E	Ac-AAAAEAAAKAAAAEAKGY-NH <sub>2</sub>	59	61	63	63
E4K4E	Ac-AAAAEAAAKAAAEAKAGY-NH <sub>2</sub>	64	63	60	61

\*Predicted using the Scint2 program. Parameter values used for side-chain  $i,i+4$  interactions ( $p$  values) for EK<sup>+</sup> and K<sup>+</sup>E<sup>-</sup> were 2.15 and 2.09, respectively (Smith and Scholtz, 1998). The  $p$  values for E<sup>o</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>o</sup> were 1.50 and 1.45, respectively (Smith and Scholtz, 1998). Other values for  $n$ ,  $v$ ,  $w$  and  $c$  were taken from Rohl *et al.* (1996).

Gly is the helix breaker and was placed between Tyr and the rest of the sequence to prevent Tyr perturbation of the circular dichroism signal (Chakrabartty *et al.*, 1993b). Tyr was included to allow concentration determination. The N- and C-termini were blocked with an acetyl group and an amide group, respectively, minimising the destabilising interactions that may occur with the charged termini.

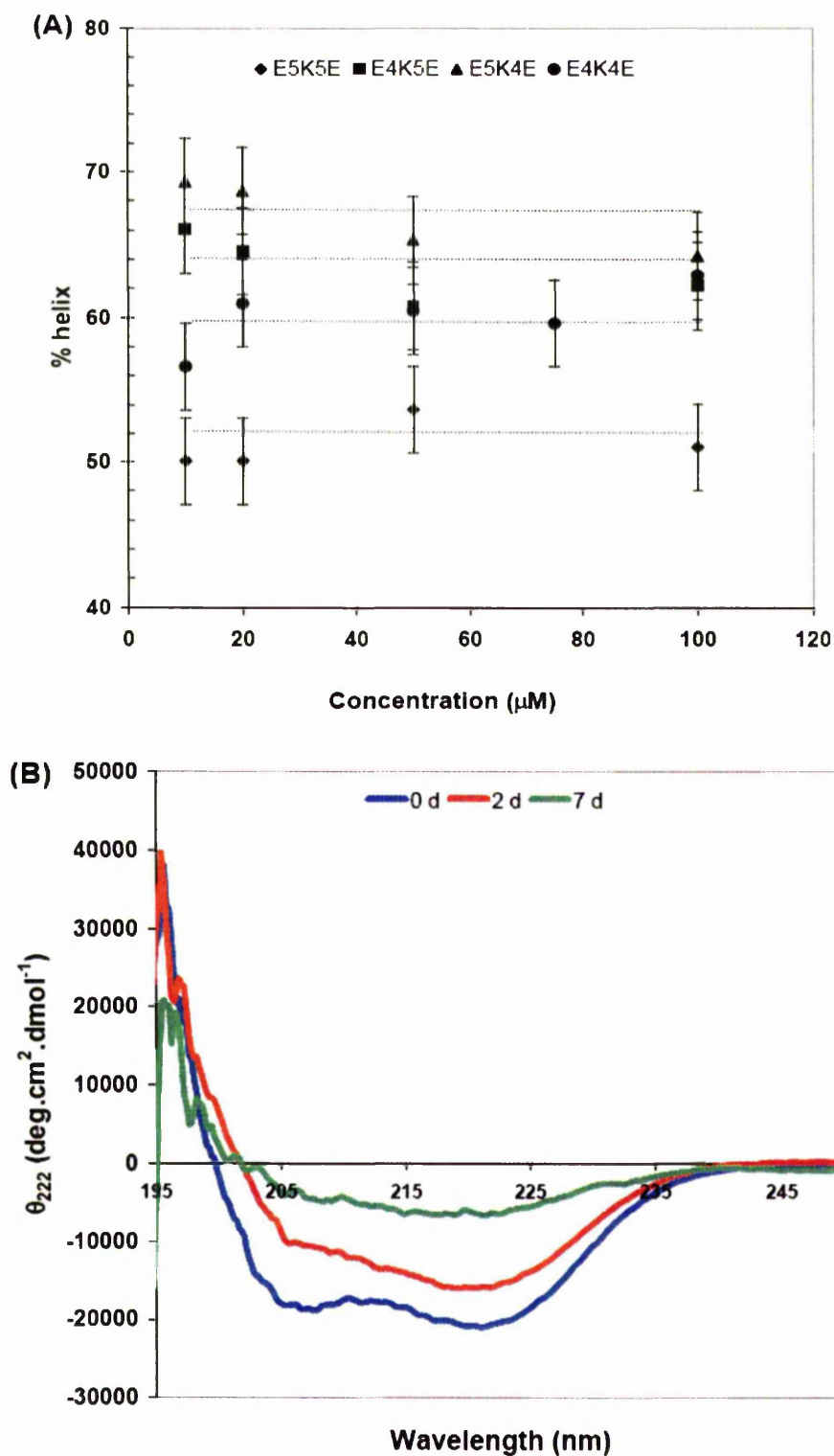
This also creates an extra hydrogen bond at each terminus. Acetyl groups also substantially stabilise the helix (Doig and Baldwin, 1995).

### 7.3. Aggregation tests of EKE peptides

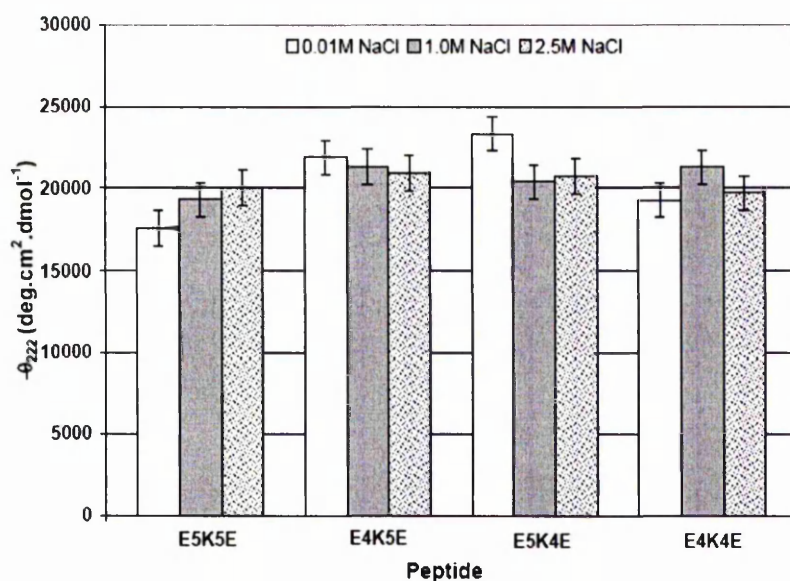
All peptides were monomeric between 5 and 100  $\mu\text{M}$  as shown by invariance of CD molar ellipticity with concentrations (Figure 7.1A). The E4K4E peptide however showed a change in secondary structure conformation from entirely helical to a mixture of helix: $\beta$ -sheet when being left for prolonged time ( $>48$  hours). This was observed by running analytical HPLC followed by mass spectra analysis on the peptide after 48 hours incubation. HPLC spectrum shows a tailing following the pure E4K4E peak. Yet both species gave identical mass when checked by MALDI (see Appendix A.30). The CD spectra in range of 190-250 nm showed a conformational shift from pure  $\alpha$ -helix structure (Figure 7.1B). The E4K4E peptide was thus collected solely on freshly dissolved pure lyophilised peptide. The reason behind the delayed aggregation could be because the E4K4E peptide contains residues assisting solubilisation (2 Glu and 3 Lys) distributed only on one face of the helix. Two, if not three, of the residues also involve in the side-chain interactions.

### 7.4. Salt screening

Results in Figure 7.2 show that the ellipticities of all EKE peptides were invariant and within error at 0.01, 1.0, 2.5M NaCl. This suggests that the interaction observed in the full ion pairs at high salt is not just a simple electrostatic interaction. Hence, a major component of the interaction between Glu and Lys cannot be screened by external salts. Hydrogen bonding thus provides the major part of the stabilising effect in the EK and KE pairs, as proposed previously by Scholtz *et al.* (1993).



**Figure 7.1.** (A) Aggregation test of EKE peptides. (B) Aggregation test of 10 $\mu\text{M}$  E4K4E peptide after prolonged incubation period. Helix content was converted from peptide CD ellipticity of peptides measured in 5mM phosphate buffer containing 10mM NaCl pH 8.2 at 273K. Error is CD measurement error of  $\pm 3\%$ . The horizontal dashed lines in (A) are to relate helicity at different concentrations and have no physical meaning.



**Figure 7.2.** Salt screening of EKE peptides. CD ellipticities were measured at pH 8.2 in 5mM phosphate buffer at 273K. Error is CD measurement error of  $\pm 1000$   $\text{deg.cm}^2.\text{dmol}^{-1}$ , which is approximately equal to  $\pm 3\%$  helicity.

### 7.5. Refitting $p$ values for EKs and KE interactions

The EK pair has been previously been studied in a different spacing and orientation in an isolated  $\alpha$ -helix (Lyu *et al.*, 1991; Scholtz *et al.*, 1993; Smith and Scholtz, 1998). The stabilising effect of EK and KE did not vary significantly with orientation. However, the results found in this study are different, with the KE pair more stabilising than the EK pair (Table 7.1). The helicities of the BKB series measured at low pH also show similar orientation preference. The terms E and B in ExKyE and BxKyB are referred in the text for describing peptides containing Glu<sup>-</sup> and Glu<sup>o</sup>, respectively.

The control peptides E5K5E and B5K5B show an excellent agreement between predicted and experimental helicity indicating that errors in previous parameter values used in the prediction for the residues in the sequences are low. The controls thus provide an independent system to measure the preferred interactions. There are, however, disagreements between the predicted and experimental helicity when  $i, i+4$  interactions are present, particularly in the KE/KB orientation. The sequence of peptides used to obtain the  $p$  values for EK/BK and KE/KB interactions are Ac-



AAQAEEAQAKAAQAAY-NH<sub>2</sub> and Ac-AAQAAKAQAEAAQAAY-NH<sub>2</sub>, respectively (Smith and Scholtz, 1998). The Gln near the helix termini in these peptides may interact with Glu and Lys in an  $i,i+3$  fashion that could give an error in the estimation of interaction energy. The EKE series do not contain these  $i,i+3$  interactions. In addition, Gln at the N3 position could form a significant interaction with the peptide backbone (Chapter 4).

The original  $p$  values of  $EK^+$  (2.15),  $K^+E^-$  (2.09),  $E^0K^+$  (1.50) and  $K^+E^0$  (1.45)  $i,i+4$  interaction were thus refitted. The  $p$  value is the equilibrium constant of an  $i,i+4$  interaction and can be used to calculate the free energy of interaction as  $\Delta G_{(i,i+4)} = -RT \ln p$  (Stapley *et al.*, 1995). More detail explanation about the  $p$  value can be seen in Chapter 2. The  $p$  values for each pair were fitted individually by varying them until the calculated helicity agreed with the experimental results. Refitting  $p$  values for the  $EK^+$  and  $K^+E^-$  interactions gives values of 3.0 and 6.6, respectively, energetically equivalent to -0.60kcal/mol and -1.02kcal/mol, giving an additive  $\Delta G$  of -1.62 kcal/mol. Refitting  $p$  values for the  $E^0K^+$  and  $K^+E^0$  interactions gives values of 1.7 and 2.5, respectively, corresponding to -0.29kcal/mol and -0.50kcal/mol, giving an additive  $\Delta G$  of -0.79kcal/mol (Table 7.2).

Using the refitted values, the predicted helicities are 74% and 69% for the corresponding E4K4E and B4K4B peptides. These predictions are significantly higher than the experimental helicity of 60% and 61% for the E4K4E and B4K4B, respectively. These large differences prove qualitatively that there is a large destabilising effect as a result of pairwise coupling in the E4K4E and B4K4B peptides. This is clear from the raw data, where there is remarkably a decrease in helix content in E4K4E compared to E5K4E and E4K5E, despite the introduction of an additional salt bridge.

For the predicted helicity to agree with the experiment in the E4K4E and B4K4B peptides, the  $p$  value of each pair was then decreased by a multiplication factor (mf) of 0.39 and 0.65, respectively (Figure 7.3). This decreases the statistical weight of all conformations within 9 consecutive helical residues from Glu<sub>N-term</sub> to Glu<sub>C-term</sub> simultaneously by  $(mf)^2$ . The  $p$  values for the  $EK^+$  and  $K^+E^-$  pairs in the

E4K4E peptide became 1.2 and 2.6, respectively, energetically equivalent to -0.09kcal/mol and -0.51kcal/mol. The  $p$  values for the  $E^\circ K^+$  and  $K^+E^\circ$  pairs in the B4K4B peptide became 1.1 and 1.6, respectively, energetically equivalent to -0.05kcal/mol and -0.26kcal/mol. This calculation is, however, based on the assumption that mf can be applied equally to EK and KE.

**Table 7.2.** Refitted  $p$  values and energetics of interaction in the EKE peptides.

Interaction	Refitted $p$ value ( $\pm 3\%$ helicity error)	$\Delta G$ pair (kcal/mol) <sup>a</sup>	Multiplication factor (mf) ( $\pm 3\%$ helicity error)	Refitted $p$ values $\times$ (mf) in triplet ( $\pm 3\%$ helicity error)	$\Delta G$ when in triplet <sup>b</sup> (kcal/mol)	$\Delta G$ Anti-cooperativity $-RT\ln(mf)^2$ (kcal/mol)
KE	6.6 (4.4 to 11.7)	-1.02 (-0.81 to -1.33)		2.6 (1.5 to 5.2)	-0.51 (-0.23 to -0.90)	
EK	3.0 (2.2 to 4.5)	-0.60 (-0.43 to -0.81)		1.2 (0.8 to 2.0)	-0.09 (0.15 to -0.37)	
EKE		-1.62 <sup>d</sup> (-1.24 to -2.15)	0.39 (0.34 to 0.45)		-0.60 <sup>d</sup> (-0.08 to -1.27)	1.02 (0.87 to 1.16)
KB	2.5 (1.9 to 3.6)	-0.50 (-0.35 to -0.69)		1.6 (1.1 to 2.6)	-0.26 (-0.04 to -0.52)	
BK	1.7 (1.3 to 2.4)	-0.29 (-0.13 to -0.47)		1.1 (0.7 to 1.8)	-0.05 (0.17 to -0.31)	
BKB		-0.79 <sup>d</sup> (-0.48 to -1.16)	0.65 (0.57 to 0.74)		-0.31 <sup>d</sup> (0.12 to -0.83)	0.47 (0.33 to 0.60)

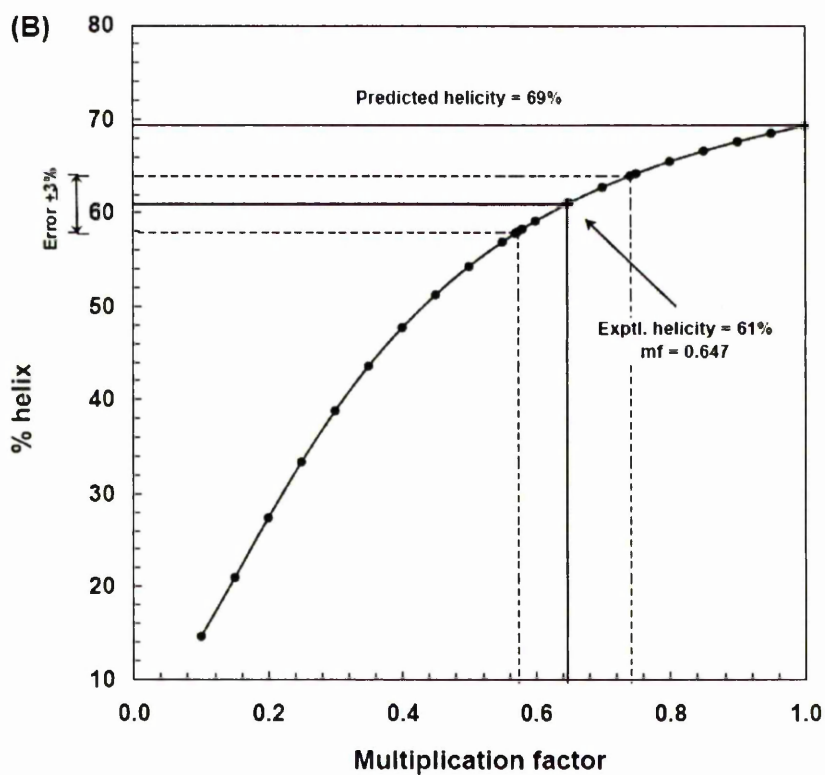
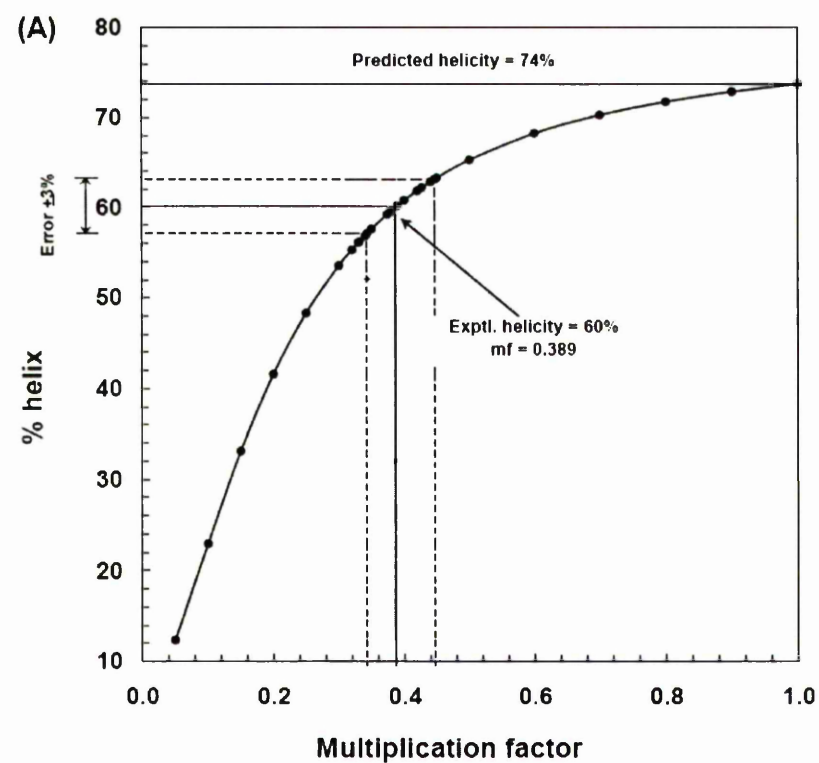
<sup>a</sup>Calculated using refitted  $p$  values.

<sup>b</sup>Calculated using refitted  $p$  values $\times$ (mf).

<sup>c</sup>B represents Glu<sup>o</sup>.

<sup>d</sup>Additive  $\Delta G$  of the pairs (kcal/mol)

The free energy resulting from the EK and KE coupling was calculated as  $-RT \ln (mf)^2$ , giving  $\Delta G$  of 1.02kcal/mol (0.87 to 1.16kcal/mol) and 0.47kcal/mol (0.33 to 0.60kcal/mol) for E4K4E and B4K4B, respectively. These values show an anti-cooperativity effect in the triplet as they are lower compared to the sum of the energies of the individual interactions. The error in  $p$  was calculated by repeating the fitting procedure using experimental helicities increased or decreased by 3% to account for experimental error in the measurements of helicity.



**Figures 7.3.** Adjustment of  $p$  values for the EK and KE pairs when present simultaneously in the triplet to get the predicted helicity to agree with experiment (A) In the EKE triplet. (B) In the BKB triplet.

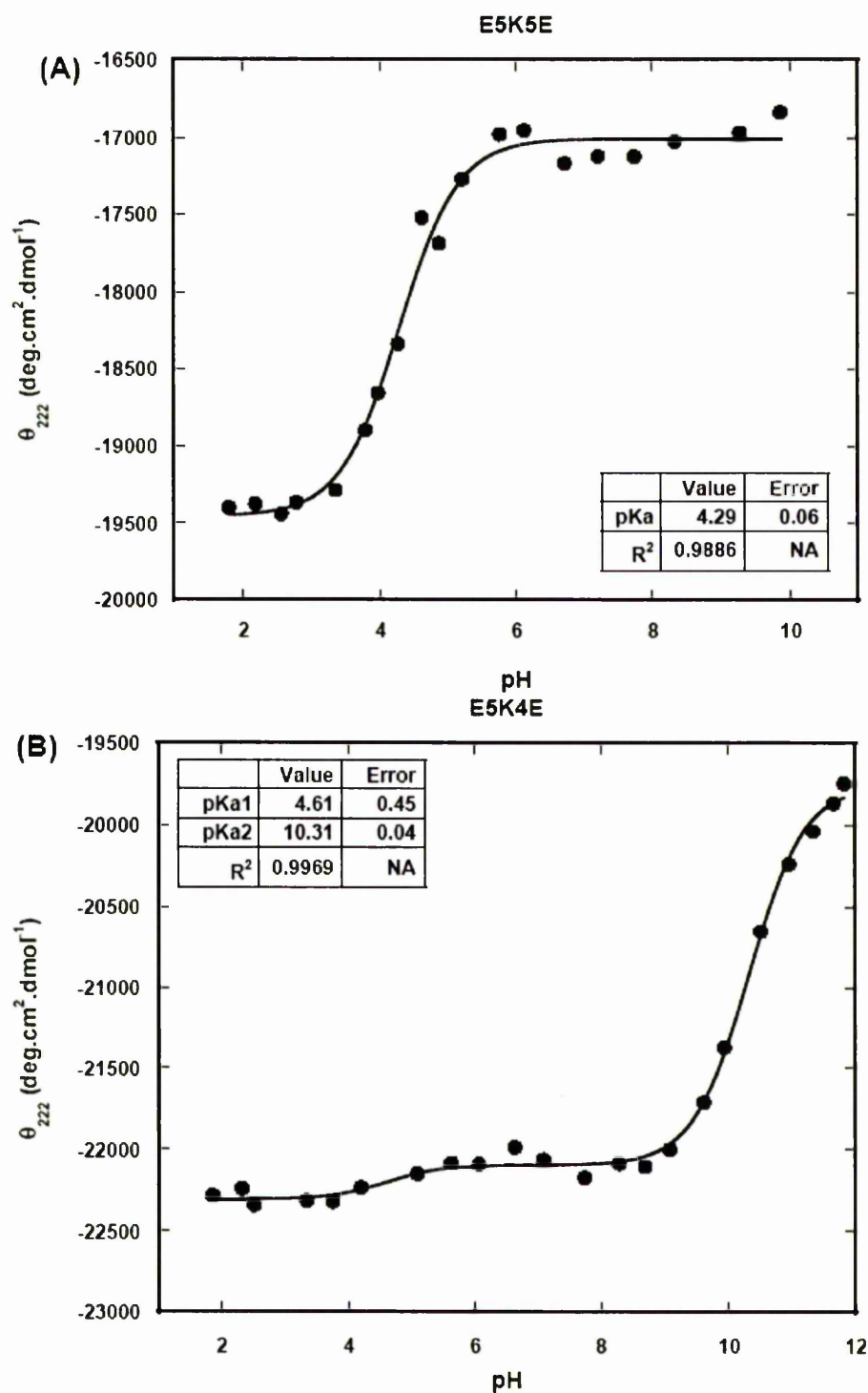
## 7.6. pH titration of EKE peptides

The helicities of the EKE peptides were followed as a function of pH using far UV CD (Figure 7.4). pKa values for the ionisable side-chain groups were evaluated between approximately pH 2-11 by curve fitting the CD data to the Henderson-Hasselbach equations (see Materials and Methods section 3.6). The upper range of accessible pH is given by Lys deprotonation, which leads to peptide aggregation.

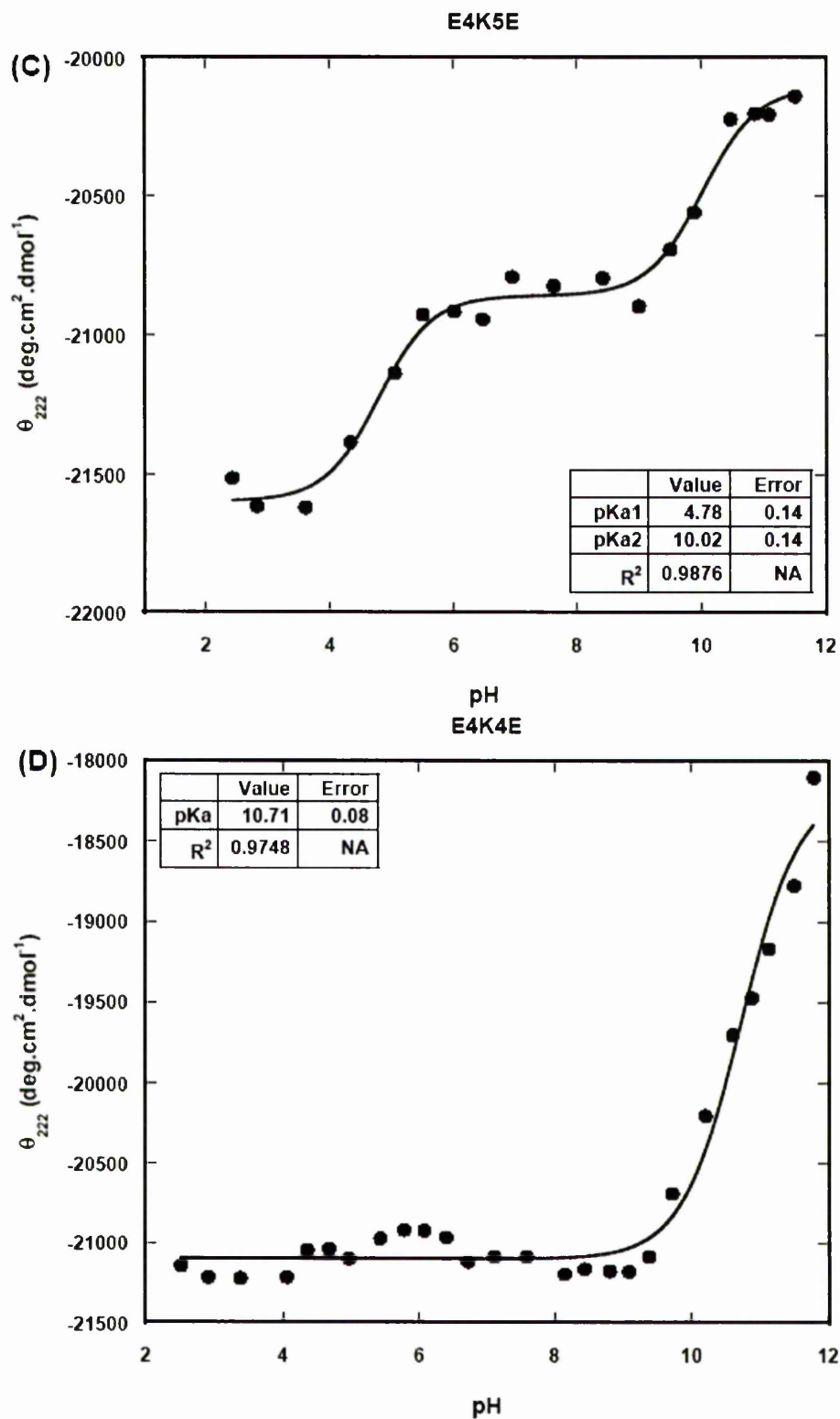
The pKa values of the two Glu residues in E5K5E cannot be separated. Only one observable transition across the pH range is present as the pKa values of the two Glu residues are similar. The pKa of  $4.29 \pm 0.06$  is comparable to the pKa value of Glu of 4.5 in a model compound N-acetyl acid amide (Nozaki and Tanford, 1967). The unperturbed pKa value for Glu residues is reasonable, as the side-chains are not involved in any interactions, yet the helix content is strongly influenced by the ionisation state. Protonation of Glu residues increase the helix content of E5K5E by about 7%. Previous works found that helix propagation parameters of a single Glu substitutions in the helix interior give a difference in free energy (about 0.15kcal/mol) for protonated and unprotonated Glu (Scholtz *et al.*, 1993; Rohl *et al.*, 1996).

Only one Glu transition was observed in the E4K5E and E5K4E peptides. The pKa values for E4K5E and E5K4E are  $4.78 \pm 0.14$  and  $4.61 \pm 0.45$ , respectively. It was expected that one of the Glu residues would have had a different pKa value to the Glu in E5K5E as it interacts with Lys. The difference in ellipticities upon protonation is very small, making it difficult to fit the data to calculate the pKa values, however. When the interaction is present, the destabilising effect of having a charged Glu in the helix is decreased, showed by the low ellipticity difference between charged and uncharged Glu. The Glu pKa in E4K5E and E5K4E surprisingly appears to be shifted to higher values than in E5K5E, though the high experimental errors in the pKa values make this conclusion uncertain. A decrease in Glu pKa would be expected within a salt-bridge, as the nearby positive charge would favour Glu<sup>-</sup>. The transition above pH 10 can be attributed to protonation of Tyr.

Titration curve of the E4K4E peptide appears to have a weak transition between pH 4 and pH 6, similar to that observed in Figure 7.4B. Repeating the experiment three times did not give significant difference, however. The fittings were not convincing when using equations for 2 or 3 apparent pKa values. Instead the data was fitted using equation for 1 apparent pKa. In general, the  $\theta_{222}$  does not significantly vary from pH 2-9, showing that the helix content of the peptide does not change when either Glu side-chain is protonated. This is due to a decrease in helix content when Glu is deprotonated, as  $\text{Glu}^-$  has a lower helix preference than  $\text{Glu}^0$ , which is counteracted by stronger salt-bridges with  $\text{Glu}^-$  in the helix. These two effects are close to equal and opposite in sign, giving no net change in  $\theta_{222}$ .



**Figure 7.4.** pH titration of EKE peptides (A) E5K5E, (B) E5K4E. pKa1 and pKa2 are associated with Glu and Tyr, respectively(continued next page).



**Figure 7.4.** (continued) pH titration of EKE peptides (C) E4K5E and (D) E4K4E.

## 7.7. Rotamer distribution of Glu and Lys residues

It has been demonstrated previously that rotamer preference is helpful in rationalising cooperativity in the RFM triplet in an  $\alpha$ -helical peptide (Chapter 6). The peptide only needs to pay the cost of restricting the Phe residue into a *trans* (*t*) conformation once in the triplet, rather than twice when the interactions are separate. The distribution of  $\chi_1$  rotamer preference of Glu and Lys residues was therefore examined. This PDB derived rotamer analysis, however, is based on the assumption that the rotamer distributions determined in the presence of tertiary contacts in proteins are applicable to peptides. Only the  $\chi_1$  rotamer population is presented as  $\chi_2$  is predominantly in the *t* rotamer in all cases.

### 7.7.1. $\chi_1$ rotamer preference of Glu and Lys in the EK and KE pairs

As summarised in Table 7.3, the Glu  $\chi_1$  population in an EK pair is spread almost equally between  $g^+$  (48%) and *t* (47%). The Lys  $\chi_1$  tends to be in  $g^+$  (61%). In the KE pair, the Lys  $\chi_1$  favours the *t* rotamer (61%). The Glu  $\chi_1$  is overwhelmingly in  $g^+$ . Individual Glu and Lys rotamer preferences in coil and helix were included for comparison. Rotamer definition has been previously described in Chapter 3.

### 7.7.2. $\chi_1$ rotamer preference of Glu and Lys in EKE triplet

The EKE triplet spaced  $i, i+4, i+8$  was found in 58 cases in the data set (Table 7.4). This number is relatively small, making quantitative conclusion on rotamer preferences in the triplet difficult. Almost half of  $\chi_1$  rotamer combination were  $g^+, t, g^+$  for the corresponding Glu, Lys and Glu residues in the EKE triplet, while one fifth are in a  $g^+, g^+, g^+$  combination. This suggests that the N-terminal Glu in the EK pair switches from equally  $\chi_1$   $g^+$  (48%) and *t* (47%) to mostly  $g^+$  when in the EKE triplet. The Lys  $\chi_1$  also switches from mostly  $g^+$  (61%) to *t* in the triplet. The Glu and Lys  $\chi_1$  preference in the KE pair seems to be retained in the triplet. The majority of Lys in the EKE triplet adopts  $\chi_1$  *t*, as for Lys in the KE pair. The preference of Glu  $\chi_1$   $g^+$  in the EKE triplet is also comparable to Glu  $\chi_1$   $g^+$  in the KE pair. Overall, this suggests that within the EKE triplet, the KE interaction is present, while the EK interaction is



not. The central Lys cannot adopt the preferred rotamers for both salt-bridges simultaneously.

**Table 7.3.**  $\chi_1$  rotamer population of Glu and Lys in  $\alpha$ -helices.

Residue	Rotamer distribution <sup>a</sup>			No. of observation
	$g^-$	$t$	$g^+$	
Glu	$0.04 \pm 0.003^b$	$0.39 \pm 0.01$	$0.57 \pm 0.01$	5947
Lys	$0.03 \pm 0.003$	$0.44 \pm 0.01$	$0.53 \pm 0.01$	4409
Glu in EK pair <sup>c</sup>	$0.05 \pm 0.01$	$0.47 \pm 0.02$	$0.48 \pm 0.02$	468
Lys in EK pair <sup>c</sup>	$0.02 \pm 0.01$	$0.37 \pm 0.02$	$0.61 \pm 0.02$	468
Lys in KE pair <sup>c</sup>	$0.02 \pm 0.01$	$0.62 \pm 0.02$	$0.36 \pm 0.02$	411
Glu in KE pair <sup>c</sup>	$0.03 \pm 0.01$	$0.27 \pm 0.02$	$0.70 \pm 0.02$	411

<sup>a</sup>Data were from domain search of 1135 non-redundant proteins containing 4778 helices of a total 65472 residues in ASTRAL database (*SCOP 1.63 Sequence Resources*). Rotamer definitions:  $-120^\circ < \chi_1 < 0^\circ = g^+$ ;  $0^\circ < \chi_1 < 120^\circ = g^-$ ;  $-120^\circ < \chi_1 < 240^\circ = t$ .

<sup>b</sup>Errors were calculated as  $\sigma pi = [pi*(1 - pi)/Ni]^{1/2}$ , where  $pi$  is the rotamer populations and  $Ni$  is the total number of observation.

<sup>c</sup>The pairs were observed in *hhhhh* conformations.

**Table 7.4.**  $\chi_1$  rotamer distribution of the interacting residues in EKE triplet in  $i, i+4, i+8$  spacing.

$\chi_1$ rotamer combination*			Total observed in <i>hhhhhhhhh</i> conformation	Percentage
Glu	Lys	Glu		
$g^+$	$g^-$	$g^+$	1	2%
$g^-$	$g^+$	$g^-$	1	2%
$g^+$	$g^+$	$g^+$	12	21%
$t$	$g^+$	$t$	5	9%
$g^+$	Not defined	$g^+$	1	2%
$g^-$	$t$	$g^-$	2	3%
$g^+$	$t$	$g^+$	27	47%
$t$	$t$	$t$	9	16%
TOTAL			58	100%

\*Data were from domain search of 1135 non-redundant proteins containing 4778 helices of a total 65472 residues in ASTRAL database (*SCOP 1.63 Sequence Resources*). Rotamer definitions:  $-120^\circ < \chi_1 < 0^\circ = g^+$ ;  $0^\circ < \chi_1 < 120^\circ = g^-$ ;  $-120^\circ < \chi_1 < 240^\circ = t$ .

## 7.8. Discussion

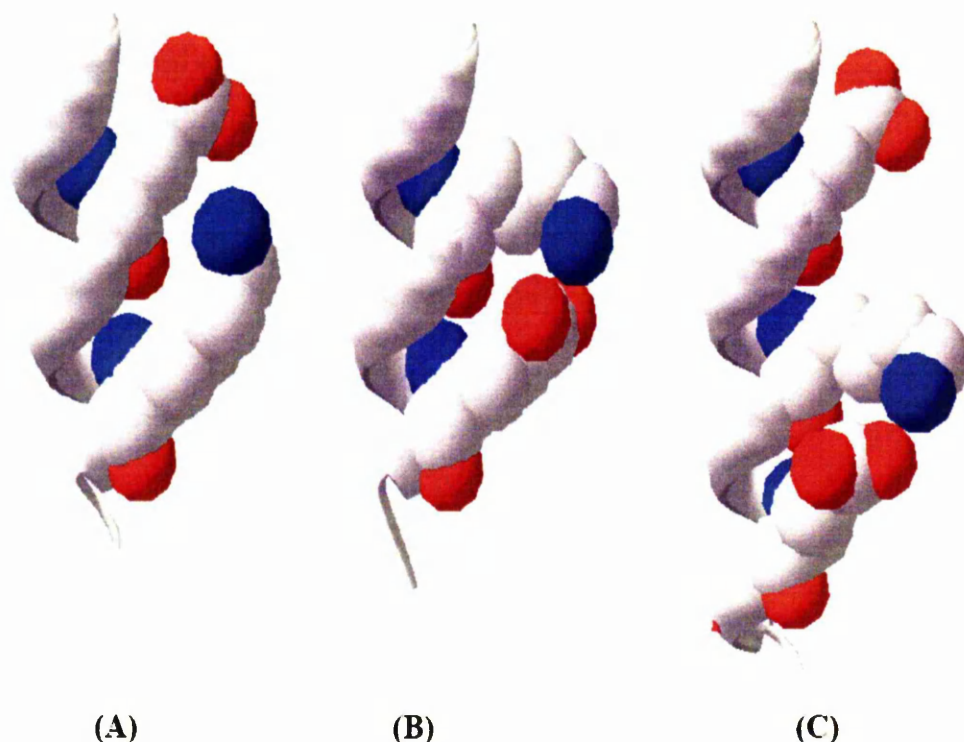
The strength of salt bridges is determined by several factors, namely (1) the geometry and distance of the interactions (Kumar and Nussinov, 2002); (2) the degree of exposure to solvent (Takano *et al.*, 2000); (3) the effect of neighbouring residues (Makhataдзе *et al.*, 2003). The salt bridges in the  $\alpha$ -helical peptides studied here are solvent-exposed and relatively in isolation from interaction with neighbouring residues. The EKE peptide series are a useful model for studying the solvent-exposed salt bridges as they simplify generalisation of the energetic role of the interacting pairs.

The E4K5E peptide has a lower helix content than E5K4E, suggesting that the EK interaction is weaker than KE. This is confirmed by helix-coil theory analysis, which gives free energies of -1.02kcal/mol and -0.60kcal/mol for  $K^+E^-$  and  $E^-K^+$ , respectively. Previously reported  $\Delta G$  values show that the stabilising free energy for EK and KE pairs are fairly similar (Lyu *et al.*, 1991; Scholtz *et al.*, 1993; Smith and Scholtz, 1998), contradicting the findings in this study.

One could argue that in the E4K4E peptide the Lys tends to interact with only one of the Glu residues and the unpaired Glu would not give any affect on the overall stability. The helicity of E4K4E would thus be indifferent in comparison either with E4K5E or E5K4E. The results in this study, however, are intriguing due to the complexity of the charges involved. It appears that in the E4K4E peptide the Lys can interact with only one of the Glu residues. Thus when E4K4E is compared to E5K4E, the addition of the second EK salt-bridge has little stabilising effect. In fact, there is a slight decrease in stability that can be attributed to moving a Glu further from the N-terminus of the peptide, weakening any helix-dipole interactions. Glu is a strongly favoured residue at the N1, N2 and N3 positions, which it can populate if the helix frays. The lower helicity in E4K4E compared to E4K5E is more difficult to explain. It is suggested that when this change is made, the Lys moves to form a KE interaction, rather than the weaker EK in E4K5E. In addition, the Glu is moved further from the C-terminus, where it will form a destabilising interaction with the helix dipole. The combination of these three effects, which are difficult to separate, results in a small decrease in helix content. The energetic analysis with helix coil

theory shows that the events occurring in E4K4E give a stabilising energy of approximately 1.0kcal/mol when the interacting residues are fully charged.

Examples of interactions are shown in Figure 7.5. When the EKE triplet is present, KE dominates so the Lys predominantly adopts the rotamer that points it towards the i+4 Glu (Figure 7.5C). This implies that essentially only a single salt bridge is present in the triplet, which is, to some extent, confirmed by the rotamer analysis in tables 7.3 and 7.4. The rotamer analysis presented in this study contradicts the previous finding that the geometry of the interactions between acidic and basic residues is very similar in simple and complex salt bridges (Musafia *et al.*, 1995). The authors proposed that adding one residue to a simple interaction shows a minor change in geometry, in term of atomic distance, partly because half of the salt bridge interaction is already constrained in a geometry that satisfies a complex salt bridge. The KE pair is indeed constrained, but the rotamer preferences in EK pairs change when in the triplet.



**Figure 7.5.** Structural examples of (A) EK interaction between Glu62 and Lys66 in PDB-1A48 (B) KE interaction between Lys175 and Glu179 in PDB-1AOP (C) EKE interaction between Glu110 and Lys114-Glu118 in PDB-1L2P.

As part of the destabilising free energy can be explained from the switching rotamer preference of the interacting pairs, it is thus tempting to compare the results in this study to previously reported triplets through rotamer preference and structural analysis. Table 7.5 shows the energetics of the complex salt bridges in the RER triplet in  $i, i+4, i+8$  spacing (Olson *et al.*, 2001b). The interaction is cooperative as the  $\Delta G$  of the pairwise coupling of RE and ER is greater then the sum of the  $\Delta G$  of individual pair.

**Table 7.5.** Energetics of RER triplet (Olson *et al.*, 2001b).

Peptide	Sequence	Fraction helix	$\Delta\Delta G$ (kcal/mol)	$\Delta\Delta G_{\text{complex}}$ (kcal/mol)
R5E5R	Ac-OOAAAAAARAAAAEAAAAAARAAAAOOY-NH <sub>2</sub>	0.57	0	
R5E4R	Ac-OOAAAAAARAAAAEAAAAAARAAAAOOY-NH <sub>2</sub>	0.60	-0.223	
R4E5R	Ac-OOAAAAAARAAAAEAAAAAARAAAAOOY-NH <sub>2</sub>	0.635	-0.623	
R4E4R	Ac-OOAAAAAARAAAAEAAAAAARAAAAOOY-NH <sub>2</sub>	0.72	-2.36	-1.5

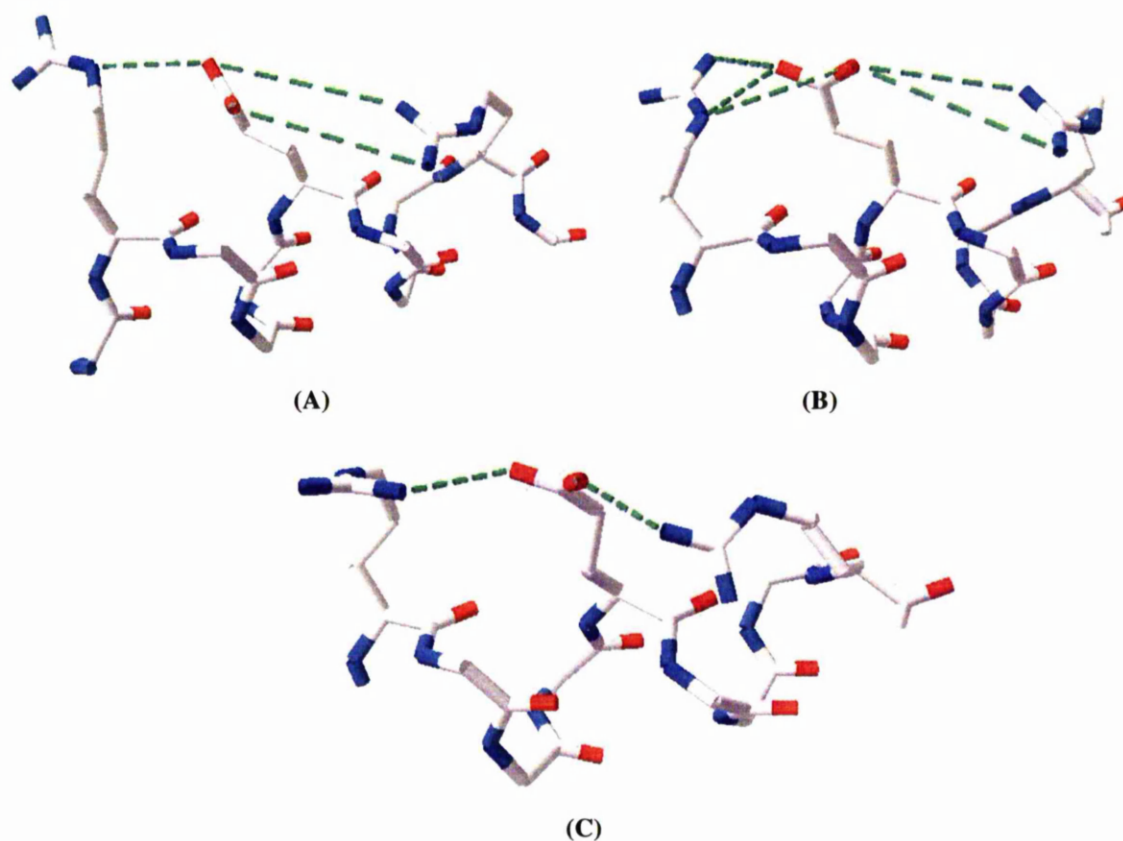
Analysis of the  $\chi_1$  rotamer preference in table 7.6 reveals that the Arg favours  $\chi_1 t$  in the RE pair, whilst the Glu prefers  $\chi_1 g^+$ . In the ER pair, the Glu favours  $\chi_1 t$  whilst the Arg prefers  $\chi_1 g^+$ . It seems that there is a conflicting rotamer preference for the Glu when in the triplet. As the triplet is cooperative, the rotamer preference for Glu in both pairs should be distributed somewhat similarly. The PDB search on the RER triplet in *hhhhhhhhh* conformation gives 15 structures, which is a small number for conducting a statistical analysis. However, there are a dominant five structures that have RER rotamers in  $g^+, t, g^+$  combinations and four structures with  $t, g^+, t$  combinations.

**Table 7.6.**  $\chi_1$  rotamer distribution of Glu and Arg in RE and ER pairs spaced  $i, i+4$ .

Residue	$\chi_1$ rotamer distribution*			No. observed in <i>hhhhh</i> conformations
	$g^-$	$t$	$g^+$	
Arg in RE pair	0.03 $\pm$ 0.01	0.63 $\pm$ 0.03	0.35 $\pm$ 0.03	328
Glu in RE pair	0.02 $\pm$ 0.01	0.27 $\pm$ 0.02	0.70 $\pm$ 0.03	328
Glu in ER pair	0.03 $\pm$ 0.01	0.53 $\pm$ 0.03	0.44 $\pm$ 0.03	360
Arg in ER pair	0.04 $\pm$ 0.01	0.38 $\pm$ 0.03	0.59 $\pm$ 0.03	360

\*Data from domain search of 1135 non-redundant proteins containing 4778 helices in ASTRAL database (SCOP 1.63 Sequence Resources). Rotamer definitions:  $-120^\circ < \chi_1 < 0^\circ = g^+$ ;  $0^\circ < \chi_1 < 120^\circ = g^-$ ;  $-120^\circ < \chi_1 < 240^\circ = t$ . Errors were calculated as  $\sigma_i = [p_i(1-p_i)/N_i]^{1/2}$ , where  $p_i$  is the rotamer populations and  $N_i$  is the total number of pair found in helices.

Structural analysis using the Swiss-PdbViewer program (v.3.7) of the RER triplet in the proteins shows that the Glu in the middle can form multiple hydrogen bonds to both Arg flanking it, regardless to the rotamer adopted. This could explain why this triplet is still stabilising despite conflicting rotamer preference for the Glu. In the E4K4E triplet, the hydrogen bonding can only be formed by only one pair at the same time because of the distance restriction.



**Figure 7.6.** Structural examples of R4E4R triplet in  $\alpha$ -helix (A) PDB-1G2A where  $\chi_1$  is in  $t, g^+, t$  combination (B) PDB-1MSK where  $\chi_1$  is in  $g^+, g^+, g^+$  combination (C) PDB-1QJB where  $\chi_1$  is in  $g^+, t, g^+$  combination. All  $\chi_2$  rotamers are in  $t$ .

This comparison suggests that rotamer preference cannot be used solely to analyse the cooperativity or anti-cooperativity within networks of amino acids. Other factors such as charge effects, dipole interactions, atomic distance and neighbouring residues should also be considered.

## 7.8. Conclusions

The presence of EK and KE interactions minimises the destabilising effect of having negatively charged Glu in the EKE peptides. The stabilising interactions are mostly from hydrogen bonding rather than from pure electrostatic interactions. When both interactions are present simultaneously, however, they destabilise the helix by 1.0kcal/mol when fully charged. The destabilising effect of the pairwise coupling is attributed not only to the change in the rotamer preference, but also to interactions with helix dipole at both termini, although they cannot be separated fully. In the EKE system studied above, the addition of a salt-bridge can be worth essentially nothing. This is attributed from the inability of the central Lys to form two interactions simultaneously. Even in the simplest systems, non-covalent interactions are non-additive and can be cooperative or anti-cooperative, depending on how compatible the conformational preferences of the interactions are. Although variations in the strength of salt-bridges have been noted and attributed to solvation effects, on the surface of a helical peptide, changes in solvation are minor. The results from this study would give a new insight in the role of complex salt bridges as well as contribute to protein engineering studies. The energetics of interactions may be useful in protein structure prediction.

## CHAPTER 8

### GENERAL CONCLUSIONS AND FUTURE WORK

This thesis has given a range of new information on factors that affect  $\alpha$ -helix stability. The amino acids at the N3 position of the  $\alpha$ -helix N-terminus show distinct preferences. Here, a complete empirical study on this site using peptides of similar sequence has given an accurate energetic scale. In combination with energetics of amino acids at other positions, it could be useful in the areas of protein modelling. The structural uniqueness of the N3 is partly attributed to the interactions with helix dipole at the N-terminus. The unsatisfied amide groups at the N1, N2 and N3 positions may also play a part, although these interactions could not be separated here. Other techniques, such as NMR, may resolve this problem by looking at the interactions at an atomic level, but the approaches are beyond the scope of this thesis.

The behaviour of the CXXC motif is difficult to interpret in the intact proteins due to the presence of neighbouring charges and other interactions. In isolation, the motif is found to be highly reducing in a synthetic  $\alpha$ -helical peptide, which is very similar to the nature of human thioredoxin. The redox potential could be altered by substituting the proximal residues, especially those in between the Cys residues. In this way, small peptides containing the motif with various redox potentials could be employed to study thiol-disulfide exchange mechanisms in the intact proteins. This approach would simplify the techniques and data interpretation compared to proteins. The helix stabilising capacity of the motif through disulfide bond formation and Zinc binding can also be used for *de novo* protein designs, as well as altering protein stability. Protein crystal surveys on the motif have also given a new insight on its high preference in the helix N-terminus, although the high preference might contain flaws due to the low frequency of the interacting residues.

It is widely known that amino acid side-chains interact within a complex network of multiple non-covalent bonds, which may reinforce or weaken each other. Again, the energetics of such networks is difficult to quantify due to the presence of other interactions in the whole proteins. The simplest system to investigate these

effects is to study triplets of side-chains in an isolated environment, as demonstrated in this thesis. This approach is not new, but is still in its early stages. Hence, it has the potential to study amino acid networks, particularly in more complex networks. Although the coupling effects shown in the RFM and EKE triplets were observed in isolated  $\alpha$ -helix peptides, they can also be applied to rationalise similar interactions in intact proteins. I propose that rotamer preferences can be used, in part, to rationalise the stabilising or destabilising effects of the pairwise couplings. However, they may be accompanied by other factors such as charge effects, dipole interactions, atomic distances and neighbouring residues. Stabilisation via shared rotamer, as shown in the RFM triplet, could be a general property of proteins. Any bond which restricts a side-chain into a conformation that is favourable for forming additional interactions will show cooperativity in folding and increase protein stability. The effect is substantial, as fixing just a single side-chain into its preferred conformation in the RFM triplet is worth 0.75kcal/mol. In contrast, the destabilising effect of the EKE pairwise coupling is more complex and attributed not only to the change in the rotamer preference, but also to interactions with the helix dipole at both termini, although they cannot be separated fully. Previously, solvation effects were considered to give variations in the strength of salt-bridges. The results in the EKE peptides study suggest that solvation effects on the surface of a helical peptide are small. The results from this study have given a new perception in the role of amino acid networks. Even in the simplest systems, non-covalent interactions are non-additive and can be cooperative or anti-cooperative, depending on how compatible the conformational preferences of the interactions are.

The energetics of interactions studied in this thesis are difficult to interpret without the use of helix-coil theory. This is largely due to the multi-state conformations of a sequence can be adopted in an  $\alpha$ -helical peptide. Once again, the helix-coil theory has been a powerful tool to resolve the energetics from CD-derived helicities. Some modifications to the original theory were made, notably in the energetic analysis of the CXXC motif. The unprecedented N-cap and N3 side-chain interaction in the CXXC motif requires changing the parameter values of the interacting and intervening residues used in the original algorithm. The stabilising or destabilising effect of pairwise coupling was also dealt with in a similar way.



This thesis is by no means intended to resolve the protein folding problem. However, it has provided some new perspectives related to  $\alpha$ -helix stability which may be used to generalise its folding mechanism. A powerful algorithm to predict  $\alpha$ -helix structures from a given sequence has already been established. More empirical data is, however, needed to improve the predictions. Prediction algorithms for other secondary structures are not as well developed as that for the  $\alpha$ -helix. Prediction of the whole protein structures is even more poorly understood. These are areas that still need improvement in the future, particularly in explaining interactions between the secondary structures which requires incorporation of additional empirical data. In the field of bioinformatics, sequence similarities could be enough to predict structures. It is thus interesting to see how empirical and bioinformatics data can be combined to develop a powerful tool in the area of protein structure prediction.

The protein prediction problem has become increasingly distinct from the protein folding problem. The former focuses with predicting the final structure, while the latter focuses on folding pathways. A fascinating question to ask is how useful the enormous efforts that have been devoted to understanding folding pathways and folding energetics will be in structure prediction. In my opinion, integrating these two areas is a big task, but would be very useful if it can be achieved. We can then reciprocally learn about sequence and structure and their relationship with folding pathways and folding energetics. At the same time, improved understanding of kinetics and stability should help us in the design of prediction algorithms. It is thus necessary to find a way to integrate the two fields.

## REFERENCES

- Albert, J.S. and Hamilton, A.D. 1995. Stabilization of helical domains in short peptides using hydrophobic interactions. *Biochemistry* **34**: 984-990.
- Alberts, I.L., Nadassy, K. and Wodak, S.J. 1998. Analysis of zinc binding sites in protein crystal structures. *Protein Sci* **7**: 1700-1716.
- Andersen, N.H. and Tong, H. 1997. Empirical parameterization of a model for predicting peptide helix coil equilibrium populations. *Protein Sci.* **6**: 1920-1936.
- Anderson, D.E., Bechtel, W.J. and Dahlquist, F.W. 1990. pH-induced denaturation of proteins: A single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29**: 2403-2408.
- Andrew, C.D., Bhattacharjee, S., Kokkoni, N., Hirst, J.D., Jones, G.R. and Doig, A.J. 2002a. Stabilizing interactions between aromatic and basic side chains in  $\alpha$ -helical peptides and proteins. Tyrosine effects on helix circular dichroism. *J. Am. Chem. Soc.* **124**: 12706-12714.
- Andrew, C.D., Penel, S., Jones, G.R. and Doig, A.J. 2001. Stabilizing nonpolar/polar side-chain interactions in the  $\alpha$ -helix. *Proteins* **45**: 449-455.
- Andrew, C.D., Warwicker, J., Jones, G.R. and Doig, A.J. 2002b. Effect of phosphorylation on  $\alpha$ -helix stability as a function of position. *Biochemistry* **41**: 1897-1905.
- Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181**: 223-230.
- Aqvist, J., Luecke, H., Quirocho, F.A. and Warshel, A. 1991. Dipoles located at helix termini of proteins stabilize charges. *Proc. Nat. Acad. Sci. U.S.A.* **88**: 2026-2030.
- Armstrong, K.M. and Baldwin, R.L. 1993. Charged histidine affects  $\alpha$ -helix stability at all positions in the helix by interacting with the backbone charges. *Proc. Nat. Acad. Sci. U.S.A.* **90**: 11337-11340.
- Aurora, R. and Rose, G.D. 1998. Helix capping. *Protein Sci.* **7**: 21-38.
- Aurora, R., Srinivasan, R. and Rose, G.D. 1994. Rules for  $\alpha$ -helix termination by glycine. *Science* **264**: 1126-1130.
- Baker, E.N. and Hubbard, R.E. 1984. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44**: 97-179.
- Baldwin, R.L. and Rose, G.D. 1999a. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem.Sci.* **24**: 26-33.
- Baldwin, R.L. and Rose, G.D. 1999b. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24**: 77-83.
- Barlow, D.J. and Thornton, J.M. 1988. Helix geometry in proteins. *J. Mol. Biol.* **201**: 601-619.
- Bell, J.A., Bechtel, W.J., Sauer, U., Baase, W.A. and Matthews, B.W. 1992. Dissection of helix capping in T4 lysozyme by structural and thermodynamic analysis of six amino acid substitutions at Thr 59. *Biochemistry* **31**: 3590-3596.

- Berezhkovskiy, L., Pham, S., Reich, E.-P. and Deshpande, S. 1999. Synthesis and kinetics of cyclization of MHC class II derived cyclic peptide vaccine for diabetes. *J. Peptide Res.* **54**: 112-119.
- Blaber, M., Zhang, X.J., Lindstrom, J.D., Pepiot, S.D., Baase, W.A. and Matthews, B.W. 1994. Determination of  $\alpha$ -helix propensity within the context of a folded protein: Sites 44 and 131 in bacteriophage-T4 lysozyme. *J. Mol. Biol.* **235**: 600-624.
- Blaber, M.W., Baase, W.A., Gassner, N. and Matthews, B.W. 1993. Structural basis of amino acid  $\alpha$ -helix propensity. *Science* **260**: 1637-1640.
- Bolin, K.A. and Millhauser, G.L. 1999.  $\alpha$  and  $3_{10}$ : The split personality of polypeptide helices. *Acc. Chem. Res.* **32**: 1027-1033.
- Bouvier, M. and Taylor, J.W. 1992. Probing the functional conformation of neuropeptide-T through the design and study of cyclic analogs. *J. Med. Chem.* **35**: 1145-1155.
- Bracken, C., Gulyas, J., Taylor, J.W. and Baum, J. 1994. Synthesis and nuclear-magnetic-resonance structure determination of an  $\alpha$ -helical, bicyclic, lactam-bridged hexapeptide. *J. Am. Chem. Soc.* **116**: 6431-6432.
- Brenner, S.E., Koehl, P. and Levitt, R. 2000. The Astral compendium for protein structure and sequence analysis. *Nucleic Acids Res.* **28**: 254-256.
- Butterfield, S.M., Patel, P.R. and Waters, M.L. 2002. Contribution of aromatic interactions to  $\alpha$ -helix stability. *J. Am. Chem. Soc.* **124**: 9751-9755.
- Campbell, R.M., Bongers, J. and Felxi, A.M. 1995. Rational design, synthesis, and biological evaluation of novel growth-hormone releasing-factor analogs. *Biopolymers* **37**: 67-68.
- Carpino, L.A. and Han, G.Y. 1970. 9-fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group. *J. Am. Chem. Soc.* **92**: 5748-&.
- Carpino, L.A. and Han, G.Y. 1972. 9-fluorenylmethoxycarbonyl amino-protecting group. *J. Org. Chem.* **37**: 3404-&.
- Chakrabartty, A., Doig, A.J. and Baldwin, R.L. 1993a. Helix capping propensities in peptides parallel those in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 11332-11336.
- Chakrabartty, A., Kortemme, T. and Baldwin, R.L. 1994. Helix propensities of the amino-acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **3**: 843-852.
- Chakrabartty, A., Kortemme, T., Padmanabhan, S. and Baldwin, R.L. 1993b. Aromatic side-chain contribution to far-ultraviolet circular-dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* **32**: 5560-5565.
- Chakrabartty, A., Schellman, J.A. and Baldwin, R.L. 1991. Large differences in the helix propensities of alanine and glycine. *Nature* **351**: 586-588.
- Chen, S.T., Chen, H.J., Yu, H.M. and Wang, K.T. 1993. Facile synthesis of a short peptide with a side-chain constrained structure. *J. Chem. Res. (S)* **6**: 228-229.
- Chivers, P.T., Prehoda, K.E. and Raines, R.T. 1997a. The CXXC motif: A rheostat in the active site. *Biochemistry* **36**: 4061-4066.
- Chivers, P.T., Prehoda, K.E., Volkman, B.F., Kim, B.-M., Markley, J.L. and Raines, R.T. 1997b. Microscopic pKa values of *Escherichia coli* thioredoxin. *Biochemistry* **36**: 14985-14991.

- Chorev, M., Roubini, E., McKee, R.L., Gibbons, S.W., Goldman, M.E., Caulfield, M.P. and Rosenblatt, M. 1991. Cyclic parathyroid-hormone related protein antagonists - Lysine 13 to aspartic acid 17 [i to (i+40)] side-chain to side-chain lactamization. *Biochemistry* **30**: 5968-5974.
- Chou, P.Y. and Fasman, G.D. 1978. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**: 251-276.
- Cleland, W.W. 1964. Dithiothreitol, a new protective reagent for sh groups. *Biochemistry* **3**: 480-482.
- Cline, D.J., Thorpe, C. and Scheider, J.P. 2003. Effects of As(III) binding on  $\alpha$ -helical structure. *J. Am. Chem. Soc.* **125**: 2923-2929.
- Cochran, D.A.E. and Doig, A.J. 2001. Effect of the N2 residue on the stability of the  $\alpha$ -helix for all 20 amino acids. *Protein Sci.* **10**: 1305-1311.
- Cochran, D.A.E., Penel, S. and Doig, A.J. 2001. Effect of the N1 residue on the stability of the  $\alpha$ -helix for all 20 amino acids. *Protein Sci.* **10**: 463-470.
- Creamer, T.P. and Rose, G.D. 1992. Side-chain entropy opposes  $\alpha$ -helix formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl. Acad. Sci. U. S. A.* **89**: 5937-5941.
- Creighton, T.E., Zapun, A. and Darby, N.J. 1995. Mechanisms and catalysts of disulfide bond formation in proteins. *Trends Biotechnol.* **13**: 18-23.
- Daggett, V. and Fersht, A.R. 2003. Is there a unifying mechanism for protein folding? *Trends Biochem.Sci.* **28**: 18-25.
- Daopin, S., Sauer, U., Nicholson, H. and Matthews, B.W. 1991. Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry* **30**: 7142-7153.
- Dasgupta, S. and Bell, J.A. 1993. Design of helix ends. Amino acid preferences, hydrogen bonding and electrostatic interactions. *Int. J. Pept. Protein Res.* **41**: 499-511.
- Debarbieux, L. and Beckwith, J. 1999. Electron avenue: Pathways of disulfide bond formation and isomerization. *Cell* **99**: 117-119.
- Dillet, V., Dyson, H.J. and Bashford, D. 1998. Calculations of electrostatic interactions and pKas in the active site of *Escherichia coli* thioredoxin. *Biochemistry* **37**: 10298-10306.
- Dobson, C.M. 1999. Protein misfolding, evolution and disease. *Trends Biol. Sci.* **24**: 329-332.
- Dobson, C.M. 2003. Protein folding and misfolding. *Nature* **426**: 884-890.
- Doig, A.J. and Baldwin, R.L. 1995. N- and C-capping preferences for all 20 amino acids in  $\alpha$ -helical peptides. *Protein Sci.* **4**: 1325-1336.
- Doig, A.J., Chakrabartty, A., Klingler, T.M. and Baldwin, R.L. 1994. Determination of free energies of N-capping in  $\alpha$ -helices by modification of the Lifson-Roig helix-coil theory to include N- and C-capping. *Biochemistry* **33**: 3396-3403.
- Doig, A.J., MacArthur, M.W., Stapley, B.J. and Thornton, J.M. 1997. Structures of N-termini of helices in proteins. *Protein Sci.* **6**: 147-155.
- Doig, A.J. and Sternberg, M.J.E. 1995. Side-chain conformational entropy in protein-folding. *Protein Sci.* **4**: 2247-2251.

- Dudev, T. and Lim, C. 2000. Tetrahedral vs octahedral zinc complexes with ligands of biological interest: A DFT/CDM study. *J. Am. Chem. Soc.* **122**: 11146-11153.
- Dudev, T. and Lim, C. 2003. Principles governing Mg, Ca and Zn binding and selectivity in proteins. *Chem. Rev.* **103**: 773-787.
- Dyson, H.J., Gippert, G.P., Case, D.A., Holmgren, A. and Wright, P.E. 1990. 3-Dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **29**: 4129-4136.
- Elcock, A.H. 1998. The stability of salt bridges at high temperatures: Implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**: 489-502.
- Ermolenko, D.N., Richardson, J.M. and Makhatadze, G.I. 2003. Noncharged amino acid residues at the solvent-exposed positions in the middle and at the C-terminus of the  $\alpha$ -helix have the same helical propensity. *Protein Sci.* **12**: 1169-1176.
- Ermolenko, D.N., Thomas, S.T., Aurora, R., Gronenborn, A.M. and Makhatadze, G.I. 2002. Hydrophobic interaction at the C-cap position of the C-capping motif of  $\alpha$ -helix. *J. Mol. Biol.* **322**: 123-135.
- Errington, N. and Doig, A.J. 2005. A phosposerine salt bridge within an  $\alpha$ -helical peptide, the strongest  $\alpha$ -helix side-chain interaction measured to date. *Biochemistry* **44**: 7553-7558.
- Fernández-Recio, J., Vásquez, A., Civera, C., Sevilla, P. and Sancho, J. 1997. The tryptophan/histidine interaction in  $\alpha$ -helices. *J. Mol. Biol.* **267**: 184-197.
- Fodje, M.N. and Al-Karadaghi, S. 2002. Occurrence, conformational features and amino acid propensities for the  $\pi$ -helix. *Protein Eng.* **15**: 353-358.
- Forood, B., Feliciano, E.J. and Nambiar, K.P. 1993. Stabilization of  $\alpha$ -helical structures in short peptides via end capping. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 838-842.
- Gans, P.J., Lyu, P.C., Manning, M.C., Woody, R.W. and Kallenbach, N.R. 1991. The helix-coil transition in heterogeneous peptides with specific side-chain interactions - Theory and comparison with CD spectral data. *Biopolymers* **31**: 1605-1614.
- Ghadiri, M.R. and Choi, C. 1990. Secondary structure nucleation in peptides - Transition metal ion stabilized  $\alpha$ -helices. *J. Am. Chem. Soc.* **112**: 1630-1632.
- Goch, G., Maciejczyk, M., Oleszczuk, M., Stachowiak, D., Malicka, J. and Bierzynski, A. 2003. Experimental investigation of initial steps of helix propagation in model peptides. *Biochemistry* **42**: 6840-6847.
- Gong, Y., Zhou, H.X., Guo, M. and Kallenbach, N.R. 1995. Structural analysis of the N-termini and C-termini in a peptide with consensus sequence. *Protein Sci.* **4**: 1446-1456.
- Grauschopf, U., Winter, J.R., Korber, P., Zander, T., Dallinger, P. and Bardwell, J.C.A. 1995. Why is DsbA such an oxidizing disulfide catalyst? *Cell* **83**: 947-955.
- Greenfield, N.J. and Fasman, G.D. 1969. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**: 4108 - 4116.
- Gutin, A.M., Abkevich, V.I. and Shakhnovich, E.I. 1995. Is burst hydrophobic collapse necessary for protein folding? *Biochemistry* **34**: 3066-3076.
- Harper, E.T. and Rose, G.D. 1993. Helix stop signals in proteins and peptides: The capping box. *Biochemistry* **32**: 7605-7609.

- Hartl, F.U. and Hayer-Hartl, M. 2002. Protein folding - Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* **295**: 1852-1858.
- Hendsch, Z.S. and Tidor, B. 1994. Do salt bridges stabilize proteins - A continuum electrostatic analysis. *Protein Sci.* **3**: 211-226.
- Hol, W.G.J., M.Halie, L. and C.Sander. 1981. Dipoles of the  $\alpha$ -helix and  $\beta$ -sheet: Their role in protein folding. *Nature* **294**: 532-536.
- Horovitz, A., Matthews, J.M. and Fersht, A.R. 1992.  $\alpha$ -helix stability in proteins. 2. Factors that influence stability at an internal position. *J. Mol. Biol.* **227**: 560-568.
- Horovitz, A., Serrano, L., Avron, B., Bycroft, M. and Fersht, A.R. 1990. Strength and cooperativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* **216**: 1031-1044.
- Huang, C.Y., Klemke, J.W., Getahun, Z., DeGrado, W.F. and Gai, F. 2001. Temperature-dependent helix-coil transition of an alanine based peptide. *J. Am. Chem. Soc.* **123**: 9235-9238.
- Huyghues-Despointes, B.M.P. and Baldwin, R.L. 1997. Ion-pair and charged hydrogen-bond interactions between histidine and aspartate in a peptide helix. *Biochemistry* **36**: 1965-1970.
- Huyghues-Despointes, B.M.P., Klingler, T.M. and Baldwin, R.L. 1995. Measuring the strength of side-chain hydrogen bonds in peptide helices: The Gln-Asp (i, i + 4) interaction. *Biochemistry* **34**: 13267-13271.
- Huyghues-Despointes, B.M.P., Scholtz, J.M. and Baldwin, R.L. 1993a. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. *Protein Sci* **2**: 1604-1611.
- Huyghues-Despointes, B.M.P., Scholtz, J.M. and Baldwin, R.L. 1993b. Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci* **2**: 80-85.
- Inaba, K. and Ito, K. 2002. Paradoxical redox properties of DsbB and DsbA in the protein disulfide-introducing reaction cascade. *Embo J.* **21**: 2646-2654.
- Jackson, D.Y., King, D.S., Chmielewski, J., Singh, S. and Schultz, P.G. 1991. General approach to the synthesis of short  $\alpha$ -helical peptides. *J. Am. Chem. Soc.* **113**: 9391-9392.
- Janin, J., Wodak, S.J., Levit, M. and Maigret, B. 1978. Conformation of amino acids side-chain in proteins. *J. Mol. Biol.* **125**: 357-386.
- Jasanoff, A. and Fersht, A.R. 1994. Quantitative determination of helical propensities from trifluoroethanol titration curves. *Biochemistry* **33**: 2129-2135.
- Jeng, M.-F., Campbell, A.P., Begley, T., Holmgren, A., Case, D.A., Wright, P.E. and Dyson, H.J. 1994. High-resolution solution structures of oxidized and reduced *Escherichia coli* thioredoxin. *Structure* **2**: 853-868.
- Jernigan, R., Raghunathan, G. and Bahar, I. 1994. Characterisation of interactions and metal-ion binding-sites in proteins. *Curr. Opin. Struc. Biol.* **4**: 256-263.
- Jiménez, M.A., Muñoz, V., Rico, M. and Serrano, L. 1994. Helix stop and start signals in peptides and proteins. The capping box does not necessarily prevent helix elongation. *J. Mol. Biol.* **242**: 487-496.

- Kallenbach, N.R. and Gong, Y.X. 1999. C-terminal capping motifs in model helical peptides. *Bioorg. Med. Chem.* **7**: 143-151.
- Kapurniotu, A. and Taylor, J.W. 1995. Structural and conformational requirements for human calcitonin activity - Design, synthesis, and study of lactam-bridged analogs. *J. Med. Chem.* **38**: 836-847.
- Karpen, M.E., De Haset, P., L. and Neet, K.E. 1992. Differences in the amino acid distributions of  $3_{10}$ -helices and  $\alpha$ -helices. *Protein Sci.* **1**: 1333-1342.
- Kelley, R.F., Shalongo, W., Jagganadham, M.V. and Stellwagen, E. 1987. Equilibrium and kinetic measurements of the conformational transition of reduced thioredoxin. *Biochemistry* **26**: 1406-1411.
- Kennedy, R.J., Tsang, K.-Y. and Kemp, D.S. 2002. Consistent helicities from CD and template t/c data for n-templated polyalanines: Progress toward resolution of the alanine helicity problem. *J. Am. Chem. Soc.* **124**: 934-944.
- Kentsis, A. and Sosnick, T.R. 1998. Trifluoroethanol promotes helix formation by destabilizing backbone exposure: Desolvation rather than native hydrogen bonding defines the kinetic pathway of dimeric coiled coil folding. *Biochemistry* **37**: 14613-14622.
- Kim, P.S. and Baldwin, R.L. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**: 459-489.
- Kise, K.J. and Bowler, B.E. 2002. Induction of helical structure in a heptapeptide with a metal cross-link: Modification of the Lifson-Roig helix-coil theory to account for covalent cross-links. *Biochemistry* **41**: 15826-15837.
- Kortemme, T. and Creighton, T.E. 1995. Ionization of cysteine residues at the termini of model  $\alpha$ -helical peptides - Relevance to unusual thiol pKa values in proteins of the thioredoxin family. *J. Mol. Biol.* **253**: 799-812.
- Kortemme, T., Darby, N.J. and Creighton, T.E. 1996. Electrostatic interactions in the active site of the N-terminal thioredoxin-like domain of protein disulfide isomerase. *Biochemistry* **35**: 14503-14511.
- Kumar, S. and Bansal, M. 1998. Dissecting  $\alpha$ -helices: Position specific analysis of  $\alpha$ -helices in globular proteins. *Proteins* **31**: 460-476.
- Kumar, S., Ma, B.Y., Tsai, C.J. and Nussinov, R. 2000. Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins* **38**: 368-383.
- Kumar, S. and Nussinov, R. 2002. Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophys. J.* **83**: 1595-1612.
- Kyte, J. 1995. Structure in protein chemistry. *Garland, New York*.
- Lacroix, E., Viguera, A.R. and Serrano, L. 1998. Elucidating the folding problem of  $\alpha$ -helices: Local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**: 173-191.
- Lee, K.H., Benson, D.R. and Kuczera, K. 2000. Transitions from  $\alpha$ - to  $\pi$ -helix observed in molecular dynamics simulations of synthetic peptides. *Biochemistry* **39**: 13737-13747.
- Levinthal, C. 1968. Are there pathways for protein folding? *J. Chim. Phys.* **65**: 44-45.

- Liehr, S. and Chenault, H.K. 1999. A comparison of the  $\alpha$ -helix forming propensities and hydrogen bonding properties of serine phosphate and  $\alpha$ -amino- $\gamma$ -phosphonobutyric acid. *Bioorg. Med. Chem. Lett.* **9**: 2759-2762.
- Liff, M.I., Lyu, P.C. and Kallenbach, N.R. 1991. Analysis of asymmetry in the distribution of helical residues in peptides by H-1 nuclear magnetic resonance. *J. Am. Chem. Soc.* **113**: 1014-1019.
- Lifson, S. 1961. Theory of helix-coil transition in polypeptides. *J. Chem. Phys.* **34**: 1963-&.
- Lockhart, D.J. and Kim, P.S. 1993. Electrostatic screening of charge and dipole interactions with the helix backbone. *Science* **260**: 198-202.
- Lovel, S.C., Word, J.M., Richardson, J.S. and Richardson, D.C. 2000. The penultimate rotamer library. *Proteins* **40**: 389-408.
- Low, B.W. and Baybutt, R.B. 1952. The  $\pi$ -helix - A hydrogen bonded configuration of the polypeptide chain. *J. Am. Chem. Soc.* **74**: 5806-5807.
- Low, B.W. and Grenville-Wells, H.J. 1953. Generalized mathematical relations for polypeptide chain helices. The coordinates for the  $\pi$ -helix. *Proc. Nat. Acad. Sci. U.S.A.* **39**: 785-801.
- Lundstrom, J. and Holmgren, A. 1993. Determination of reduction-oxidation potential of the thioredoxin-like domains of protein disulfide-isomerase from the equilibrium with guthathione and thioreoxin. *Biochemistry* **32**: 6649-6655.
- Luo, P.Z. and Baldwin, R.L. 1999. Interaction between water and polar groups of the helix backbone: An important determinant of helix propensities. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 4930-4935.
- Luo, P.Z., Braddock, D.T., Subramanian, R.M., Meredith, S.C. and Lynn, D.G. 1994. Structural and thermodynamic characterization of a bioactive peptide model of apolipoprotein-E - Side-chain lactam bridges to constrain the conformation. *Biochemistry* **33**: 12367-12377.
- Luo, Y.Z. and Baldwin, R.L. 1998. Trifluoroethanol stabilizes the pH 4 folding intermediate of sperm whale apomyoglobin. *J. Mol. Biol.* **279**: 49-57.
- Lyu, P.C., Liff, M.I., Marky, L.A. and Kallenbach, N.R. 1990. Side-chain contributions to the stability of  $\alpha$ -helical structure in peptides. *Science* **250**: 669-673.
- Lyu, P.C., Marky, L.A. and Kallenbach, N.R. 1989. The role of ion-pairs in  $\alpha$ -helix stability-2. New designed helical peptides. *J. Am. Chem. Soc.* **111**: 2733-2734.
- Lyu, P.C., Sherman, J.C., Chen, A. and Kallenbach, N.R. 1991.  $\alpha$ -helix stabilization by natural and unnatural amino-acids with alkyl side-chains. *Proc. Natl. Acad. Sci. U. S. A.* **88**: 5317-5320.
- Lyu, P.C., Zhou, H.X.X., Jelveh, N., Wemmer, D.E. and Kallenbach, N.R. 1992. Position-dependent stabilizing effects in  $\alpha$ -helices - N-terminal capping in synthetic model peptides. *J. Am. Chem. Soc.* **114**: 6560-6562.
- Makhatadze, G.I., Loladze, V.V., Ermolenko, D.N., Chen, X.F. and Thomas, S.T. 2003. Contribution of surface salt bridges to protein stability: Guidelines for protein engineering. *J. Mol. Biol.* **327**: 1135-1148.
- Marqusee, S. and Baldwin, R.L. 1987. Helix stabilization by Glu<sup>-</sup> - Lys<sup>+</sup> salt bridges in short peptides of denovo design. *Proc. Natl. Acad. Sci. U. S. A.* **84**: 8898-8902.



- Marqusee, S. and Sauer, R.T. 1994. Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in  $\lambda$  repressor. *Protein Sci.* **3**: 2217-2225.
- Martin, J.L. 1995. Thioredoxin - A fold for all reasons. *Structure* **3**: 245-250.
- Matthews, C.R. 1993. Pathways of protein folding. *Annu. Rev. Biochem.* **62**: 653-683.
- Mayne, L., Englander, S.W., Qiu, R., Yang, J.X., Gong, Y.X., Spek, E.J. and Kallenbach, N.R. 1998. Stabilizing effect of a multiple salt bridge in a prenucleated peptide. *J. Am. Chem. Soc.* **120**: 10643-10645.
- McGregor, M.J., Islam, S.A. and Sternberg, M.J.E. 1987. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**: 295-310.
- Merrifield, R.B. 1963. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**: 2149-2154.
- Merutka, G. and Stellwagen, E. 1991. Effect of amino acid ion pairs on peptide helicity. *Biochemistry* **30**: 1591-1594.
- Millhauser, G.L. 1995. Views of helical peptides - A proposal for the position of  $3_{10}$ -helix along the thermodynamic folding pathway. *Biochemistry* **34**: 3872-3877.
- Miranda, J.J. 2003. Position-dependent interactions between cysteine and the helix dipole. *Protein Sci.* **12**: 73-81.
- Moore, E.C., Reichard, P. and Thelander, L. 1964. Enzymatic synthesis of deoxyribonucleotides. V. Purification and properties of thioredoxin reductase from *Escherichia coli* b. *J. Biol. Chem.* **239**: 3445-3452.
- Morgan, D.M., Lynn, D.G., Miller-Auer, H. and Meredith, S.C. 2001. A designed  $Zn^{2+}$ -binding amphiphilic polypeptide: Energetic consequences of  $\pi$ -helicity. *Biochemistry* **40**: 14020-14029.
- Mossner, E., Huber-Wunderlich, M., Rietsch, A., Beckwith, J., Glockshuber, R. and Aslund, F. 1999. Importance of redox potential for the in vivo function of the cytoplasmic disulfide reductant thioredoxin from *Escherichia coli*. *J. Biol. Chem.* **274**: 25254-25259.
- Mossner, E., Iwai, H. and Glockshuber, R. 2000. Influence of the pKa value of the buried, active-site cysteine on the redox properties of thioredoxin-like oxidoreductases. *FEBS Lett.* **477**: 21-26.
- Muñoz, V. and Serrano, L. 1994. Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* **1**: 399-409.
- Muñoz, V. and Serrano, L. 1995a. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245**: 275-296.
- Muñoz, V. and Serrano, L. 1995b. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J. Mol. Biol.* **245**: 297-308.
- Muñoz, V. and Serrano, L. 1995c. The hydrophobic-staple motif and a role for loop residues in  $\alpha$ -helix stability and protein folding. *Nat. Struct. Biol.* **2**: 380-385.

- Muñoz, V. and Serrano, L. 1997. Development of the multiple sequence approximation within the agadir model of  $\alpha$ -helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* **41**: 495-509.
- Musafia, B., Buchner, V. and Arad, D. 1995. Complex salt bridges in proteins - Statistical analysis of structure and function. *J. Mol. Biol.* **254**: 761-770.
- Myers, J.K. and Pace, C.N. 1996. Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **71**: 2033-2039.
- Myers, J.K., Pace, C.N. and Scholtz, J.M. 1997. Helix propensities are identical in proteins and peptides. *Biochemistry* **36**: 10923-10929.
- Myers, J.K., Pace, N. and Scholtz, J.M. 1998. Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1. *Protein Sci.* **7**: 383-388.
- Nelson, J.W., Isaacson, D. and Kallenbach, N.R. 1986. Effects of trifluoroethanol on the charge effect of RNase s-peptide. *Biophys. J.* **49**: A492-A492.
- Nelson, J.W. and Kallenbach, N.R. 1989. Persistence of the  $\alpha$ -helix stop signal in the s-peptide in trifluoroethanol solutions. *Biochemistry* **28**: 5256-5261.
- Némethy, G., Phillips, D.C., Leach, S.J. and Scheraga, H.A. 1967. A second right-handed helical structure with the parameters of the Pauling-Corey  $\alpha$ -helix. *Nature* **214**: 363-365.
- Newton, A.C. 2003. Regulation of the ABC kinases by phosphorylation: Protein kinase C as a paradigm. *Biochem. J.* **370**: 361-371.
- Nguyen, M.T., Beck, J., Lue, H., Fiunfzig, H., Kleemann, R., Koolwijk, P., Kapurniotu, A. and Bernhagen, J. 2003. A 16-residue peptide fragment of macrophage migration inhibitory factor, MIF-(50-65), exhibits redox activity and has MIF-like biological functions. *J. Biol. Chem.* **278**: 33654-33671.
- Nozaki, Y. and Tanford, C. 1967. Intrinsic dissociation constants of aspartyl and glutamyl carboxyl groups. *J. Biol. Chem.* **242**: 4731-4735.
- Olson, C.A., Shi, Z.S. and Kallenbach, N.R. 2001a. Polar interactions with aromatic side chains in  $\alpha$ -helical peptides: CH...OH-bonding and cation- $\pi$  interactions. *J. Am. Chem. Soc.* **123**: 6451-6452.
- Olson, C.A., Spek, E.J., Shi, Z.S., Vologodskii, A. and Kallenbach, N.R. 2001b. Cooperative helix stabilization by complex Arg-Glu salt bridges. *Proteins* **44**: 123-132.
- O'Neil, K.T. and DeGrado, W.F. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250**: 646-651.
- Ookura, T., Kainuma, K., Kim, H., Otaka, A., Fujii, N. and Kawamura, Y. 1995. Active site peptides with CXXC motif on map-resin can mimic protein disulfide isomerase activity. *Biochem. Biophys. Res. Com.* **213**: 746-751.
- Opella, S.J., DeSilva, T.M. and Veglia, G. 2002. Structural biology of metal-binding sequences. *Curr. Opin. Chem. Biol.* **6**: 217-223.
- Osapay, G. and Taylor, J.W. 1990. Multicyclic polypeptide model compounds .1. Synthesis of a tricyclic amphiphilic  $\alpha$ -helical peptide using an oxime resin, segment-condensation approach. *J. Am. Chem. Soc.* **112**: 6046-6051.

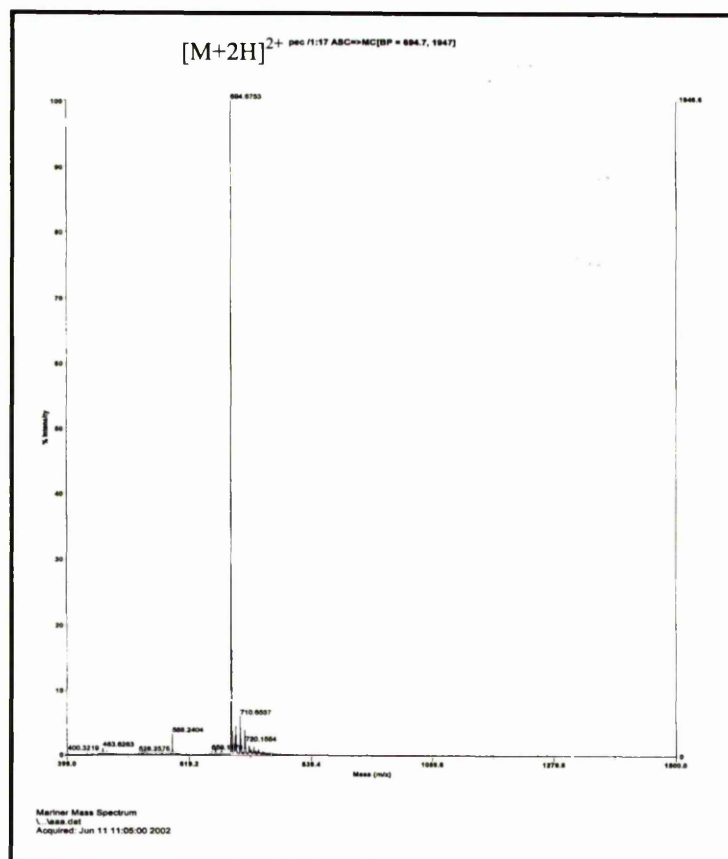
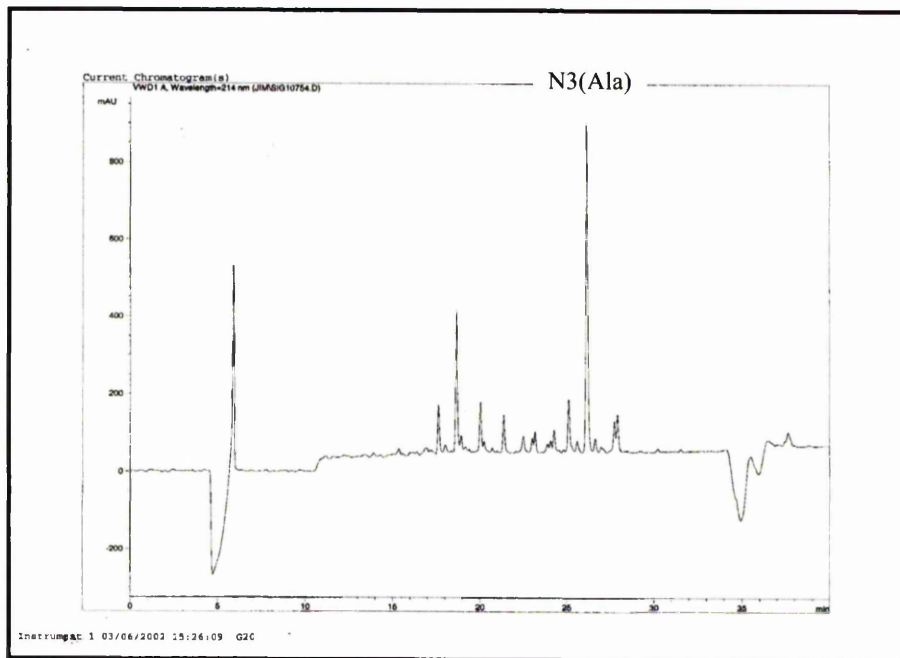
- Osapay, G. and Taylor, J.W. 1992. Multicyclic polypeptide model compounds .2. Synthesis and conformational properties of a highly  $\alpha$ -helical uncosapeptide constrained by 3 side-chain to side-chain lactam bridges. *J. Am. Chem. Soc.* **114**: 6966-6973.
- Pabo, C.O., Peisach, E. and Grant, R.A. 2001. Design and selection of novel Cys(2)His(2) zinc finger proteins. *Annu. Rev. Biochem.* **70**: 313-340.
- Pace, C.N. and Scholtz, J.M. 1998.  $\alpha$ -helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75**: 422-427.
- Padmanabhan, S. and Baldwin, R.L. 1994a. Helix-stabilizing interaction between tyrosine and leucine or valine when the spacing is  $i,i+4$ . *J. Mol. Biol.* **241**: 706-713.
- Padmanabhan, S. and Baldwin, R.L. 1994b. Tests for helix-stabilizing interactions between various nonpolar side chains in alanine-based peptides. *Protein Sci* **3**: 1992-1997.
- Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T.M. and Baldwin, R.L. 1990. Relative helix-forming tendencies of nonpolar amino acids. *Nature* **344**: 268-270.
- Park, S.H., Shalongo, W. and Stellwagen, E. 1993. Residue helix parameters obtained from dichroic analysis of peptides of defined sequence. *Biochemistry* **32**: 7048-7053.
- Pauling, L., Corey, R.B. and Branson, H.R. 1951. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **37**.
- Penel, S. and Doig, A.J. 2001. Rotamer strain energy in protein helices quantification of a major force opposing protein folding. *J. Mol. Biol.* **305**: 961-968.
- Penel, S., Hughes, E. and Doig, A.J. 1999. Side-chain structures in the first turn of the  $\alpha$ -helix. *J. Mol. Biol.* **287**: 127-143.
- Petukhov, M., Munoz, V., Yumoto, N., Yoshikawa, S. and Serrano, L. 1998. Position dependence of non-polar amino acid intrinsic helical propensities. *J. Mol. Biol.* **278**: 279-289.
- Petukhov, M., Uegaki, K., Yumoto, N. and Serrano, L. 2002. Amino acid intrinsic  $\alpha$ -helical propensities III: Positional dependence at several positions of C-terminus. *Protein Sci.* **11**: 766-777.
- Petukhov, M., Uegaki, K., Yumoto, N., Yoshikawa, S. and Serrano, L. 1999. Position dependence of amino acid intrinsic helical propensities II: Non-charged polar residues: Ser, Thr, Asn, and Gln. *Protein Sci.* **8**: 2144-2150.
- Petukhov, M., Yumoto, N., Murase, S., Onmura, R. and Yoshikawa, S. 1996. Factors that affect the stabilization of  $\alpha$ -helices in short peptides by a capping box. *Biochemistry* **35**: 387-397.
- Pickett, S.D. and Sternberg, M.J.E. 1993. Empirical scale of side-chain conformational entropy in protein-folding. *J. Mol. Biol.* **231**: 825-839.
- Presta, L.G. and Rose, G.D. 1988. Helix signals in proteins. *Science* **240**: 1632-1641.
- Prieto, J. and Serrano, L. 1997. C-capping and helix stability: The Pro C-capping motif. *J. Mol. Biol.* **274**: 276-288.
- Ptitsyn, O.B. and Rashin, A.A. 1975. A model of myoglobin self-organization. *Biophys. Chem.* **3**: 1-20.
- Qian, H. and Schellman, J.A. 1992. Helix/coil theories: A comparative study for finite length polypeptides. *J. Phys. Chem.* **96**: 3987-3994.

- Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**: 95-99.
- Ramachandran, G.N. and Sasisekharan, V. 1968. Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**: 283-438.
- Reiersen, H. and Rees, A.R. 2000. Trifluoroethanol may form a solvent matrix for assisted hydrophobic interactions between peptide side-chains. *Protein Eng.* **13**: 739-743.
- Richardson, J.S. and Richardson, D.C. 1988. Amino acid preferences for specific locations at the ends of  $\alpha$ -helices. *Science* **240**: 1648-1652.
- Rohl, C.A. and Baldwin, R.L. 1998. Deciphering rules of helix stability in peptides. In *Energetics of biological macromolecules, pt. b*, pp. 1-26.
- Rohl, C.A., Chakrabarty, A. and Baldwin, R.L. 1996. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5**: 2623-2637.
- Romero-Isart, N. and Vasak, M. 2002. Advances in the structure and chemistry of metallothioneins. *J. Inorg. Biochem.* **88**: 388-396.
- Ruan, F., Chen, Y. and Hopkins, P.B. 1990. Metal ion enhanced helicity in synthetic peptides containing unnatural, metal-ligating residues. *J. Am. Chem. Soc.* **112**: 9403-9404.
- Schellman, C. 1980. The  $\alpha$ <sub>c</sub> conformation at the ends of helices. *Protein Fold.*: 53-61.
- Scholtz, J.M., Hong, Q., York, E.J., Stewart, J.M. and Baldwin, R.L. 1991a. Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* **31**: 1463-1470.
- Scholtz, J.M., Marqusee, S., Baldwin, R.L., York, E.J., Stewart, J.M., Santoro, M. and Bolen, D.W. 1991b. Calorimetric determination of the enthalpy change for the  $\alpha$ -helix to coil transition of an alanine peptide in water. *Proc. Natl. Acad. Sci. U. S. A.* **88**: 2854-2858.
- Scholtz, J.M., Qian, H., Robbins, V.H. and Baldwin, R.L. 1993. The energetics of ion-pair and hydrogen-bonding interactions in  $\alpha$ -helical peptide. *Biochemistry* **32**: 9668-9676.
- Seale, J.W., Srinivasan, R. and Rose, G.D. 1994. Sequence determinants of the capping box, a stabilizing motif at the N-termini of  $\alpha$ -helices. *Protein Sci* **3**: 1741-1745.
- Serrano, L. and Fersht, A.R. 1989. Capping and  $\alpha$ -helix stability. *Nature* **342**: 296-299.
- Shalongo, W. and Stellwagen, E. 1995. Incorporation of pairwise interactions into the Lifson-Roig model for helix prediction. *Protein Sci.* **4**: 1161-1166.
- Shi, Z.S., Olson, C.A., Bell, A.J. and Kallenbach, N.R. 2002a. Non-classical helix-stabilizing interactions: CH...OH-bonding between Phe and Glu side chains in  $\alpha$ -helical peptides. *Biophys. Chem.* **101**: 267-279.
- Shi, Z.S., Olson, C.A. and Kallenbach, N.R. 2002b. Cation- $\pi$  interaction in model  $\alpha$ -helical peptides. *J. Am. Chem. Soc.* **124**: 3284-3291.
- Shirley, W.A. and Brooks, C.L. 1997. Curious structure in "canonical" alanine based peptides. *Proteins* **28**: 59-71.

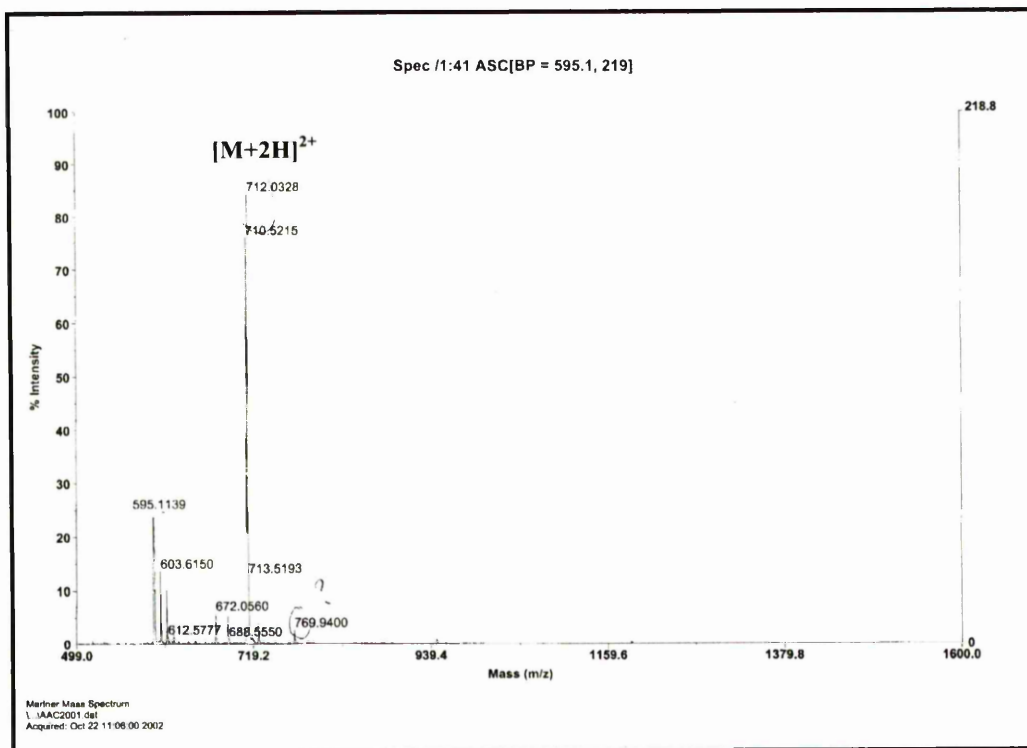
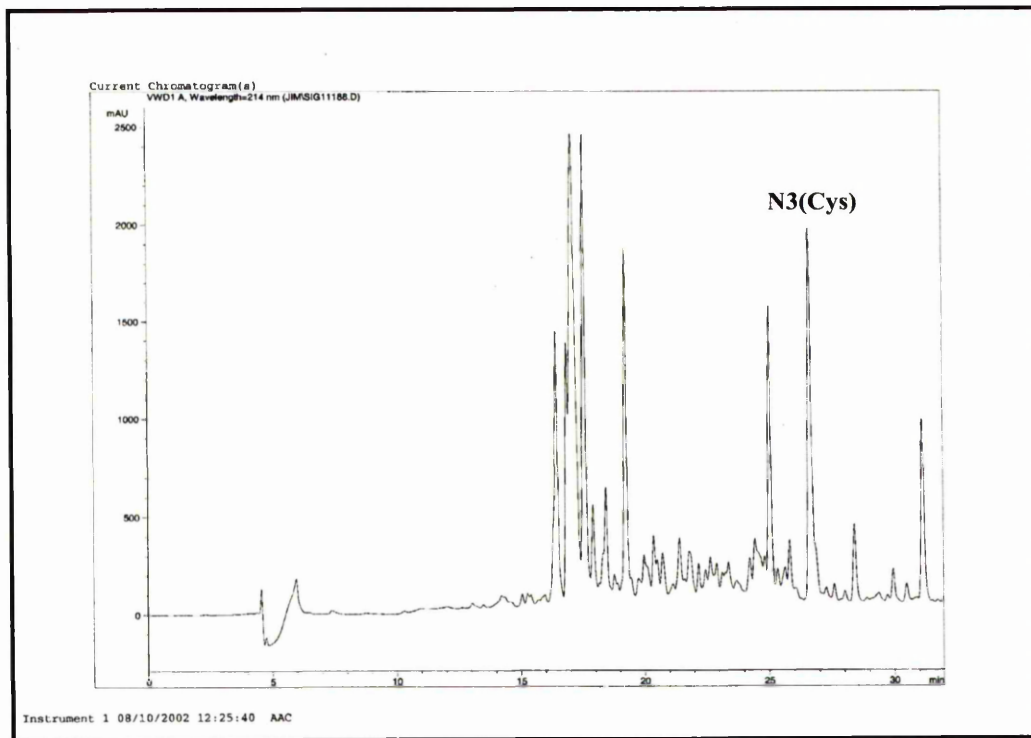
- Siedlecka, M., Goch, G., Ejchart, A., Sticht, H. and Bierzynski, A. 1999.  $\alpha$ -helix nucleation by a calcium-binding peptide loop. *Proc. Natl. Acad. Sci. USA* **99**: 903-908.
- Smith, D. 1989. Sstruc a computer program, London.
- Smith, J.S. and Scholtz, J.M. 1998. Energetics of polar side-chain interactions in helical peptides: Salt effects on ion pairs and hydrogen bonds. *Biochemistry* **37**: 33-40.
- Sonnichsen, F.D., Van Eyk, J.E., Hodges, R.S. and Sykes, B.D. 1992. Effect of trifluoroethanol on protein secondary structure: An NMR and CD study using a synthetic actin peptide. *Biochemistry* **31**: 8790-8798.
- Spek, E.J., Bui, A.H., Lu, M. and Kallenbach, N.R. 1998. Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Sci.* **7**: 2431-2437.
- Stapley, B.J. and Doig, A.J. 1997a. Free energies of amino acid side-chain rotamers in  $\alpha$ -helices,  $\beta$ -sheets and  $\alpha$ -helix N-caps. *J. Mol. Biol.* **272**: 456-464.
- Stapley, B.J. and Doig, A.J. 1997b. Hydrogen bonding interactions between glutamine and asparagine in  $\alpha$ -helical peptides. *J. Mol. Biol.* **272**: 465-473.
- Stapley, B.J., Rohl, C.A. and Doig, A.J. 1995. Addition of side-chain interactions to modified Lifson-Roig helix-coil theory - Application to energetics of phenylalanine-methionine interactions. *Protein Sci.* **4**: 2383-2391.
- Starzyk, A., Barber-Armstrong, W., Sridharan, M. and Decatur, S.M. 2005. Spectroscopic evidence for backbone desolvation of helical peptides by 2,2,2-trifluoroethanol: An isotope-edited FTIR study. *Biochemistry* **44**: 369-376.
- Stellwagen, E., Park, S.-H., Shalongo, W. and Jain, A. 1992. The contribution of residue ion pairs to the helical stability of a model peptide. *Biopolymers* **32**: 1193-1200.
- Sun, J.K. and Doig, A.J. 1998. Addition of side-chain interactions to  $3_{10}$ -helix/coil and  $\alpha$ -helix/ $3_{10}$ -helix/coil theory. *Protein Sci.* **7**: 2374-2383.
- Sun, J.K., Penel, S. and Doig, A.J. 2000. Determination of  $\alpha$ -helix N1 energies after addition of N1, N2, and N3 preferences to helix/coil theory. *Protein Sci.* **9**: 750-754.
- Szalik, L., Moitra, J., Krylov, D. and Vinson, C. 1997. Phosphorylation destabilizes  $\alpha$ -helices. *Nat. Struct. Biol.* **4**: 112-114.
- Takano, K., Tsuchimori, K., Yamagata, Y. and Yutani, K. 2000. Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry* **39**: 12375-12381.
- Thomas, S.T., Loladze, V.V. and Makhatadze, G.I. 2001. Hydration of the peptide backbone largely defines the thermodynamic propensity scale of residues at the C' position of the C-capping box of  $\alpha$ -helices. *Proc. Natl. Acad. Sci. U. S. A.* **98**: 10670-10675.
- Tissot, A.C., Vuilleumier, S. and Fersht, A.R. 1996. Importance of two buried salt bridges in the stability and folding pathway of barnase. *Biochemistry* **35**: 6786-6794.
- Toniolo, C. and Benedetti, E. 1991. The polypeptide  $3_{10}$ -helix. *Trends Biochem. Sci.* **16**: 350-353.
- Toth, G., Watts, C.R., Murphy, R.F. and Lovas, S. 2001. Significance of aromatic-backbone amide interactions in protein structure. *Proteins* **43**: 373-381.

- Tsou, L.K., Tatko, C.D. and Waters, M.L. 2002. Simple cation- $\pi$  interaction between a phenyl ring and a protonated amine stabilizes an  $\alpha$ -helix in water. *J. Am. Chem. Soc.* **124**: 14917-14921.
- Viguera, A.R. and Serrano, L. 1995a. Experimental analysis of the Schellman motif. *J. Mol. Biol.* **251**: 150-160.
- Viguera, A.R. and Serrano, L. 1995b. Side-chain interactions between sulfur-containing amino acids and phenylalanine in  $\alpha$ -helices. *Biochemistry* **34**: 8771-8779.
- Viguera, A.R. and Serrano, L. 1999. Stable proline box motif at the N-terminal end of  $\alpha$ -helices. *Protein Sci* **8**: 1733-1742.
- Vijayakumar, M., Qian, H. and Zhou, H.X. 1999. Hydrogen bonds between short polar side chains and peptide backbone: Prevalence in proteins and effects on helix-forming propensities. *Proteins* **34**: 497-507.
- Vila, J., Williams, R.L., Grant, J.A., Wojcik, J. and Scheraga, H.A. 1992. The intrinsic helix-forming tendency of L-alanine. *Proc. Nat. Acad. Sci. U.S.A.* **89**: 7821-7825.
- Vila, J.A., Ripoll, D.R. and Scheraga, H.A. 2000. Physical reasons for the unusual  $\alpha$ -helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides. *Proc. Nat. Acad. Sci. U.S.A.* **97**: 13075-13079.
- Walgers, R., Lee, T.C. and Cammers-Goodwin, A. 1998. An indirect chaotropic mechanism for the stabilization of helix conformation of peptides in aqueous trifluoroethanol and hexafluoro-2-propanol. *J. Am. Chem. Soc.* **120**: 5073-5079.
- Wan, W.Y. and Milner-White, E.J. 1999. A natural grouping of motifs with an aspartate or asparagine residue forming two hydrogen bonds to residues ahead in sequence: Their occurrence at  $\alpha$ -helical N-termini and in other situations. *J. Mol. Biol.* **286**: 1633-1649.
- Waterhous, D.V. and Johnson, W.C. 1994. Importance of environment in determining secondary structure in proteins. *Biochemistry* **33**: 2121-2128.
- Watson, W., H., Pohl, J., Montfort, W., R., Stuchlik, O., Reed, M., S., Powis, G. and Jones, D., P. 2003. Redox potential of human thioredoxin 1 and identification of a second dithiol/disulfide motif. *J. Biol. Chem.* **278**: 33408-33415.
- Weaver, T.M. 2000. The  $\pi$ -helix translates structure into function. *Protein Sci.* **9**: 201-206.
- Wedemeyer, W.J., Welker, E., Narayan, M. and Scheraga, H.A. 2000. Disulfide bonds and protein folding. *Biochemistry* **39**: 4207-4216.
- Weichsel, A., Gasdaska, J.R., Powis, G. and Montfort, W.R. 1996. Crystal structures of reduced, oxidized, and mutated human thioredoxins: Evidence for a regulatory homodimer. *Structure* **4**: 735-751.
- Wetlaufer, D.B. 1973. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **70**: 697-701.
- White, P.D. and Chan, W.C. 2000. *Fmoc solid phase peptide synthesis. A practical approach*. Oxford University Press, Oxford, pp. 9-40.
- Wojcik, J., Altmann, K.H. and Scheraga, H.A. 1990. Helix-coil stability constants for the naturally occurring amino acids in water. XXIV. Half-cysteine parameters from random poly-(hydroxybutylglutamine-co-s-methylthio-l-cysteine). *Biopolymers* **30**: 121-134.

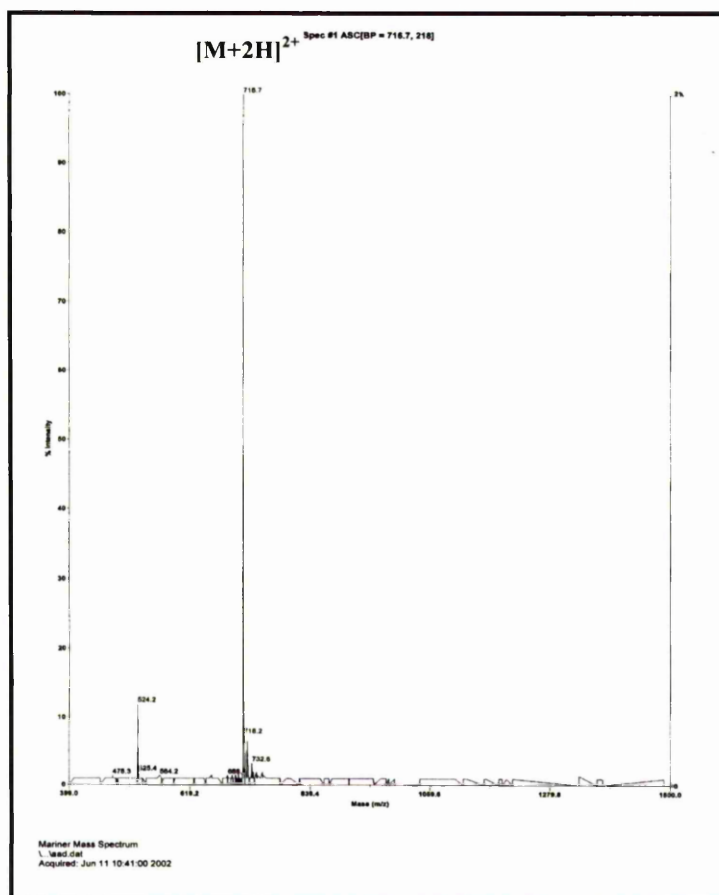
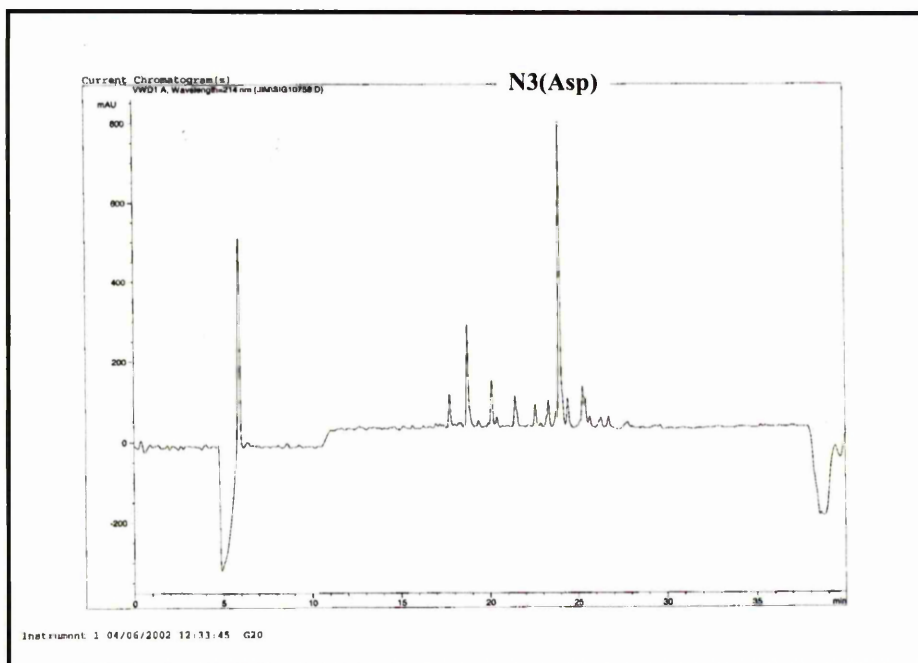
- Wunderlich, M. and Glockshuber, R. 1993. Redox properties of protein disulfide isomerase (DsbA) from *Escherichia coli*. *Protein Sci* **2**: 717-726.
- Yang, A.S. and Honig, B. 1995. Free-energy determinants of secondary structure formation. 1.  $\alpha$ -helices. *J. Mol. Biol.* **252**: 351-365.
- Yoder, G., Pancoska, P. and Keiderling, T.A. 1997. Characterization of alanine-rich peptides, Ac-(AAKAA) $_n$ -GYNH $_2$  ( $n=1-4$ ), using vibrational circular dichroism and fourier transform infrared conformational determination and thermal unfolding. *Biochemistry* **36**: 5123-5133.
- Zapun, A., Bardwell, J.C.A. and Creighton, T.E. 1993. The reactive and destabilizing disulfide bond of DsbA, a protein required for protein disulfide bond formation in vivo. *Biochemistry* **32**: 5083-5092.
- Zhang, W.T. and Taylor, J.W. 1996. Efficient solid-phase synthesis of peptides with tripodal side-chain bridges and optimization of the solvent conditions for solid-phase cyclizations. *Tetrahedron Lett.* **37**: 2173-2176.
- Zhou, H.X.X., Lyu, P.C., Wemmer, D.E. and Kallenbach, N.R. 1994.  $\alpha$ -helix capping in synthetic model peptides by reciprocal side-chain main-chain interactions - Evidence for an N-terminal capping box. *Proteins* **18**: 1-7.
- Zhou, R., Huang, X., Margulis, C.J. and Berne, B.J. 2004. Hydrophobic collapse in multidomain protein folding. *Science* **305**: 1605-1609.
- Zimm, B.H. and Bragg, J.K. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* **31**: 526-535.

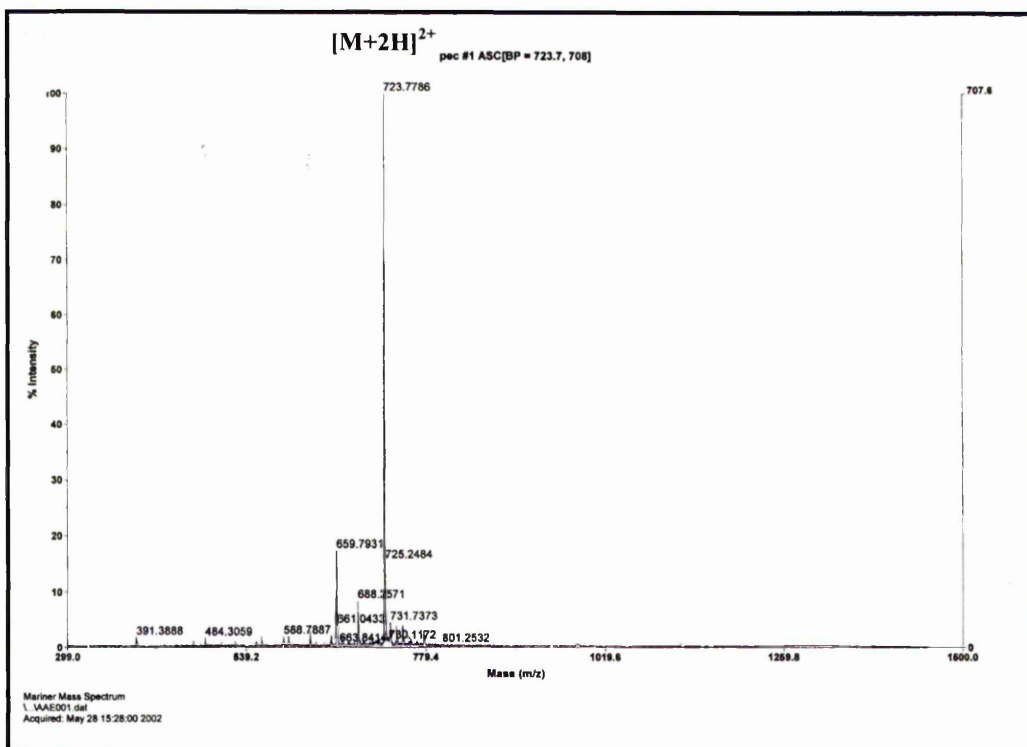
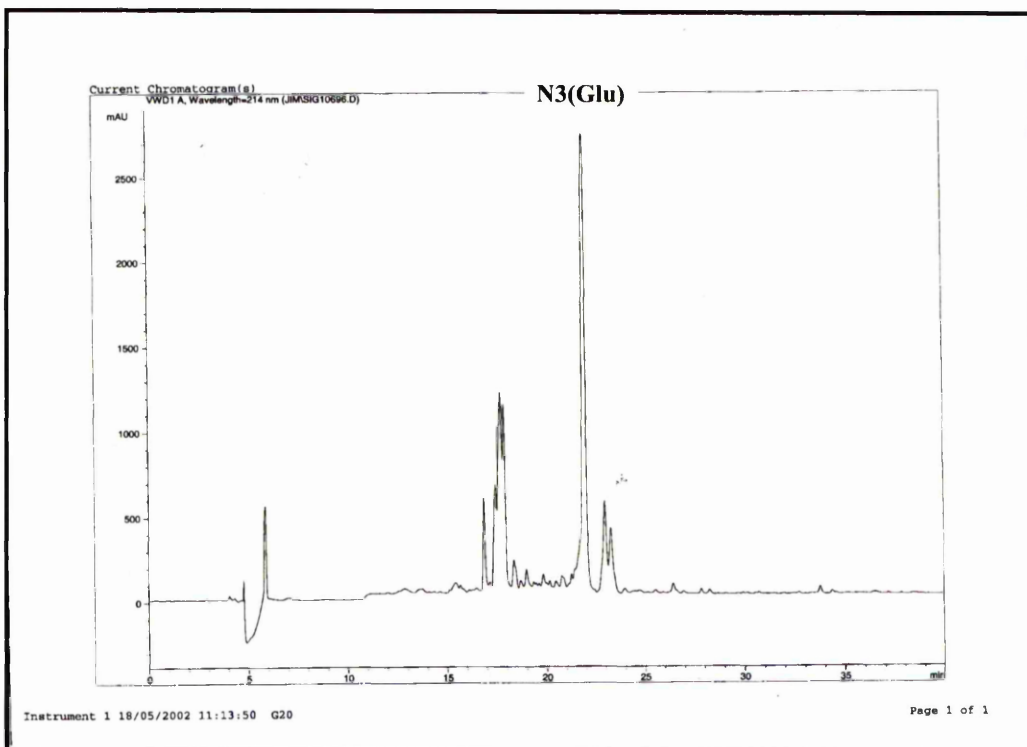
**APPENDIX A. HPLC and MS results to check identity of peptides****A.1. Identity confirmation of Ac-AAAAAAKAAAKAGY-NH<sub>2</sub> peptide**

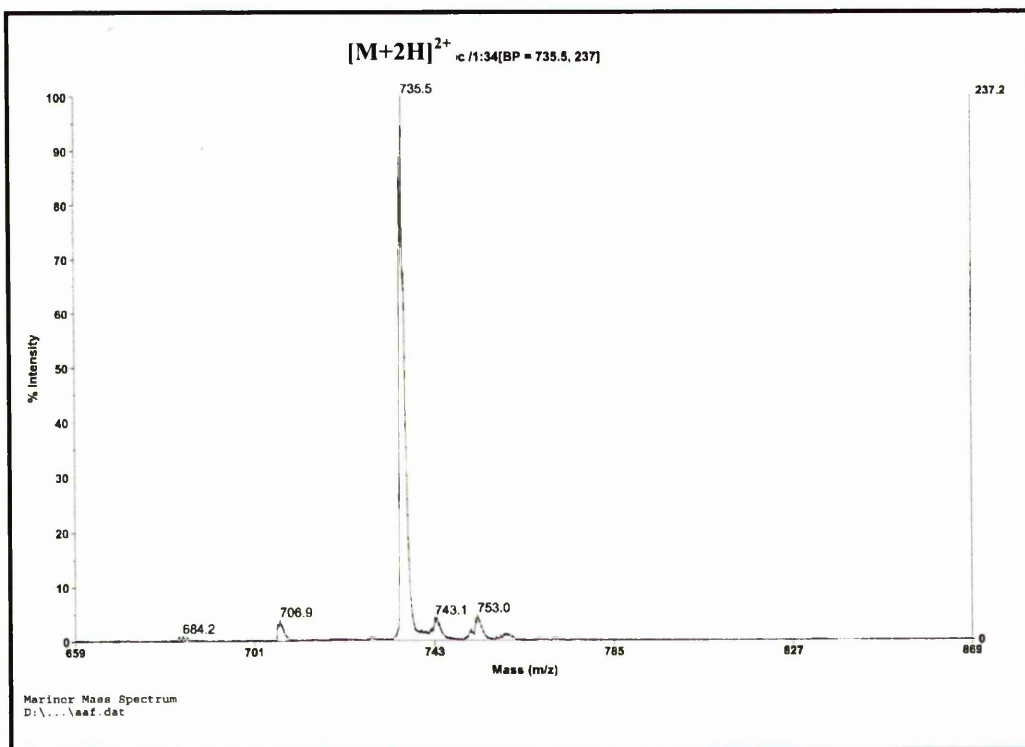
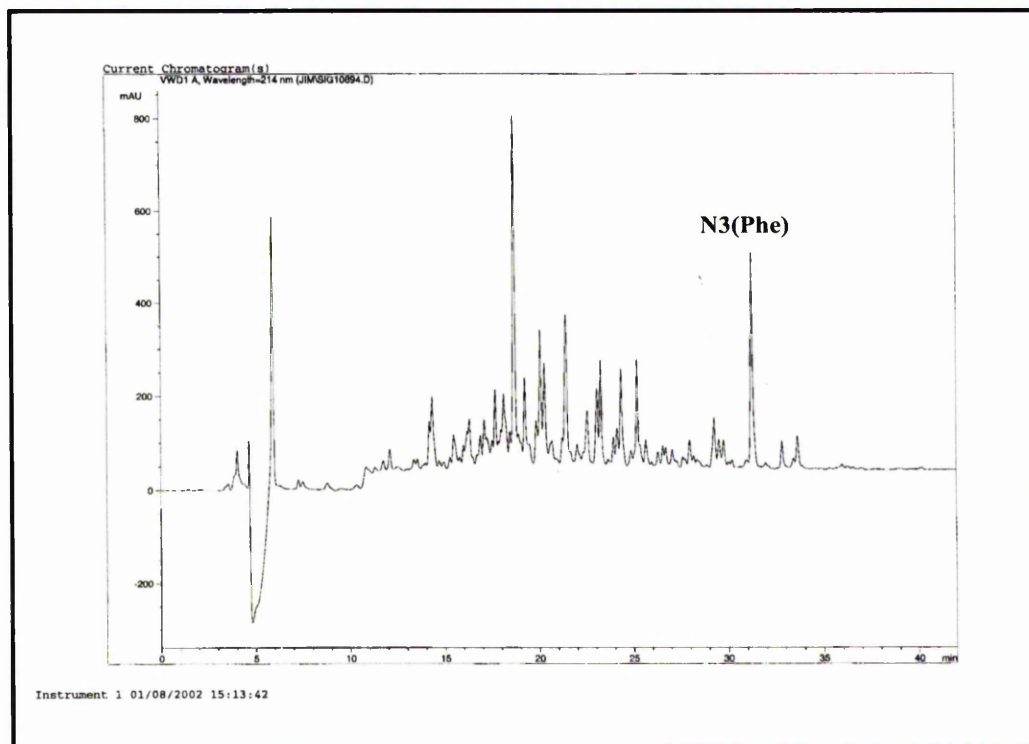


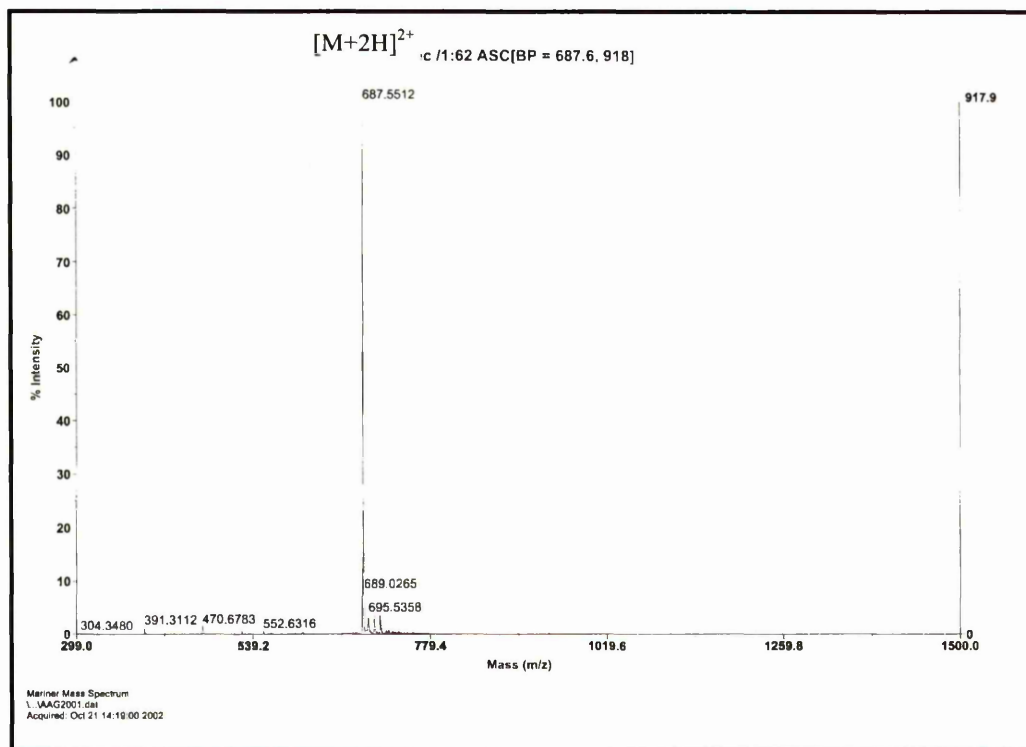
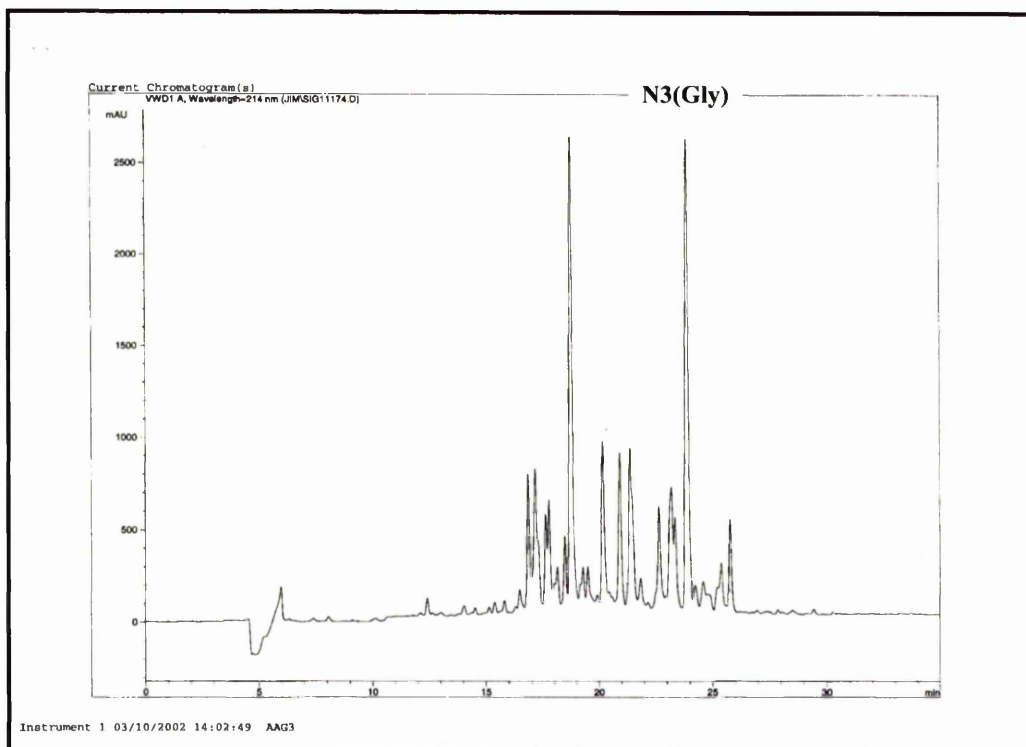
A.2. Identity confirmation of Ac-AACAAAKAAAAKAGY-NH<sub>2</sub> peptide

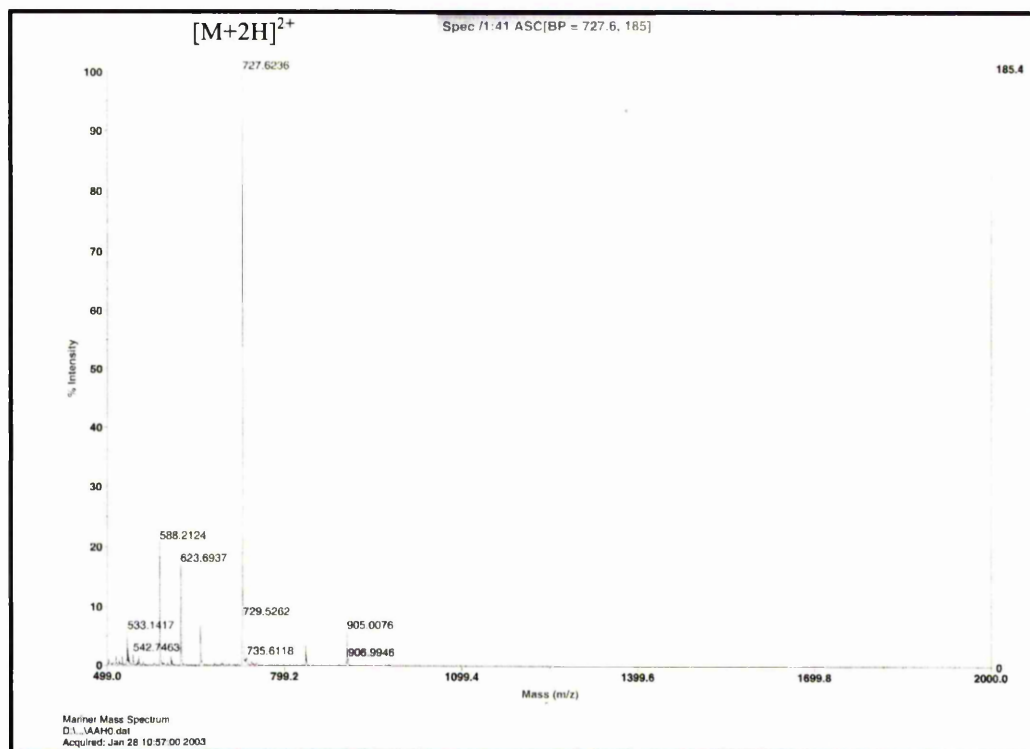
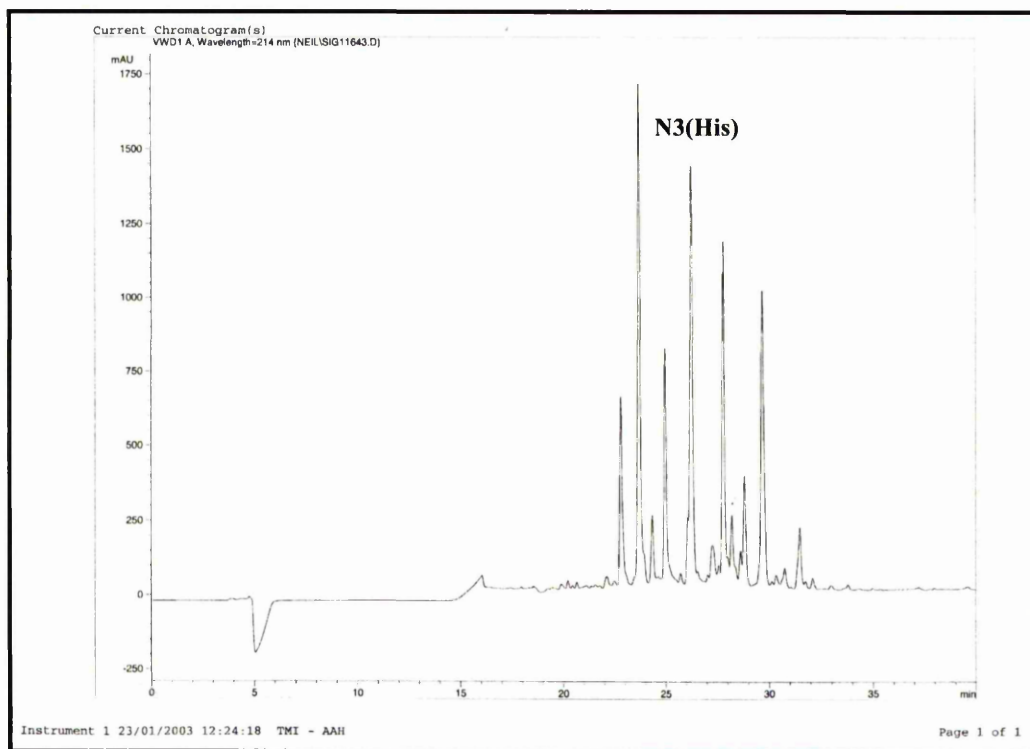
### A.3. Identity confirmation of Ac-AADAAAKAAAKAGY-NH<sub>2</sub> peptide

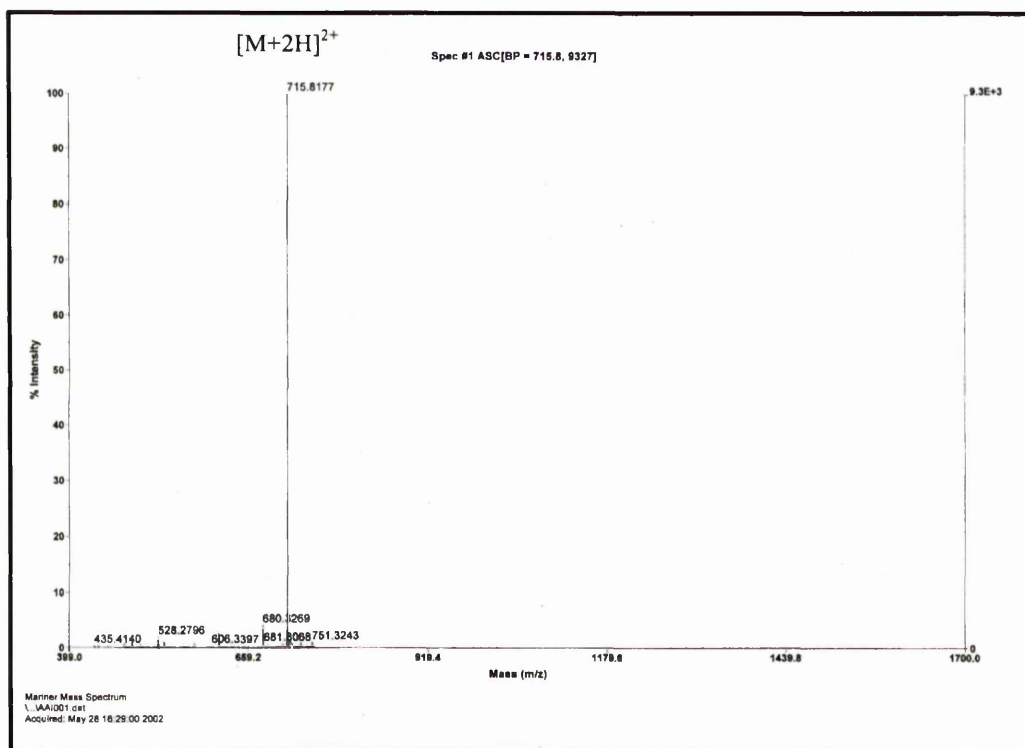
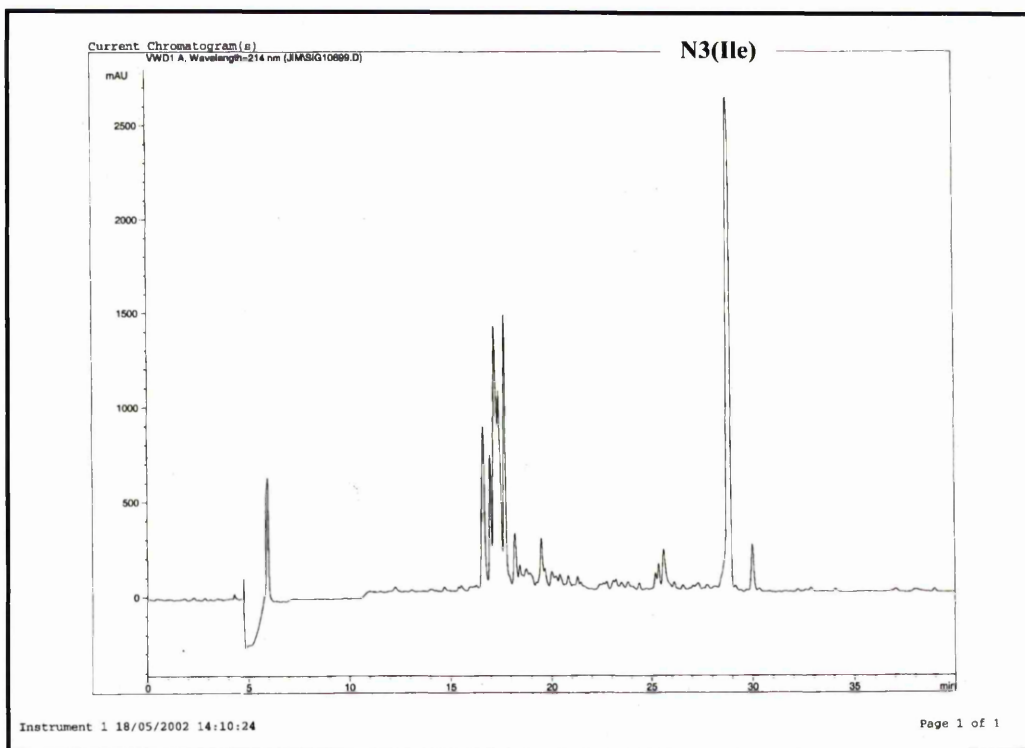


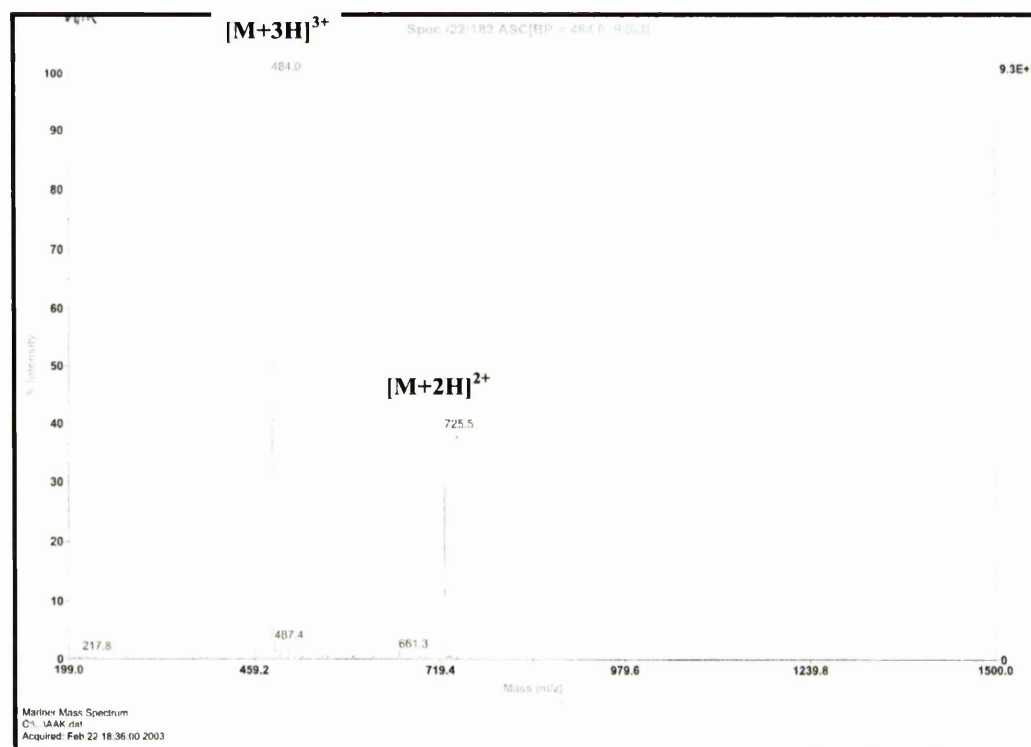
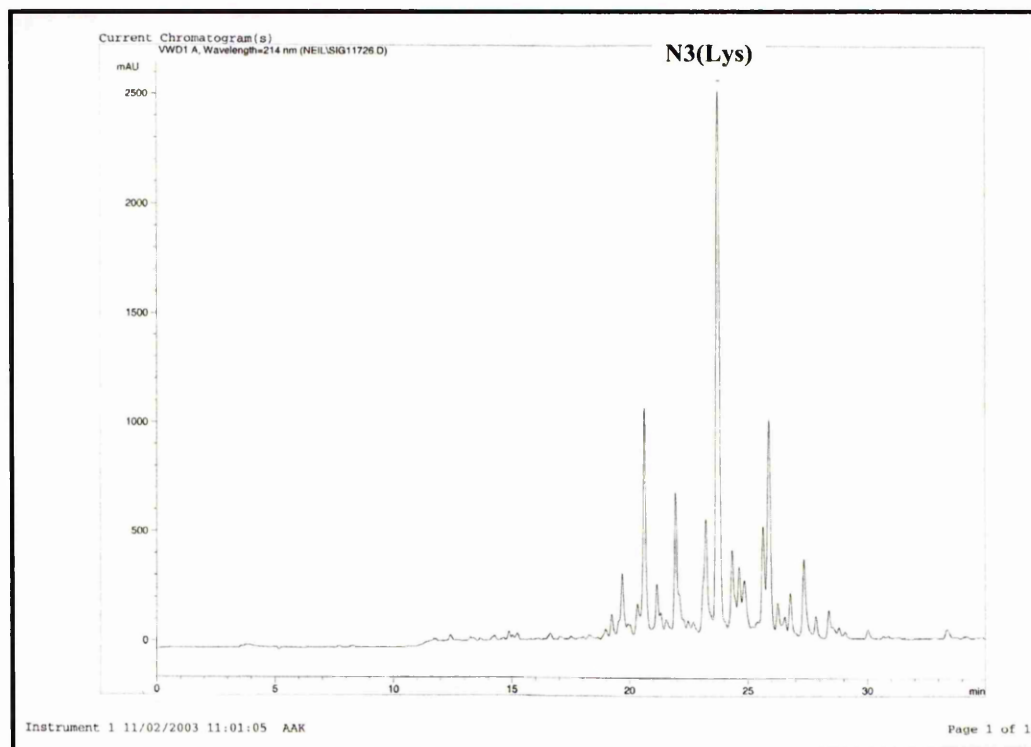
A.4. Identity confirmation of Ac-AAEAAAAKAAAKAGY-NH<sub>2</sub> peptide

A.5. Identity confirmation of Ac-AAFAAAAKAAAKAGY-NH<sub>2</sub> peptide

A.6. Identity confirmation of Ac-AAGAAA(AKAAAKAGY)-NH<sub>2</sub> peptide

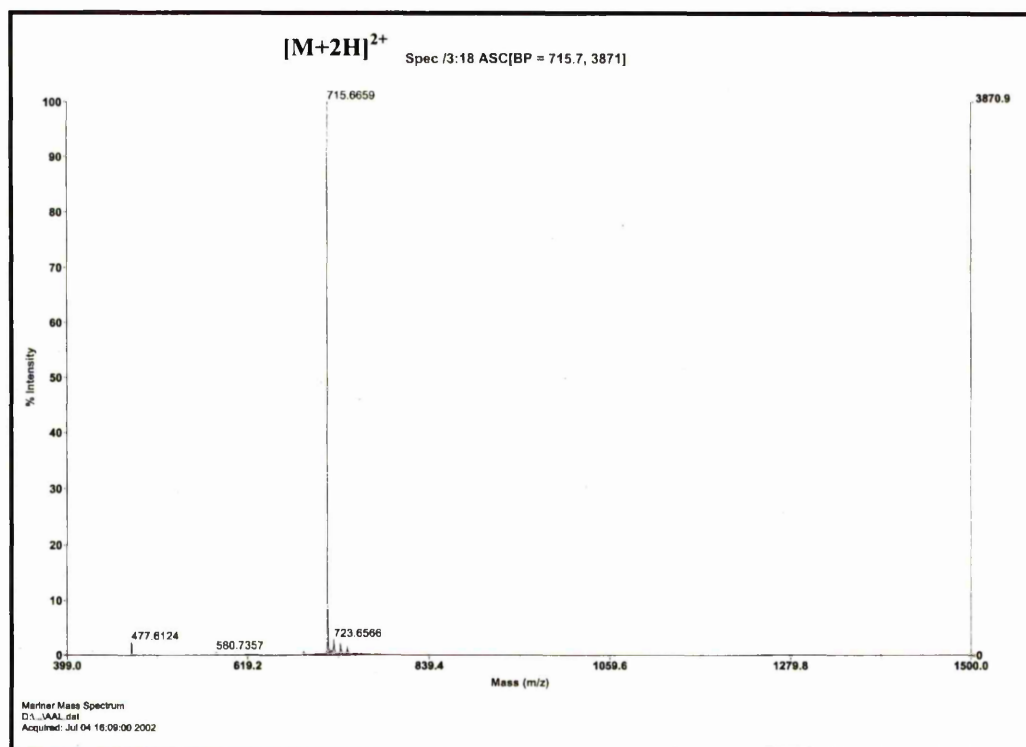
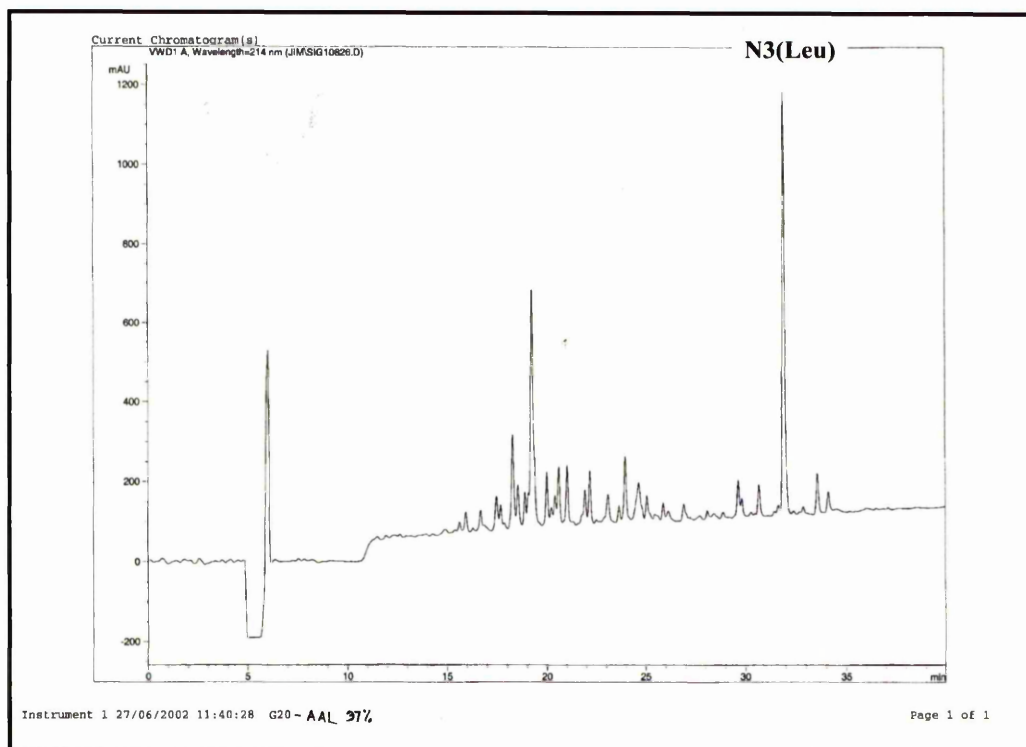
A.7. Identity confirmation of Ac-AAHAAAAKAAAKAG-NH<sub>2</sub> peptide

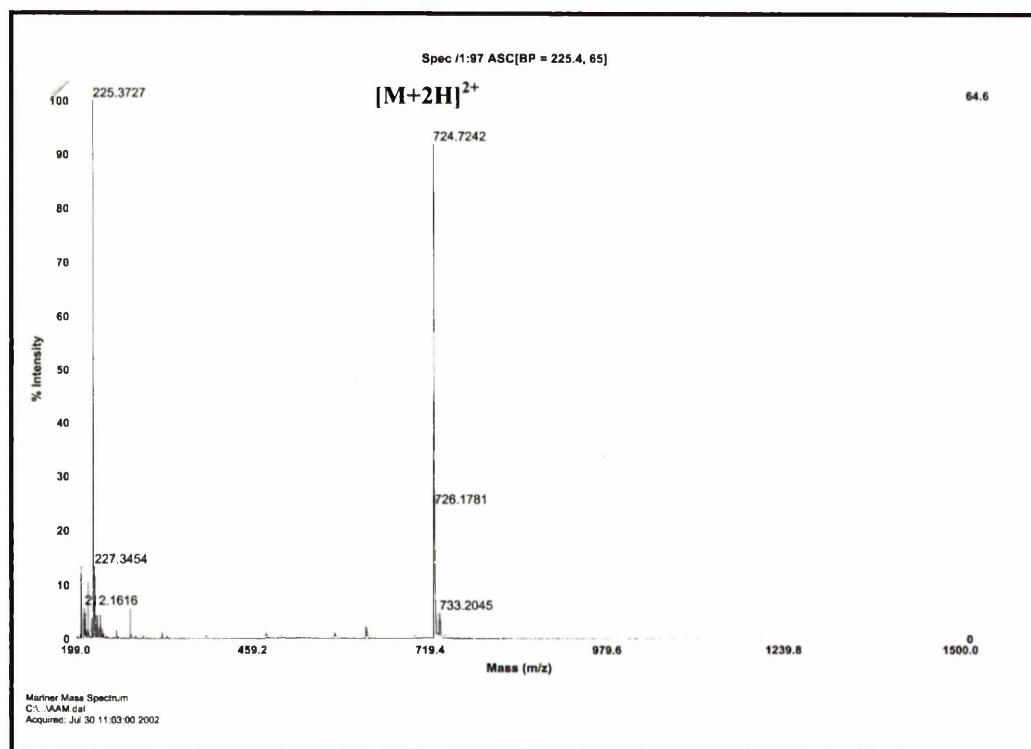
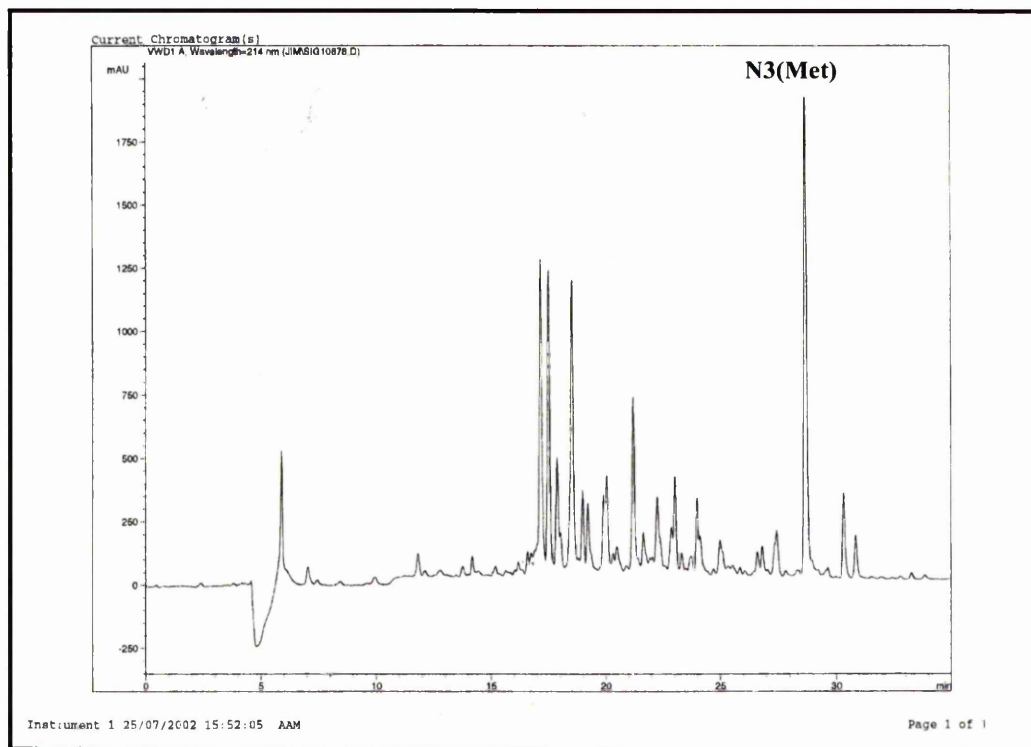
A.8. Identity confirmation of Ac-AAIAAAAKAAAKAG-NH<sub>2</sub> peptide

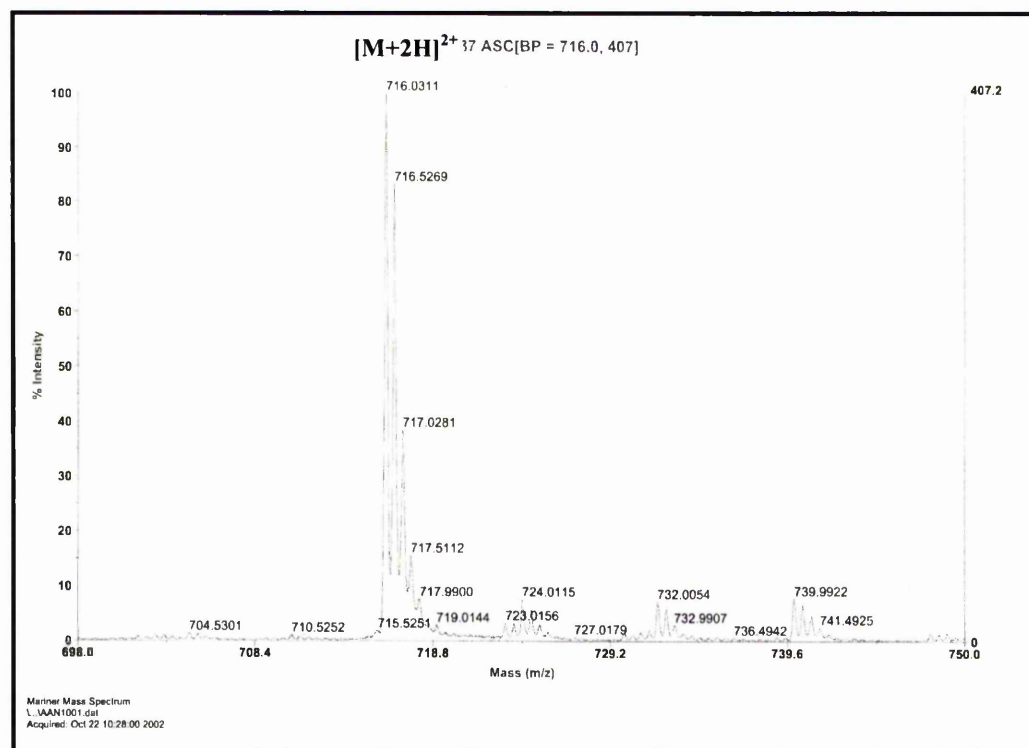
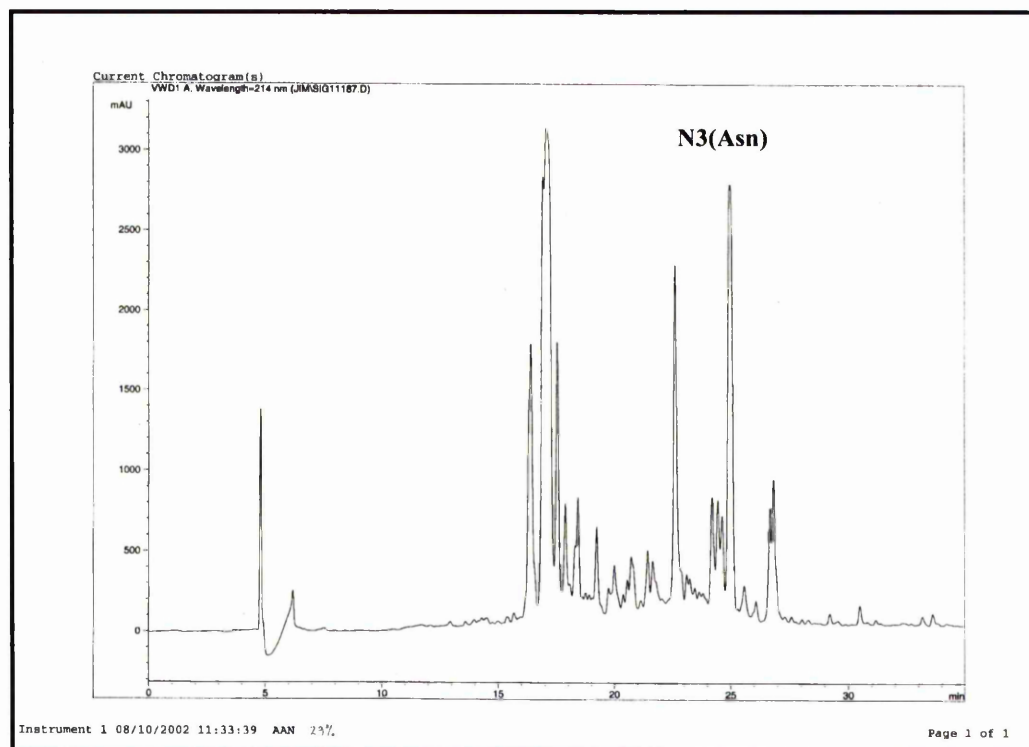
A.9. Identity confirmation of Ac-AAKAAAKAAAKAG-NH<sub>2</sub> peptide

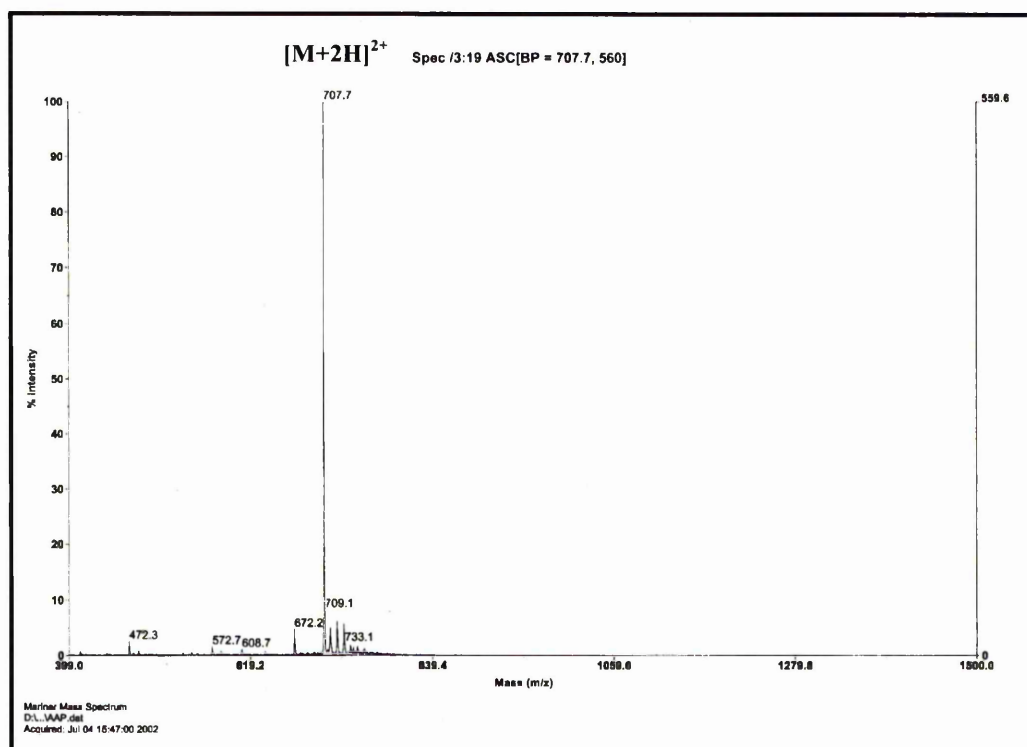
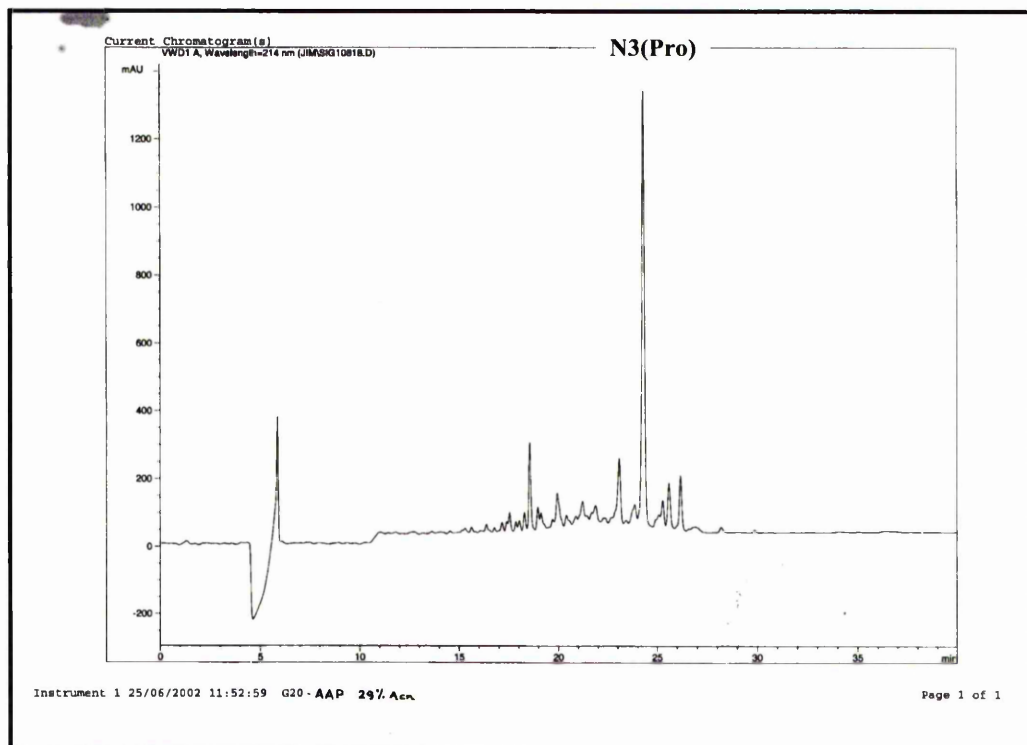


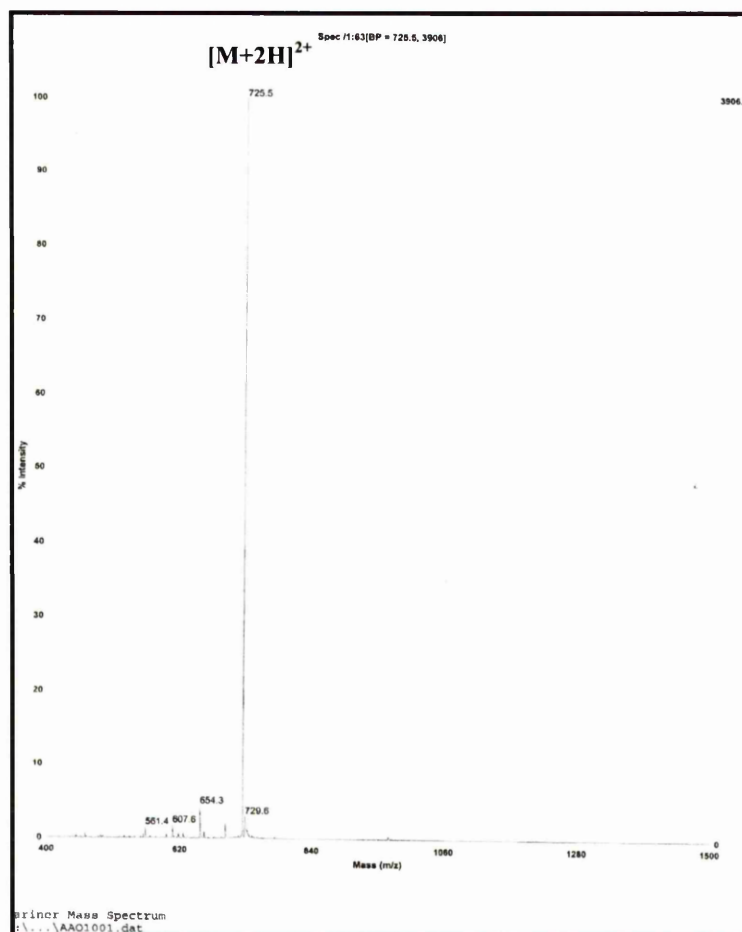
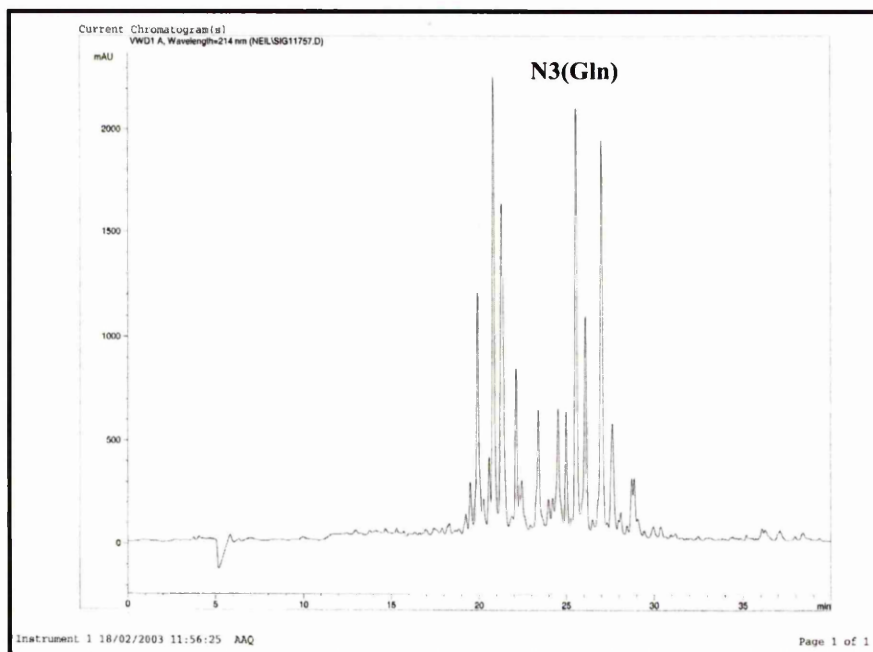
A.10. Identity confirmation of Ac-AALAAAAKAAAKAG-NH<sub>2</sub> peptide

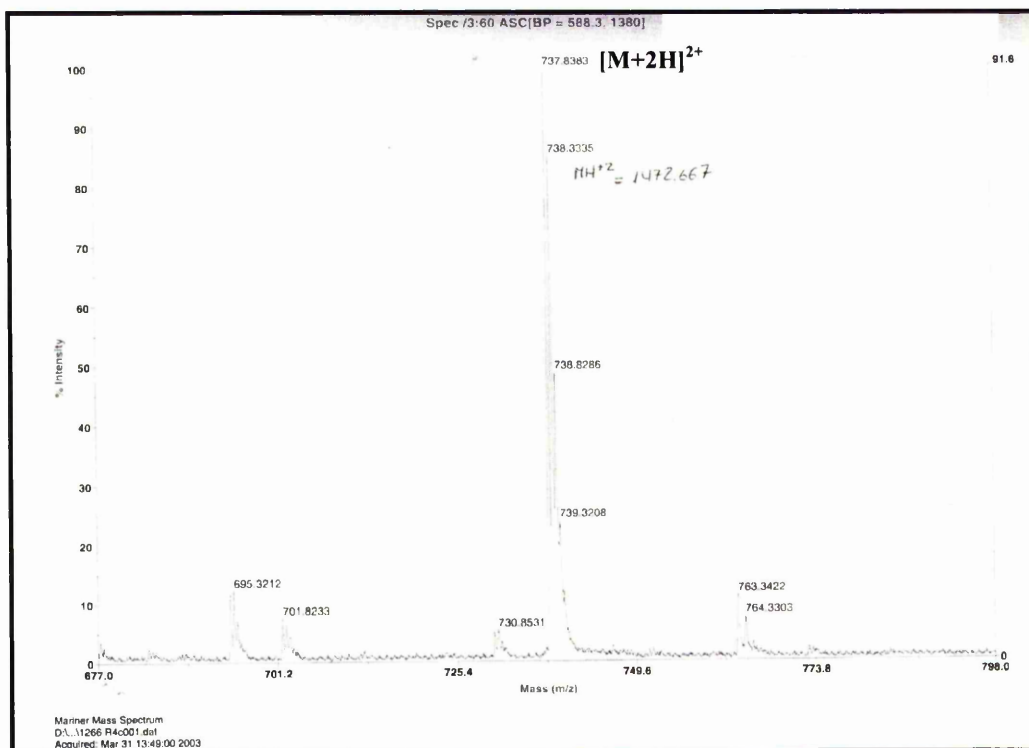
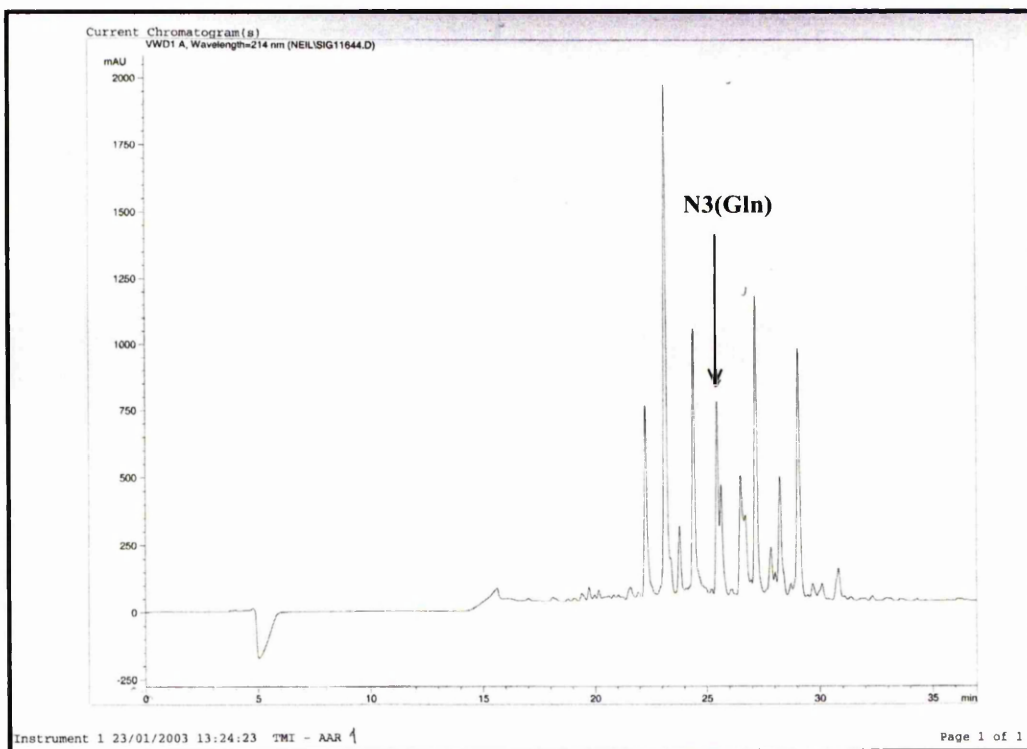


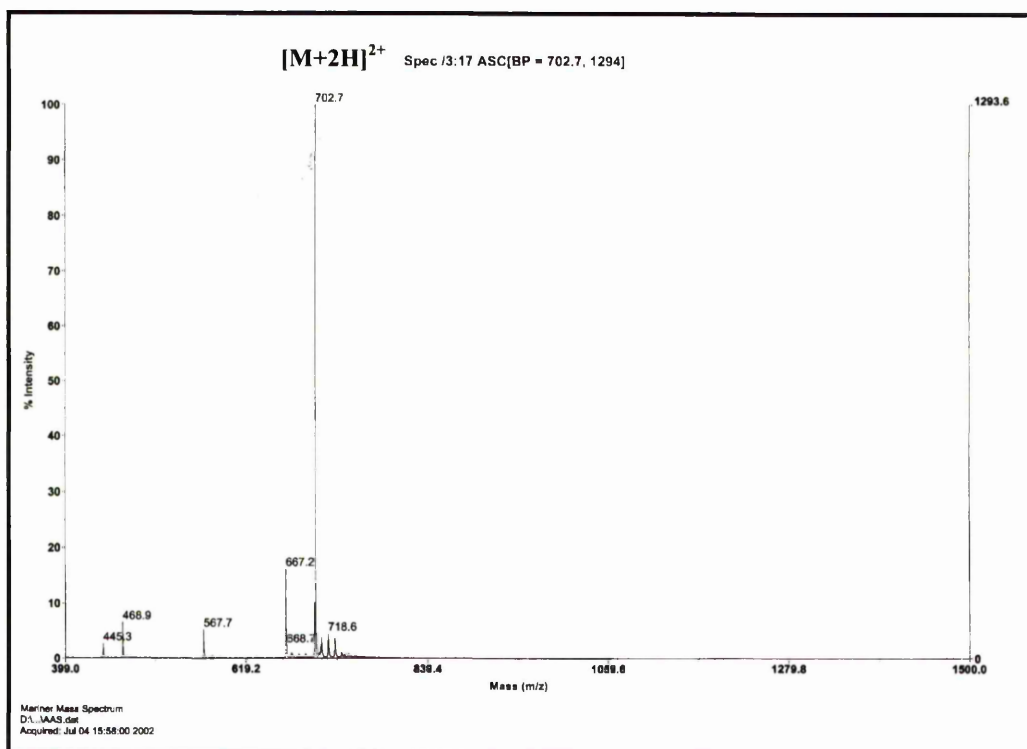
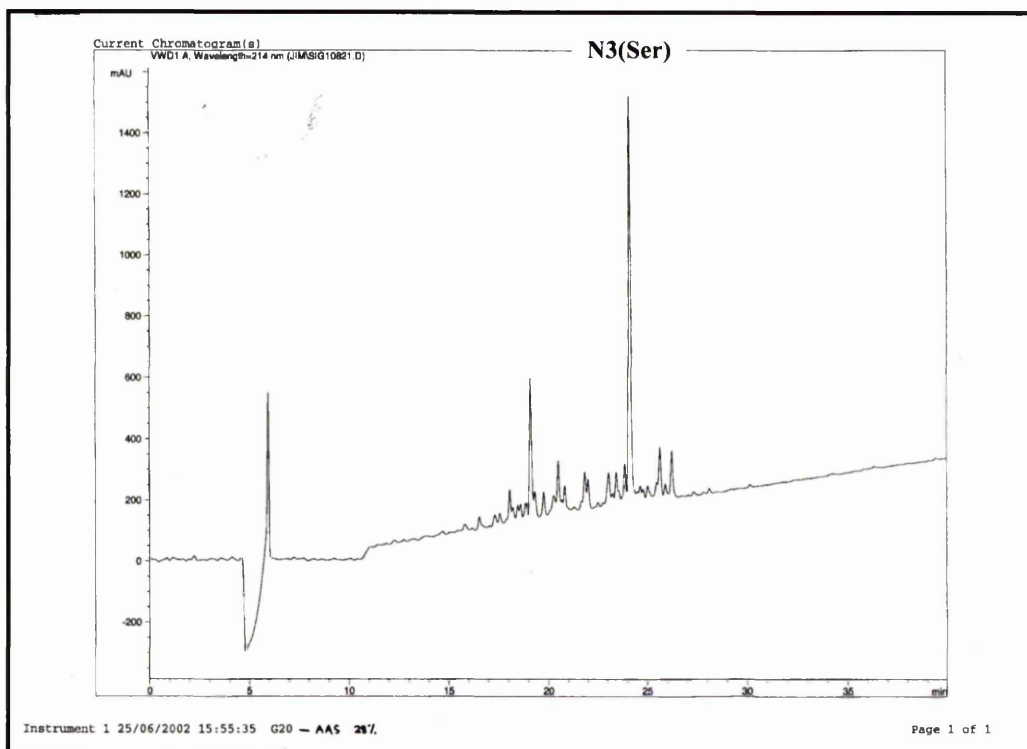
A.11. Identity confirmation of Ac-AAMAAAAKAAAKAGY-NH<sub>2</sub> peptide

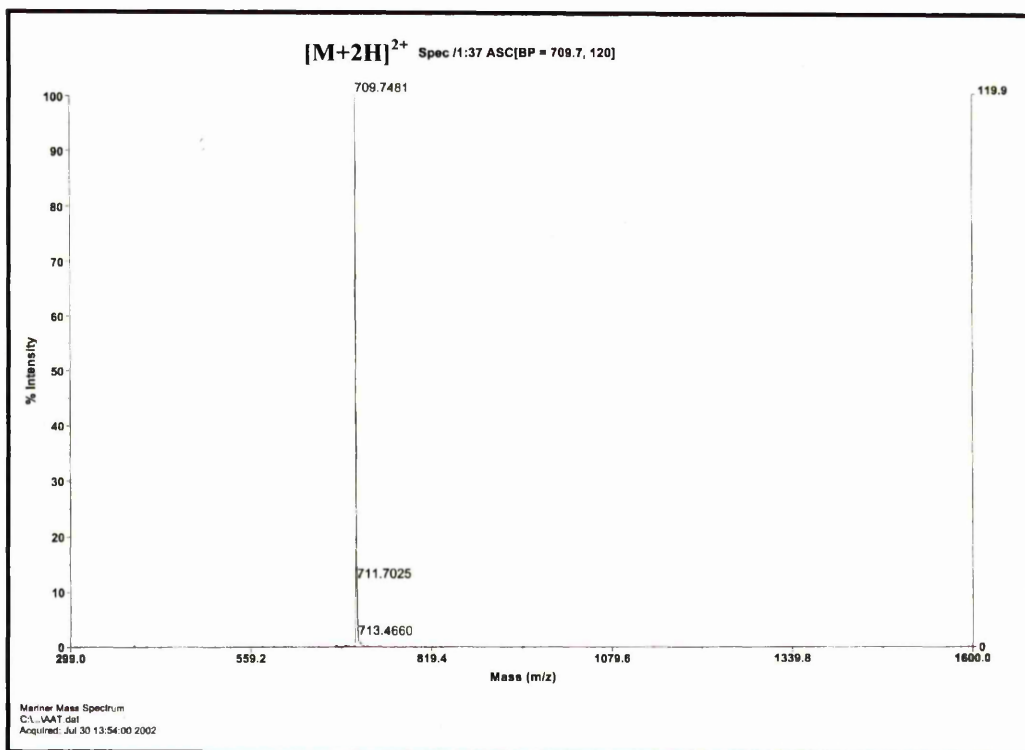
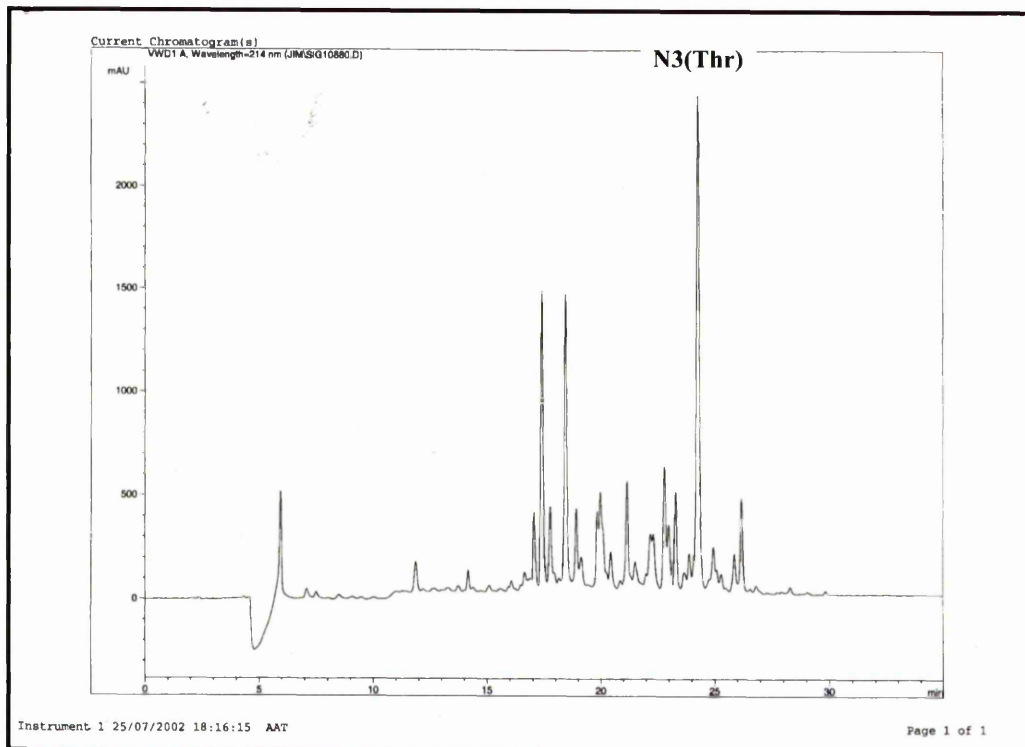
A.12. Identity confirmation of Ac-AANAAAAKAAAKAGY-NH<sub>2</sub> peptide

A.13. Identity confirmation of Ac-AAPAAAAKAAAKAG-NH<sub>2</sub> peptide

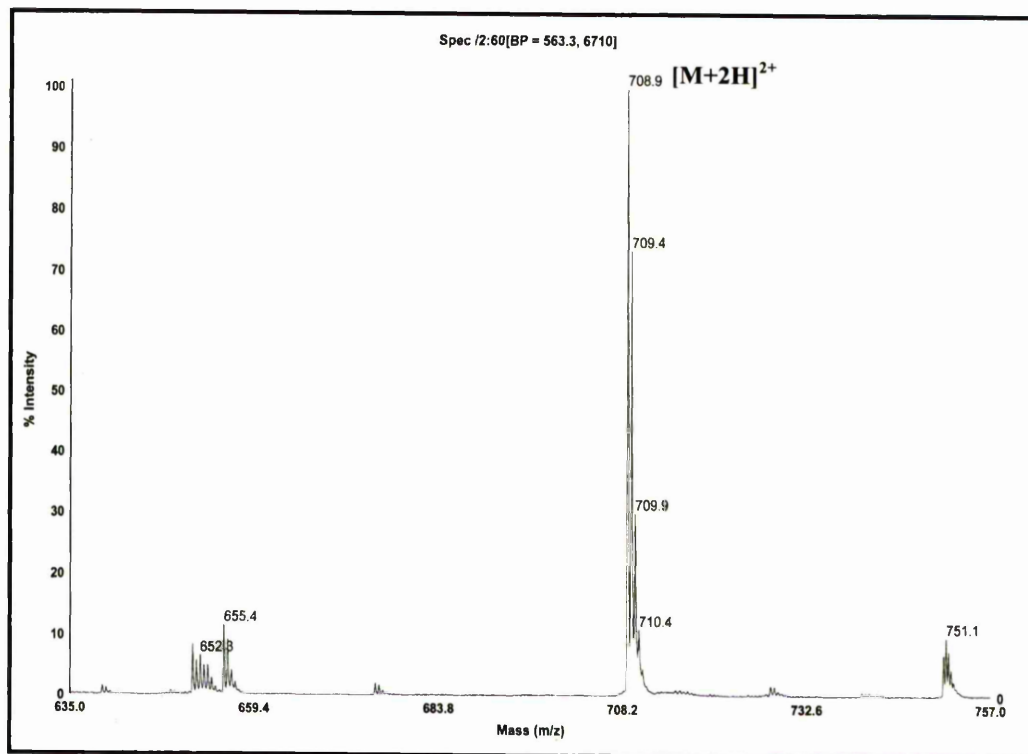
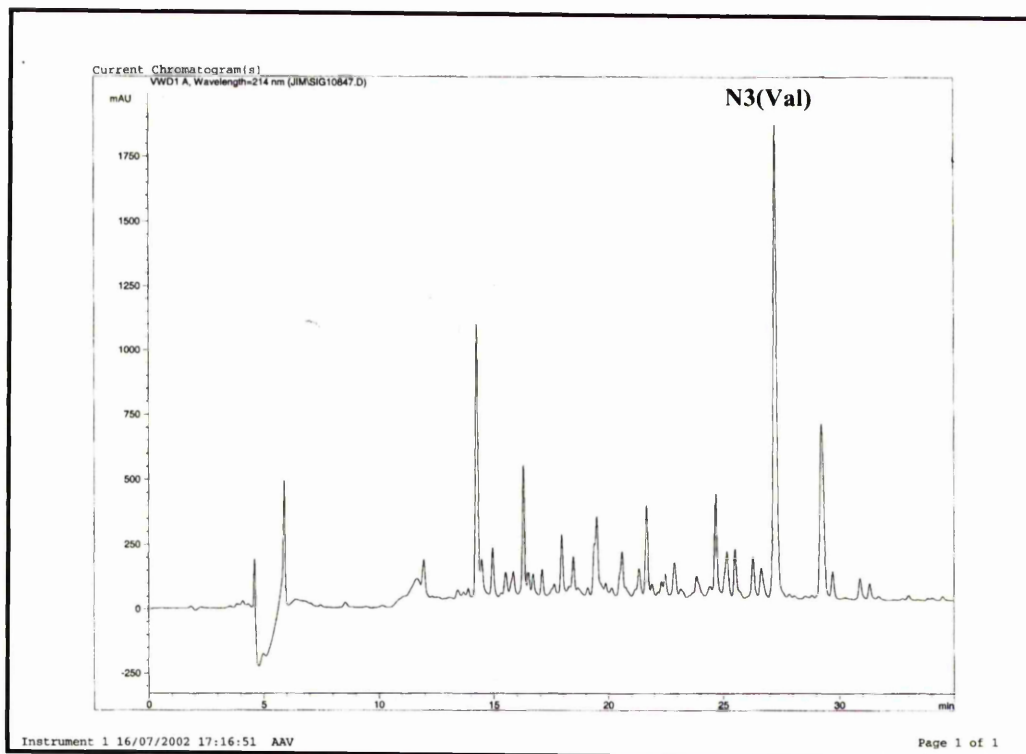
A.14. Identity confirmation of Ac-AAQAAA KAAA KAGY-NH<sub>2</sub> peptide

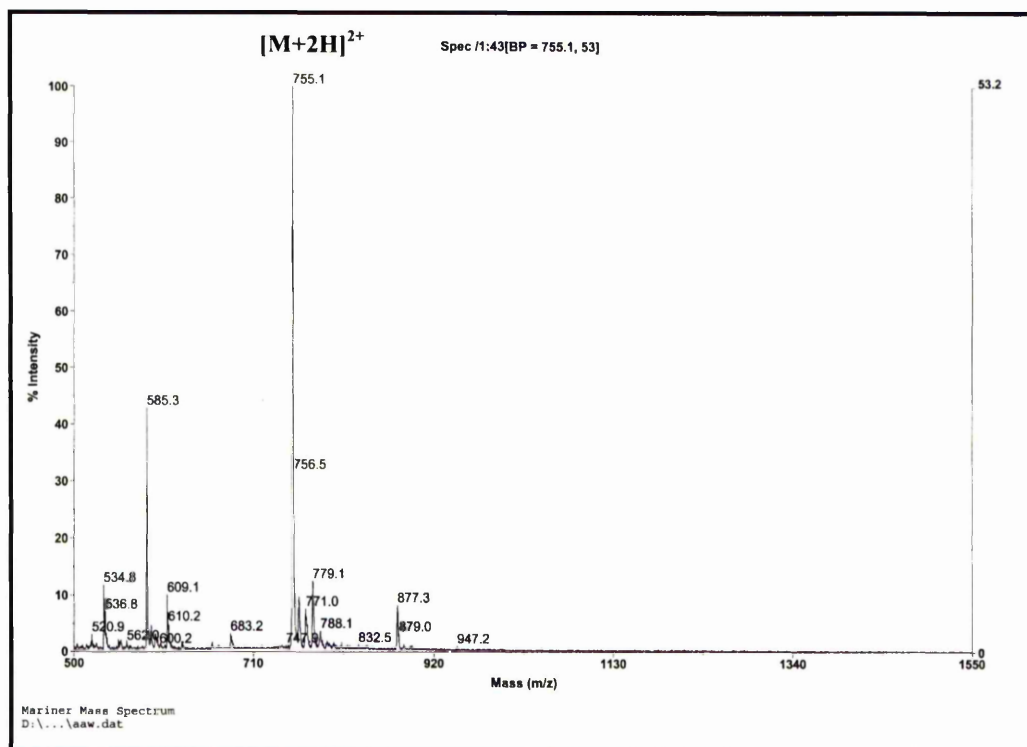
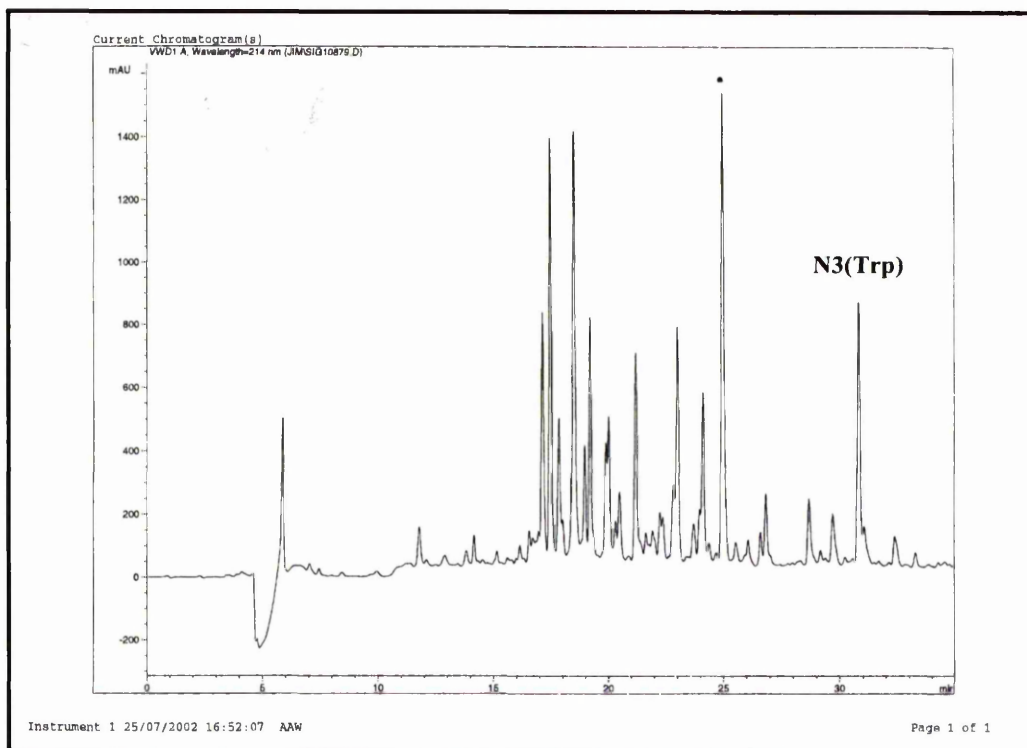
A.15. Identity confirmation of Ac-AARAAAAKAAAKAGY-NH<sub>2</sub> peptide

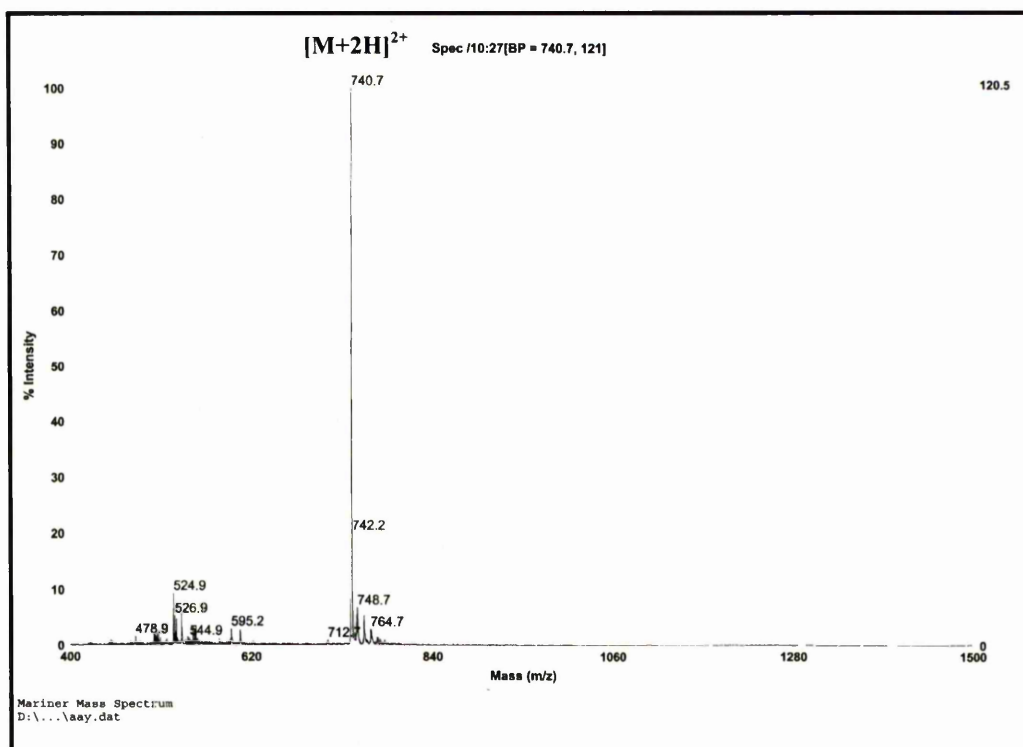
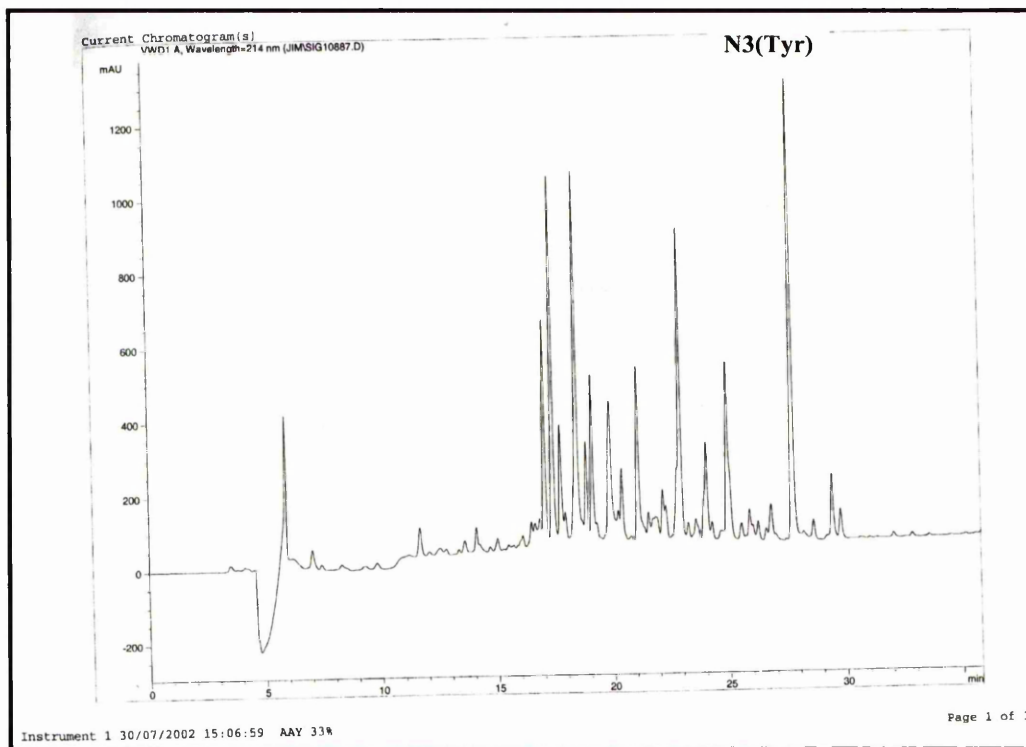
A.16. Identity confirmation of Ac-AASAAAAKAAAAKAGY-NH<sub>2</sub> peptide

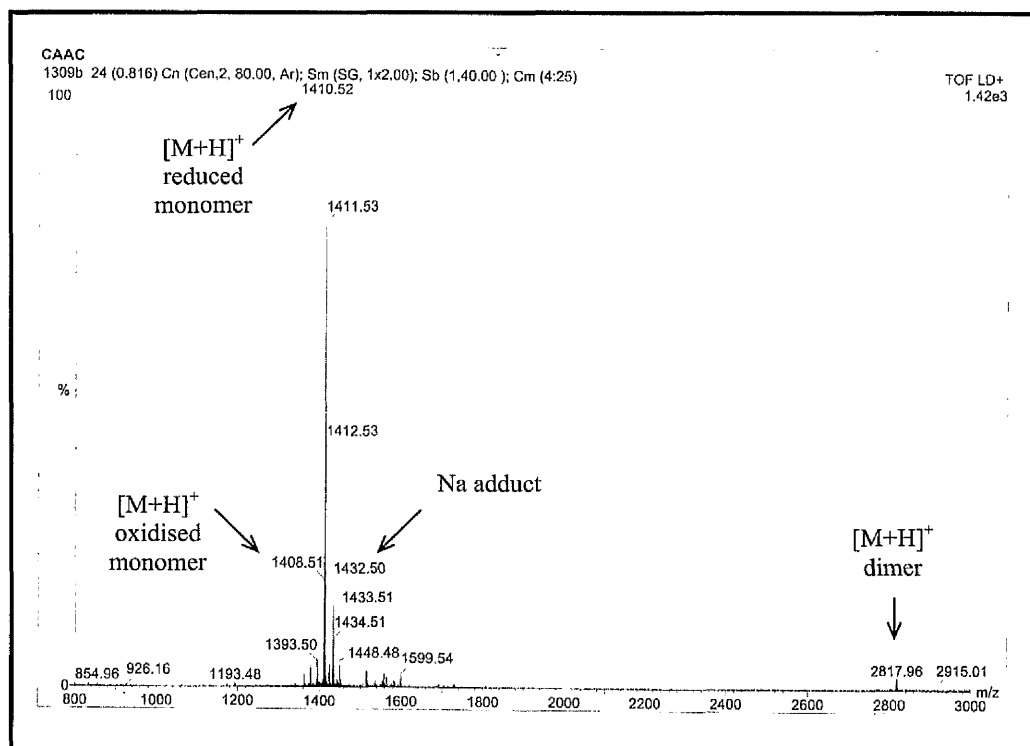
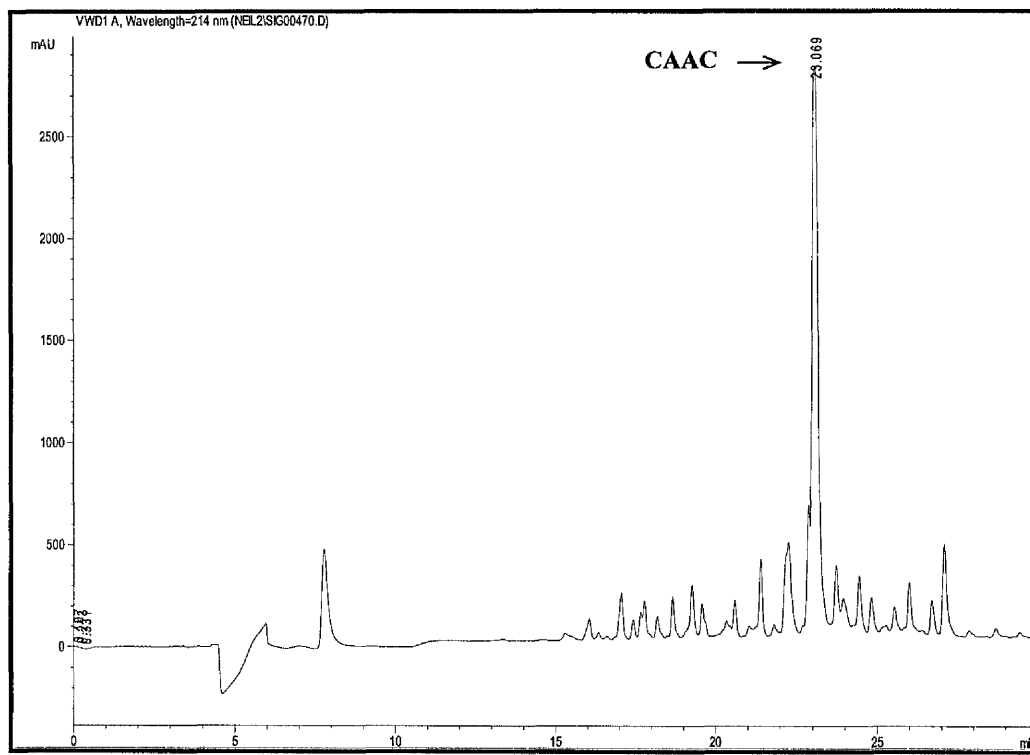
A.17. Identity confirmation of Ac-AATAAAAKAAAKAGY-NH<sub>2</sub> peptide



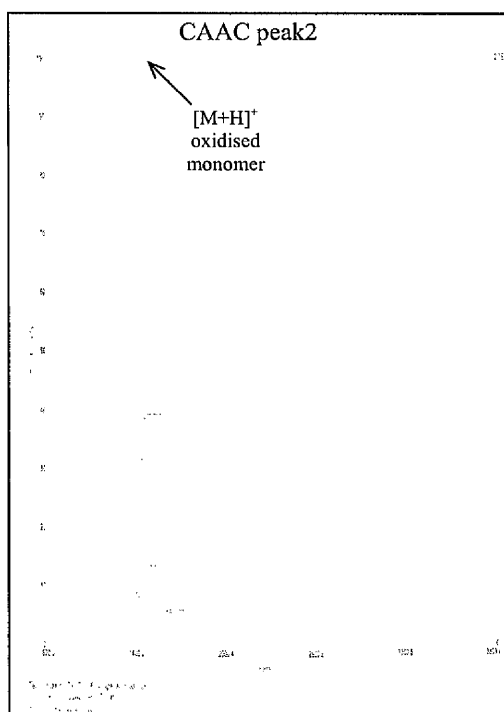
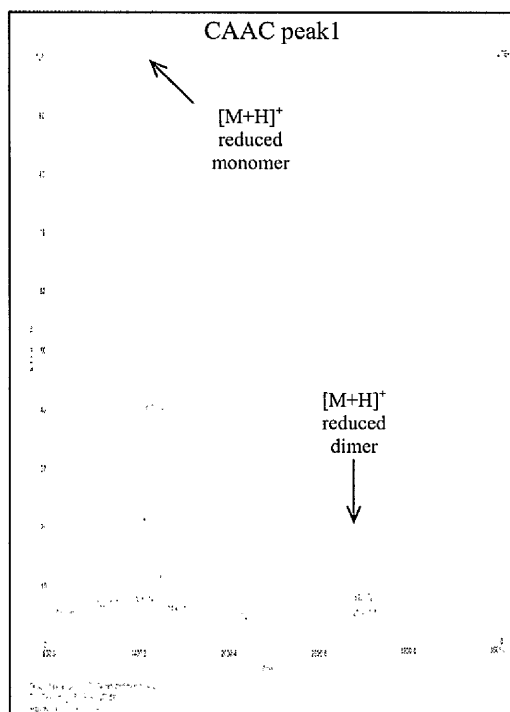
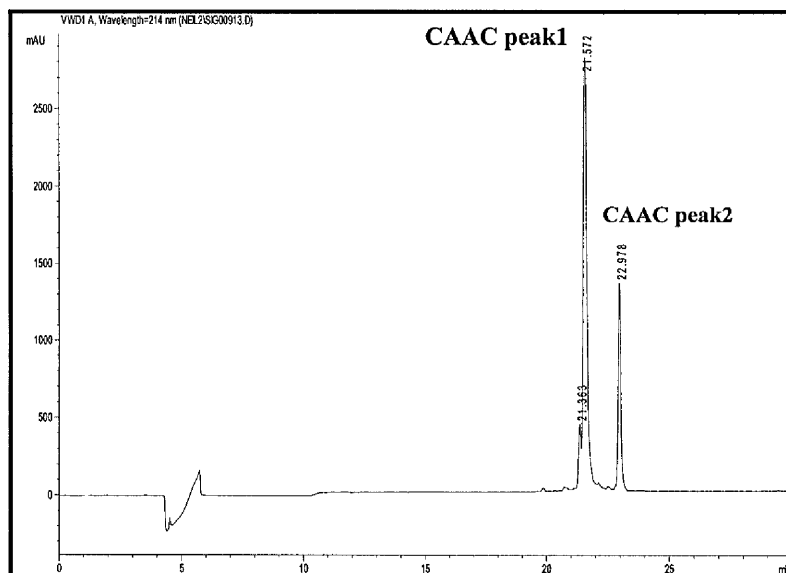
A.18. Identity confirmation of Ac-AAVAAAAKAAAKAGY-NH<sub>2</sub> peptide

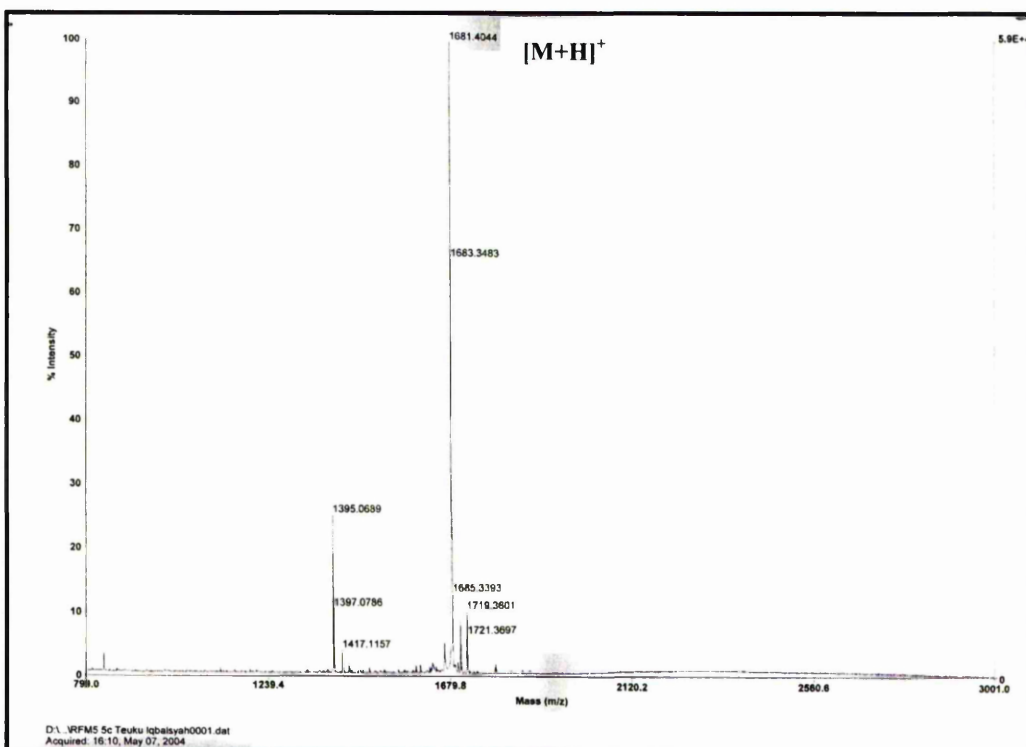
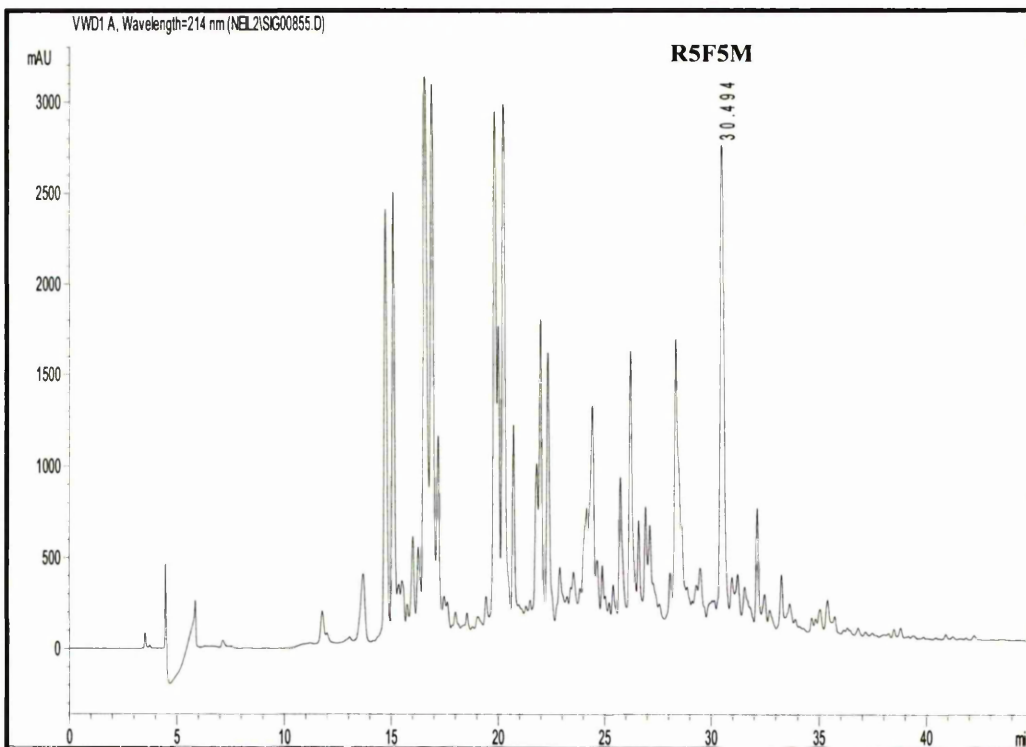
A.19. Identity confirmation of Ac-AAWAAAAKAAAAKAGY-NH<sub>2</sub> peptide

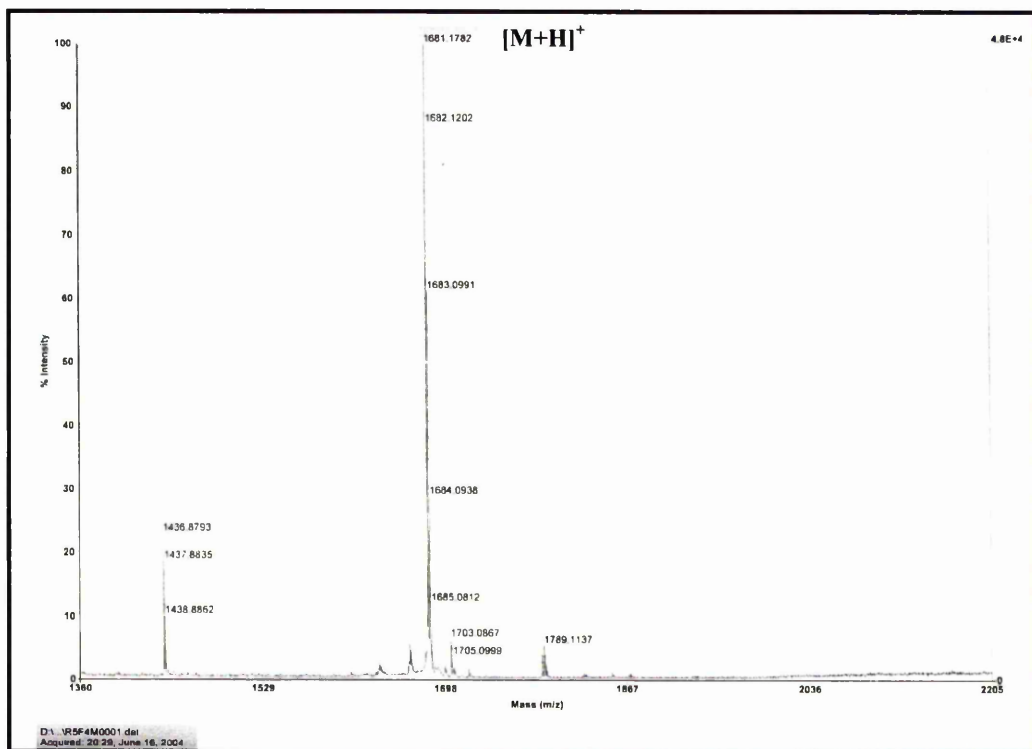
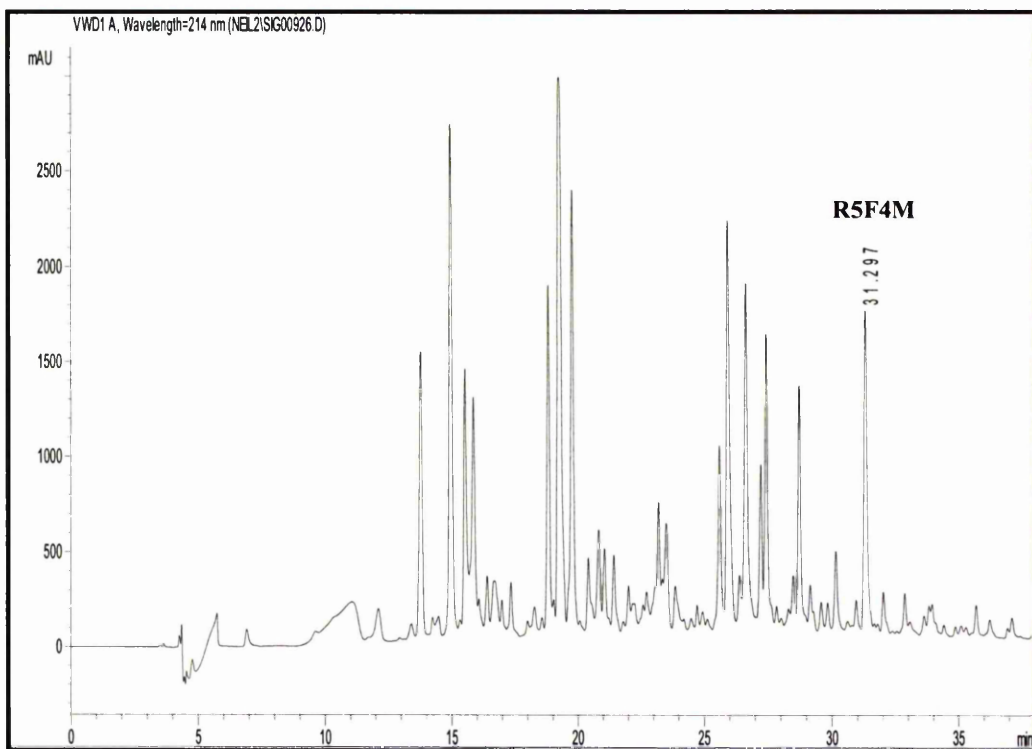
A.20. Identity confirmation of Ac-AAYAAAKAAAKAGY-NH<sub>2</sub> peptide

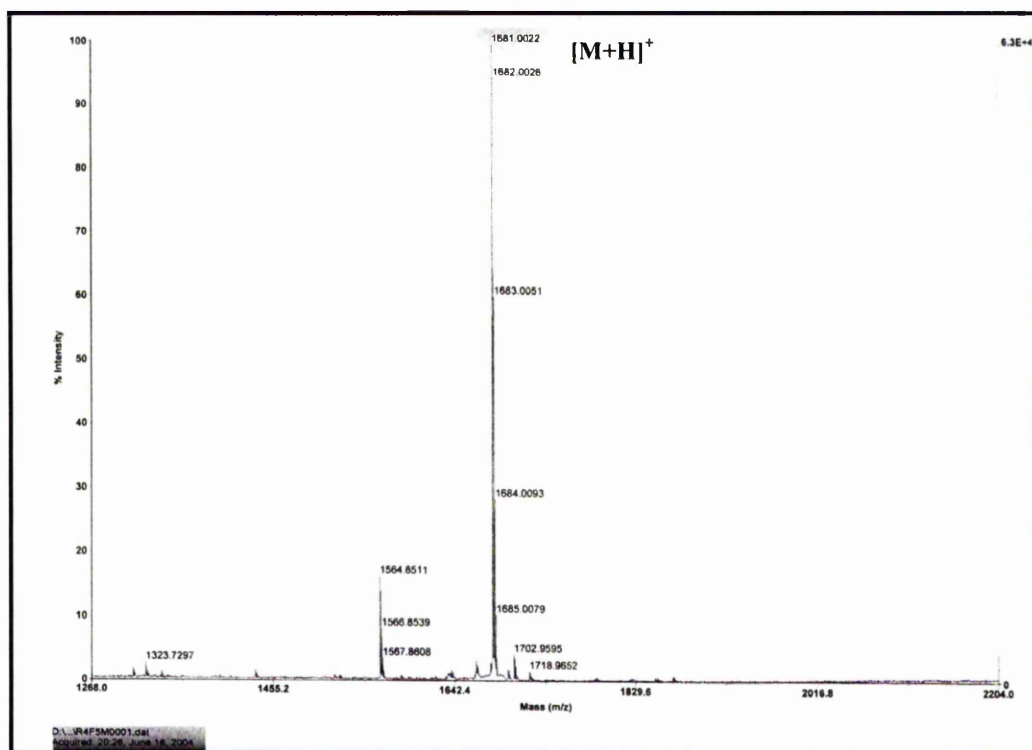
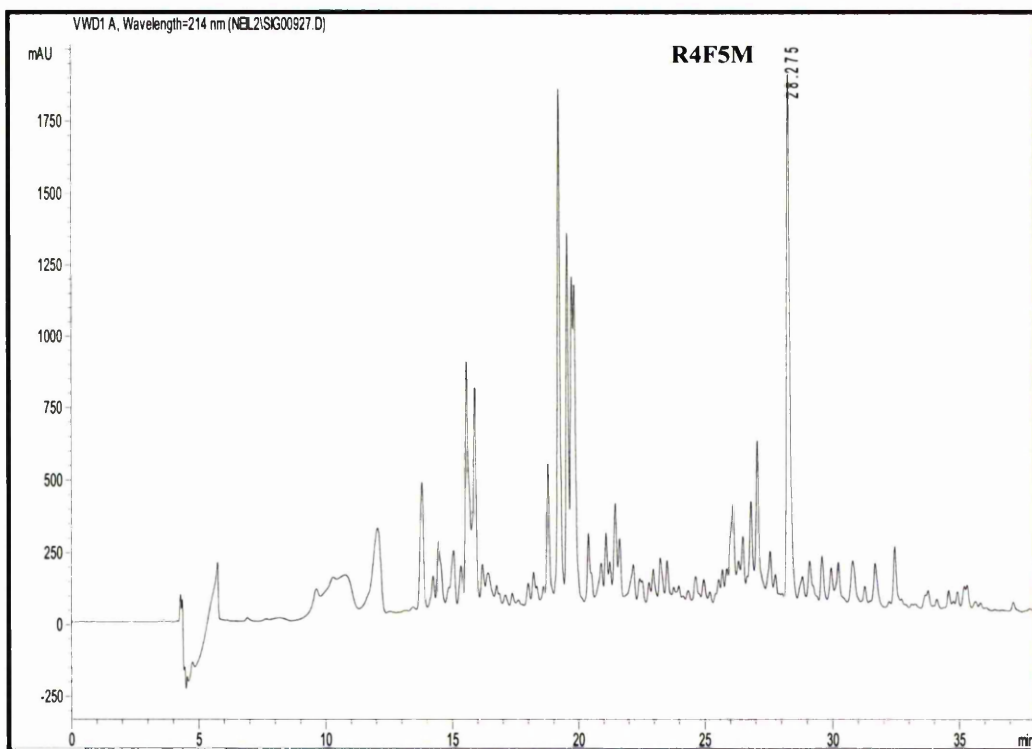
A.21. Identity confirmation of CAACAAAAKAAAAGY-NH<sub>2</sub> peptide from crude sample

# A.22. Identity confirmation of CAACAAAAKAAAAKAGY-NH<sub>2</sub> peptide from pure sample

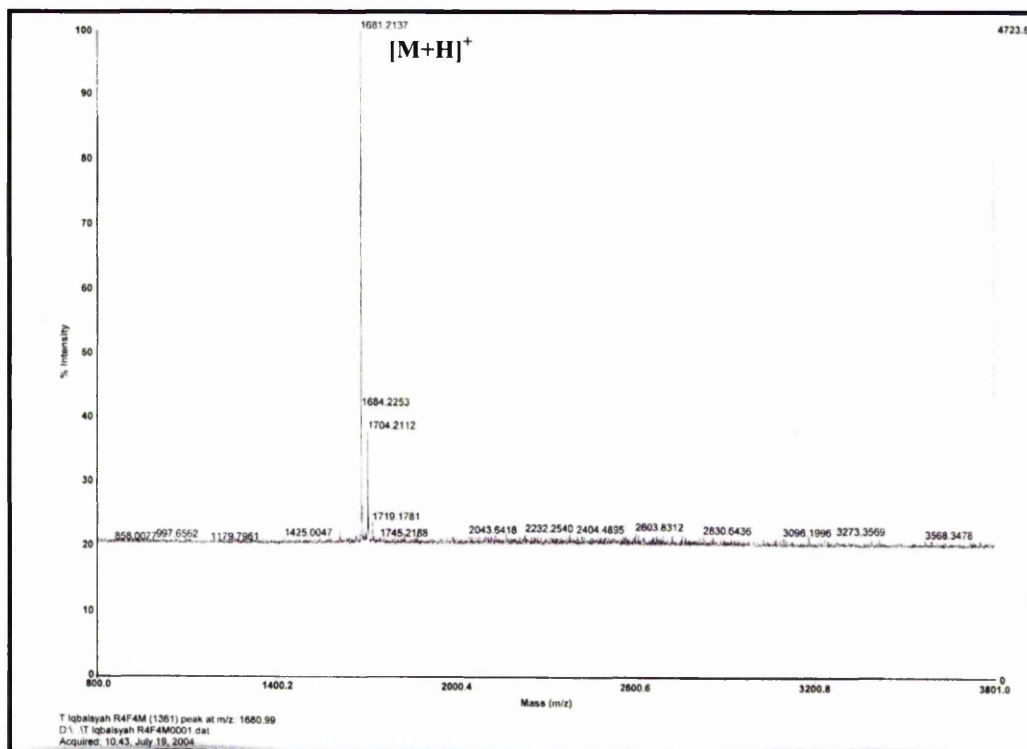
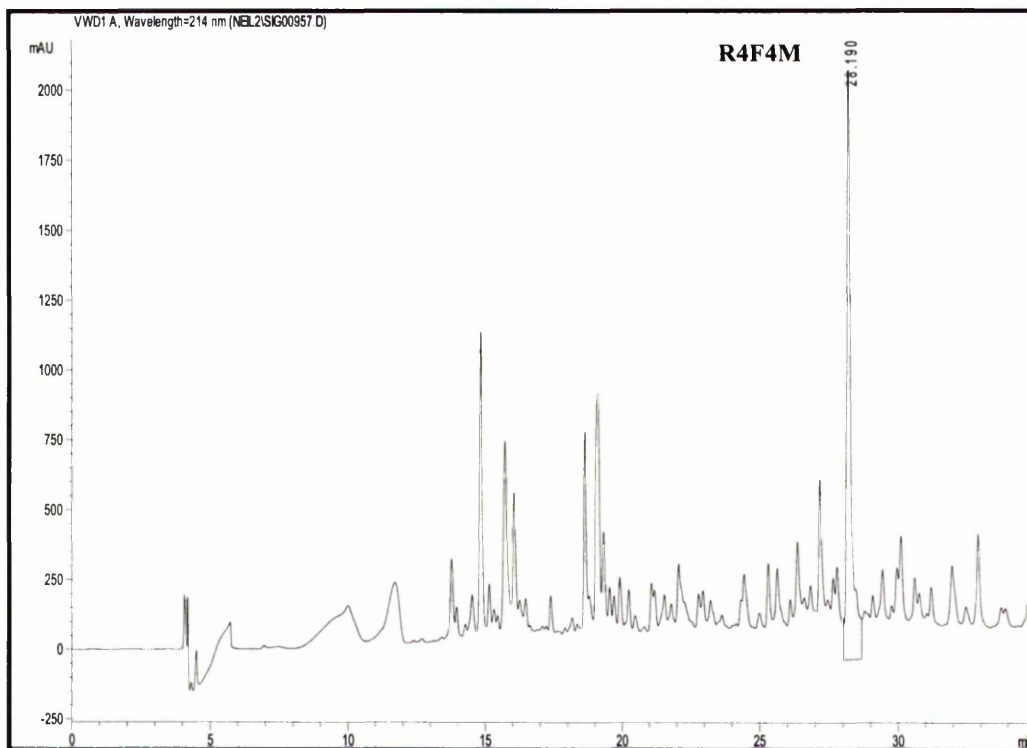


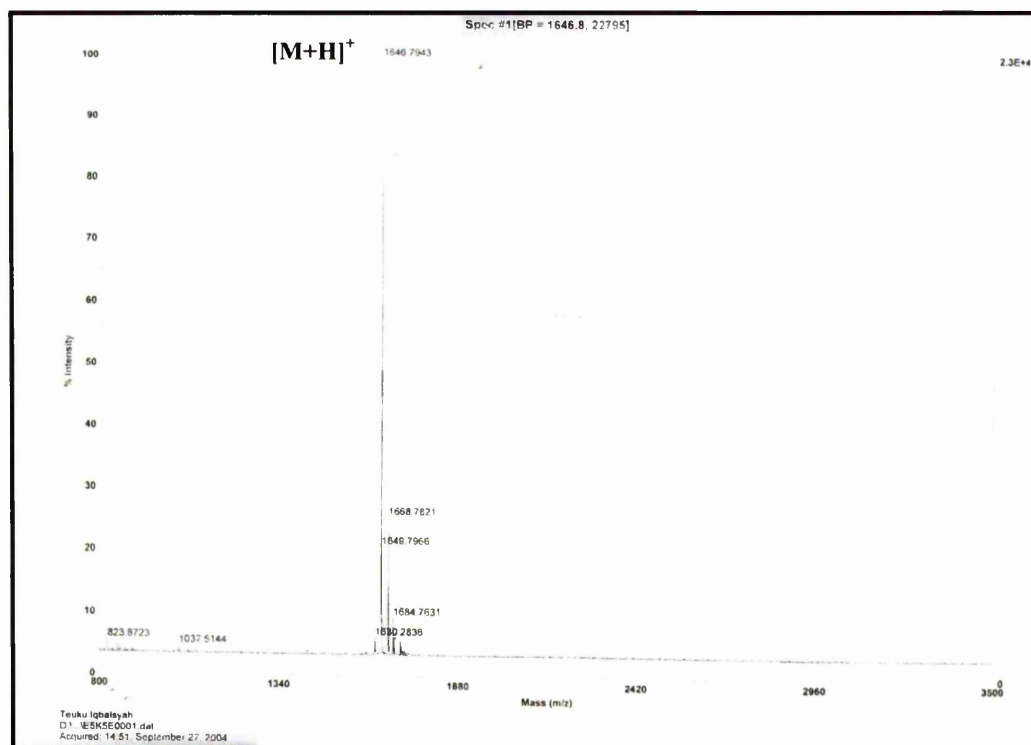
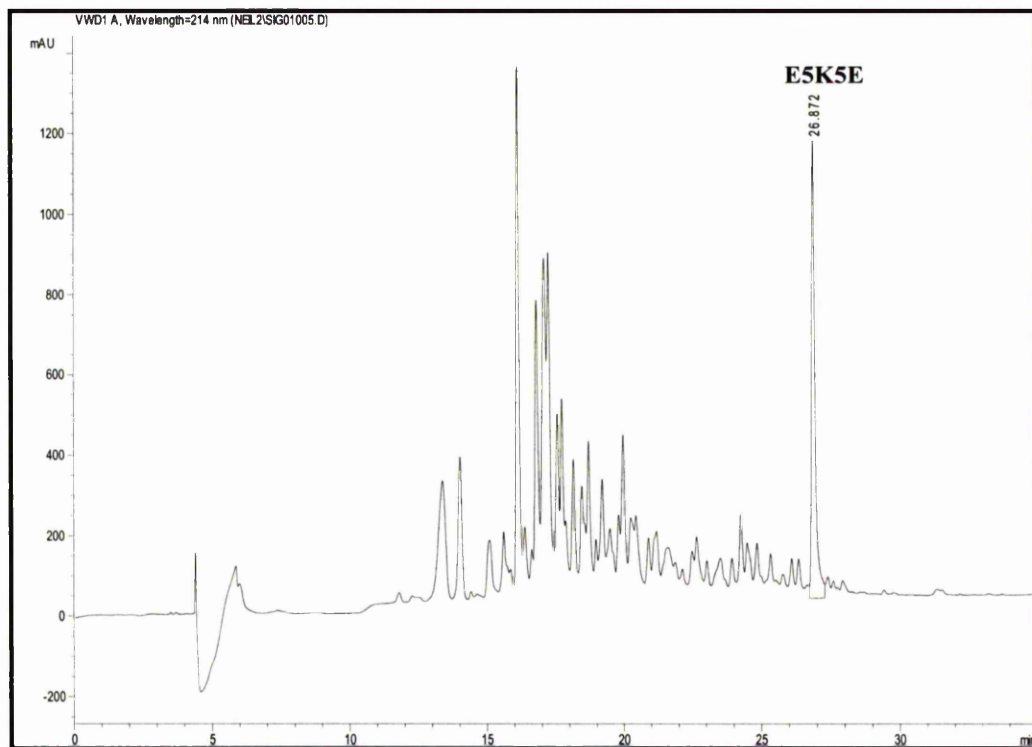
A.23. Identity confirmation of Ac-AAKRAAAAFAAAAMKGY-NH<sub>2</sub> peptide

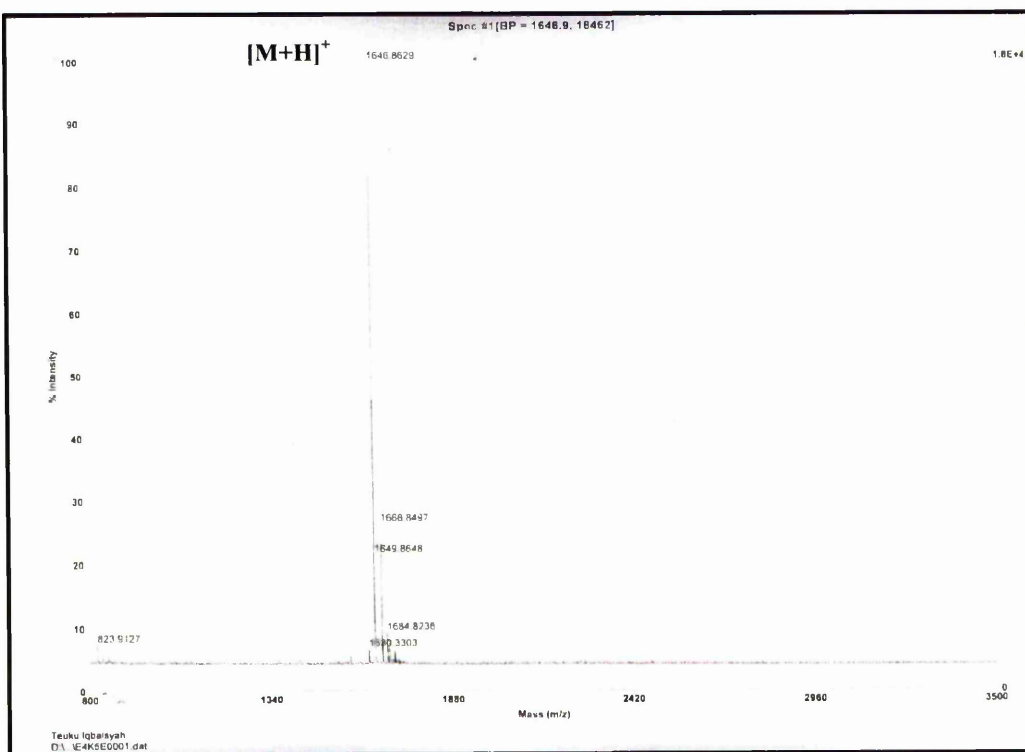
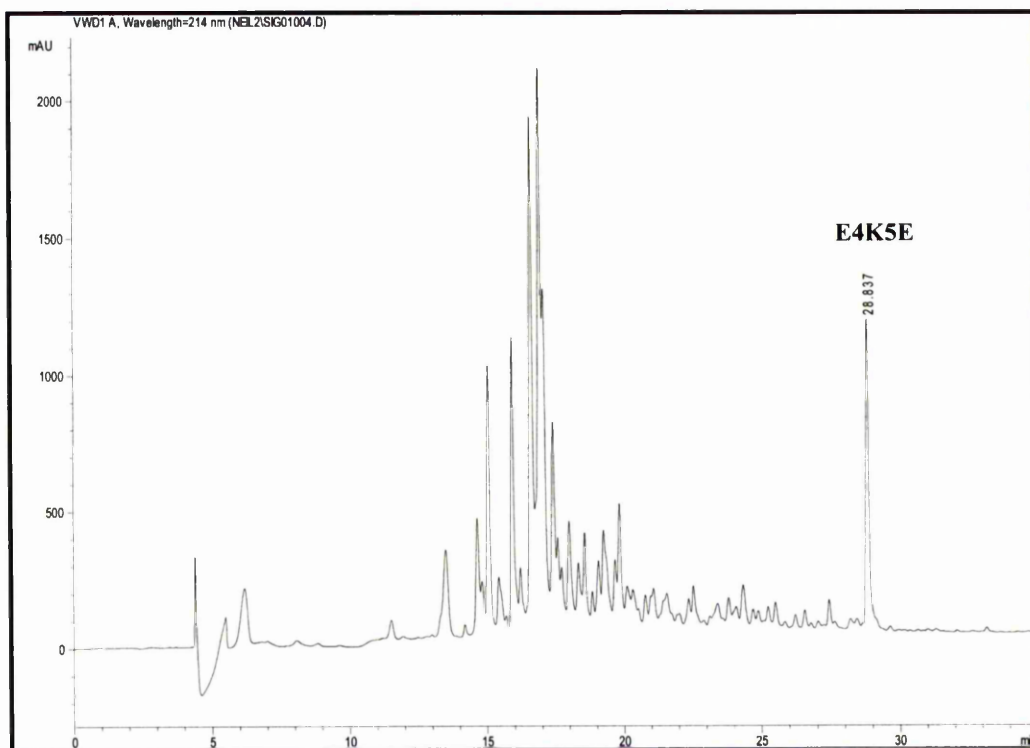
A.24. Identity confirmation of Ac-AAKRAAAAFAAAMAKGY-NH<sub>2</sub> peptide

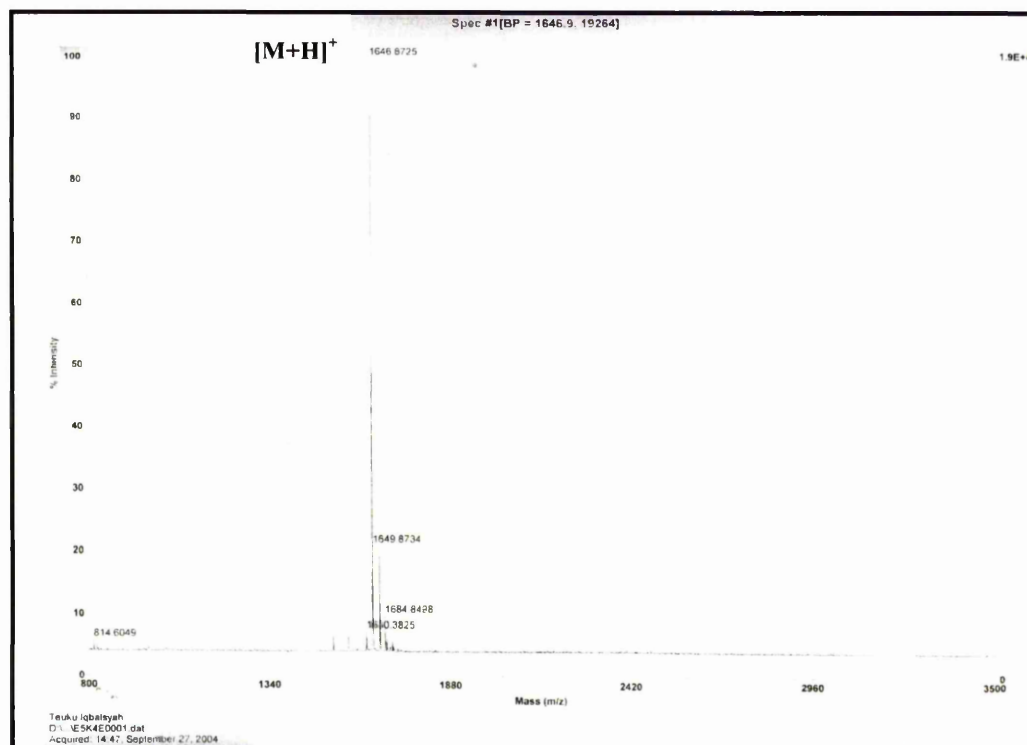
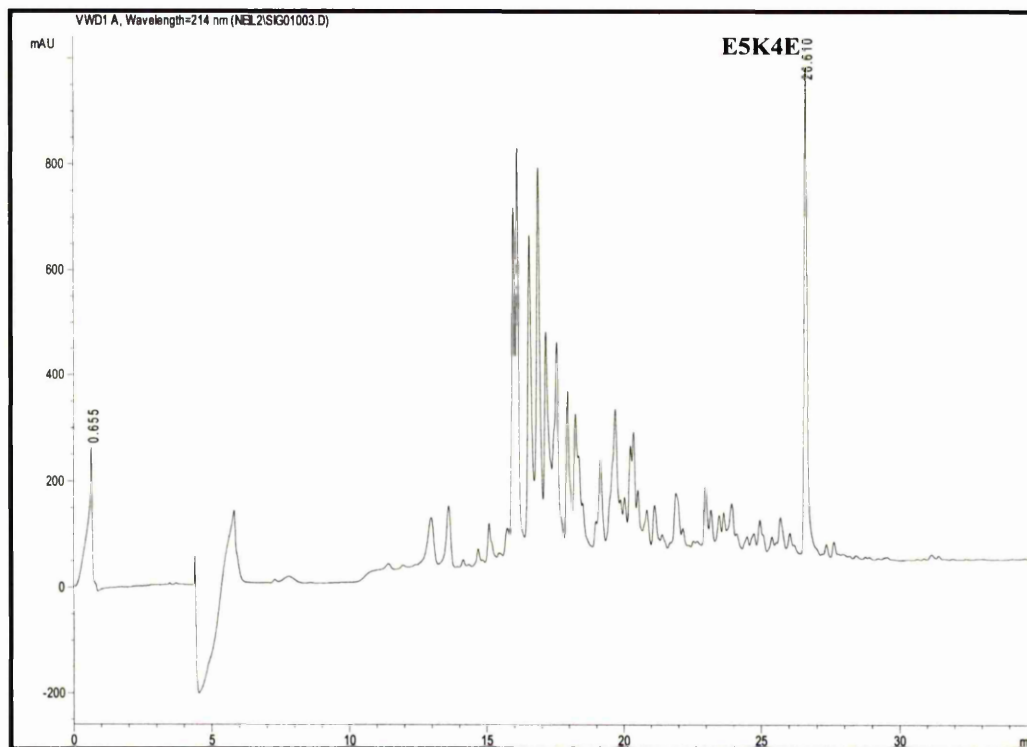
A.25. Identity confirmation of Ac-AAKARAAAFAAAAMKGY-NH<sub>2</sub> peptide



A.26. Identity confirmation of Ac-AAKARAAFAAAMAKGY-NH<sub>2</sub> peptide

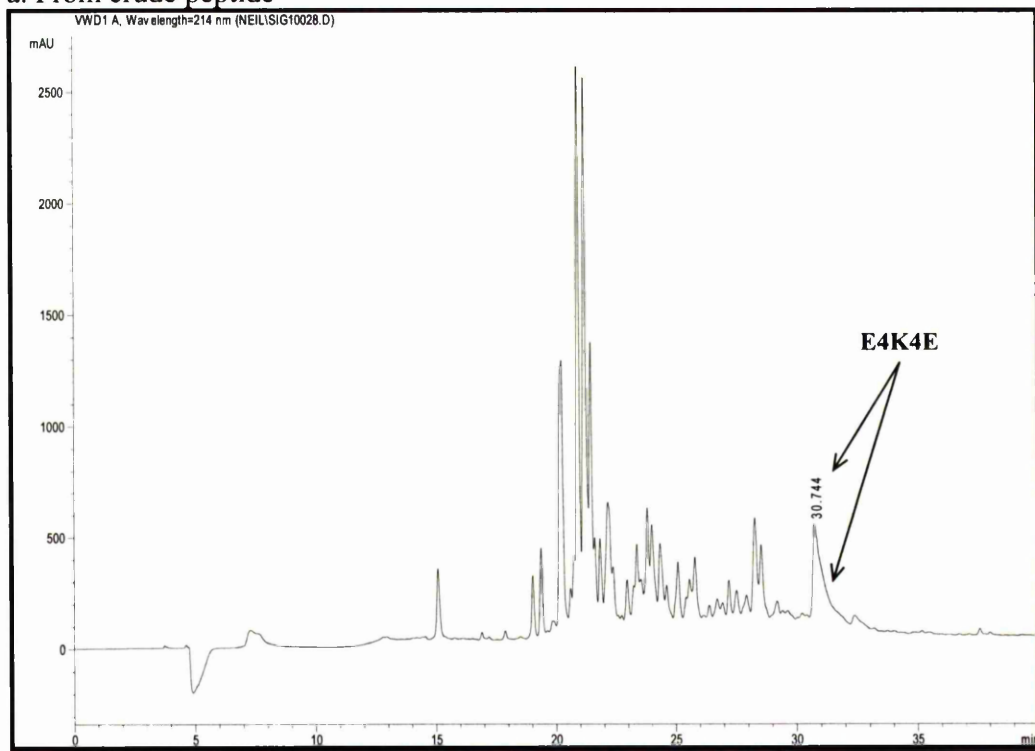
A.27. Identity confirmation of Ac-AAAEAAAKAAAAEAKGY-NH<sub>2</sub> peptide

A.28. Identity confirmation of Ac-AAAAEAAKAAAAEAKGY-NH<sub>2</sub> peptide

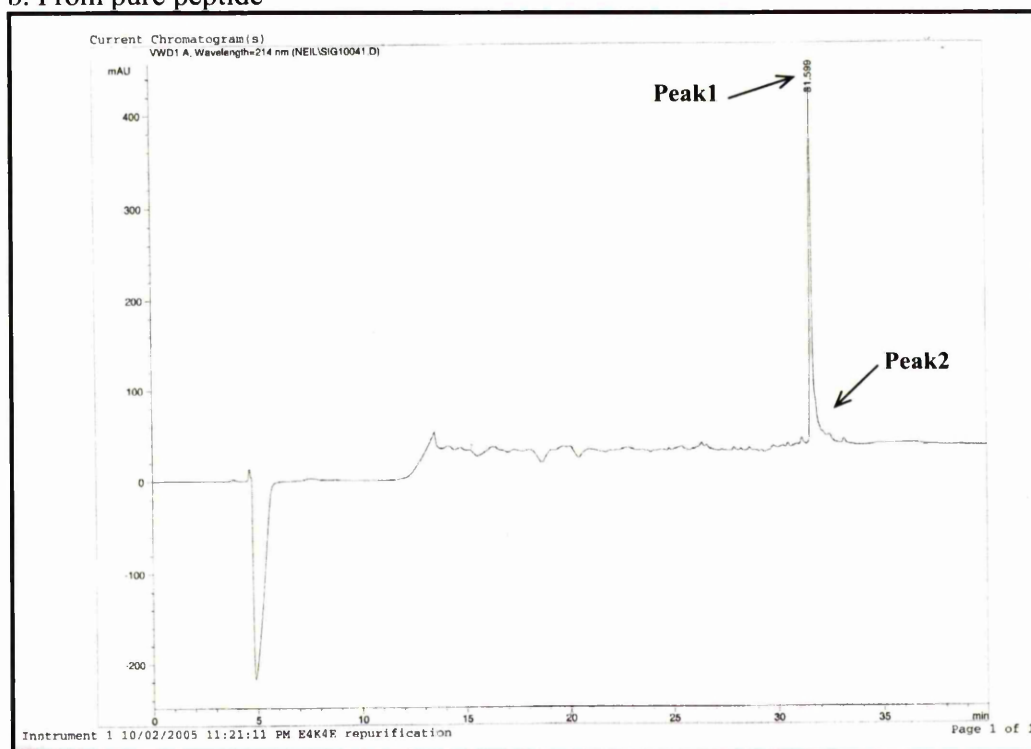
A.29. Identity confirmation of Ac-AAAEAAAAKAAAEAKAGY-NH<sub>2</sub> peptide

### A.30. Identity confirmation of Ac-AAAAEAAKAAAEAKAGY-NH<sub>2</sub> peptide

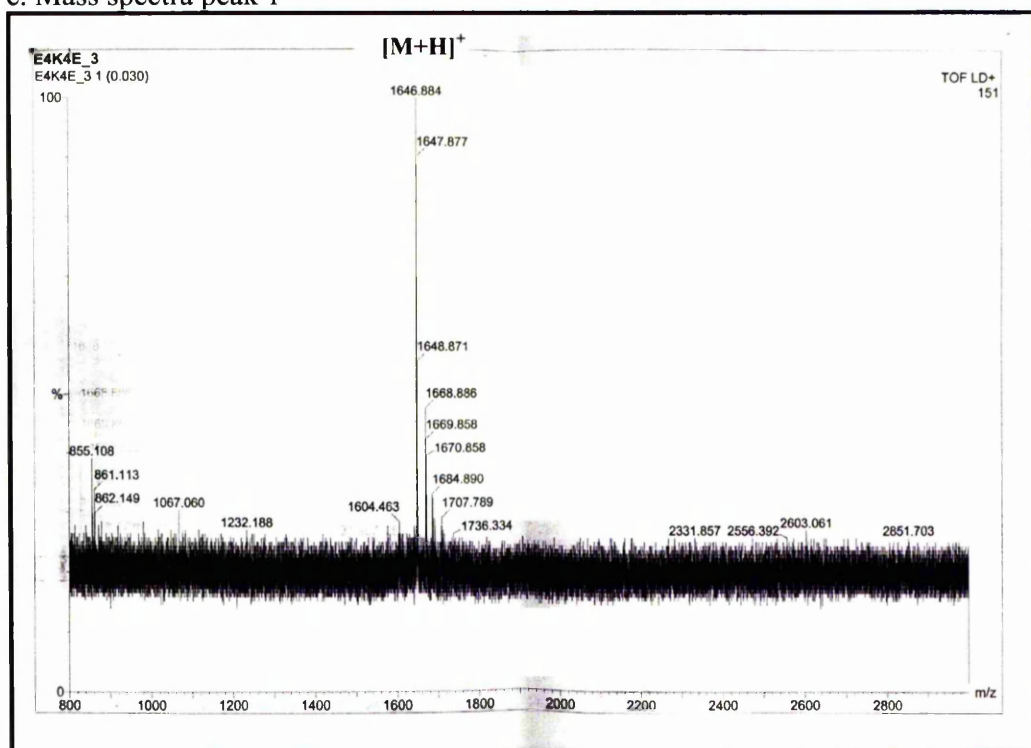
#### a. From crude peptide



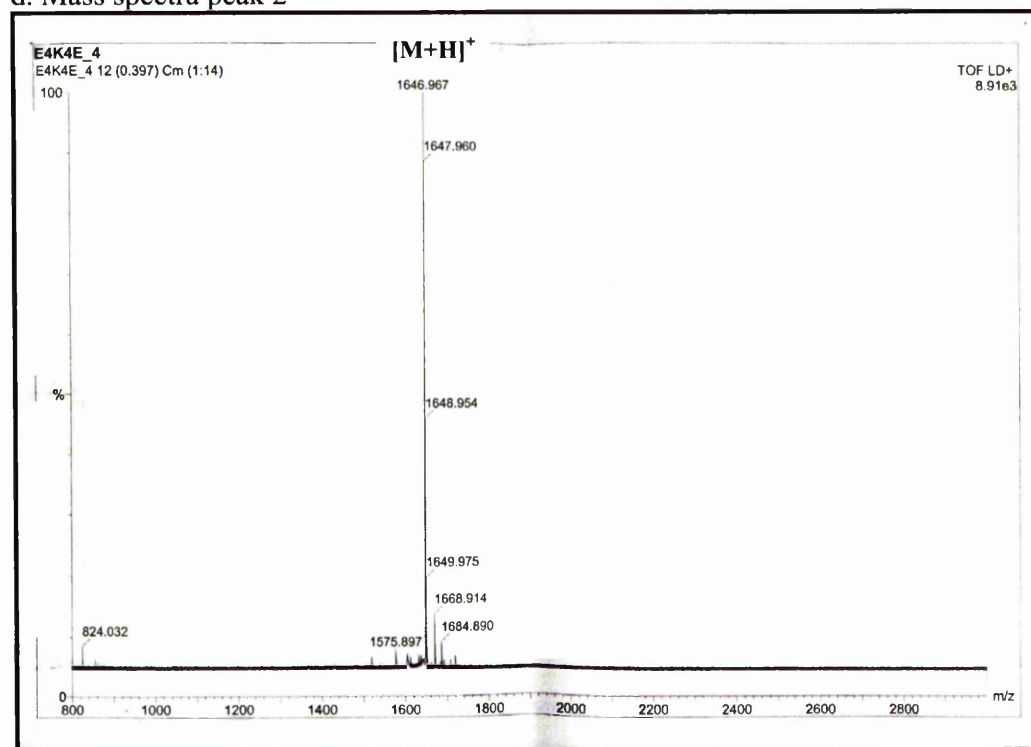
#### b. From pure peptide



## c. Mass spectra peak 1



## d. Mass spectra peak 2



**Appendix B. Protein dataset from the ASTRAL database (SCOP 1.63 Sequence****Resources)**

d16vpa_	d1aau_1	d1c0ma1	d1d3ya_	d1dtja_	d1ei7a_	d1f1ea_	d1fre_
d1a12a_	d1aau_2	d1c17m_	d1d4oa_	d1dtoa_	d1ejfa_	d1f1ma_	d1fs1a1
d1a1ia1	d1axn_	d1c1da2	d1d6ga_	d1duvg1	d1ejga_	d1f2va_	d1fs1b1
d1a1x_	d1ay7b_	d1c1ka_	d1d8ca_	d1dvka_	d1ek8a_	d1f32a_	d1fsga_
d1a26_1	d1ayl_1	d1c2aa1	d1d8ia_	d1dvoa_	d1ek9a_	d1f35a_	d1fsz_1
d1a2za_	d1ayl_2	d1c3ga1	d1daaa_	d1dwka1	d1ekra_	d1f3ua_	d1fsz_2
d1a34a_	d1ayoa_	d1c3ma_	d1dbha1	d1dwka2	d1el6a_	d1f3va_	d1ftra1
d1a3aa_	d1az9_1	d1c4ka3	d1dbxa_	d1dxea_	d1elka_	d1f44a1	d1fuia1
d1a48_	d1b25a1	d1c4za_	d1dcea2	d1dy5a_	d1elva2	d1f46a_	d1fuia2
d1a5t_1	d1b25a2	d1c5ea_	d1dcia_	d1dysa_	d1em9a_	d1f4la1	d1fvga_
d1a62_1	d1b2pa_	d1c75a_	d1dcpa_	d1dzfa1	d1emva_	d1f4la3	d1fvia2
d1a6da1	d1b33n_	d1c8za_	d1dcqa2	d1dzfa2	d1en2a1	d1f4pa_	d1fvka1
d1a6da3	d1b3aa_	d1c9la2	d1dd3c_	d1dzoa_	d1ep0a_	d1f52a1	d1fw9a_
d1a6m_	d1b3qa1	d1c9oa_	d1dd9a_	d1e0ta3	d1epua_	d1f5aa1	d1fwxa2
d1a6q_1	d1b3qa2	d1cbf_	d1de4c1	d1e19a_	d1eq6a_	d1f5aa3	d1fx2a_
d1a6q_2	d1b65a_	d1cby_	d1de4c2	d1e1aa_	d1eqja1	d1f5ma_	d1fx3a_
d1a79a1	d1b6a_2	d1cc8a_	d1deeg_	d1e20a_	d1eqoa_	d1f5na1	d1fx7a2
d1a79a2	d1b77a1	d1ccwa_	d1df0a2	d1e2aa_	d1erza_	d1f60a1	d1fx8a_
d1a8d_2	d1b79a_	d1ccwb_	d1dfca1	d1e2wa1	d1es6a_	d1f60a2	d1fxkc_
d1a8o_	d1b8aa2	d1cfza_	d1dfea_	d1e3ha1	d1etxa_	d1f60b_	d1fyea_
d1a9xa1	d1b8oa_	d1chd_	d1dfma_	d1e3ha5	d1eula1	d1f7la_	d1fyva_
d1a9xb1	d1b8za_	d1ci4a_	d1dfup_	d1e44b_	d1eula2	d1f7ua3	d1g10a_
d1aa7a_	d1bco_1	d1cipa1	d1dg6a_	d1e4cp_	d1eula1	d1f86a_	d1g13a_
d1aac_	d1bd8_	d1ciy_2	d1dhn_	d1e58a_	d1eula3	d1f8ya_	d1g2aa_
d1abv_	d1bdfa1	d1ciy_3	d1di2a_	d1e6ia_	d1eula4	d1f94a_	d1g2ra_
d1acc_	d1bdfa2	d1clia1	d1di6a_	d1e6ta_	d1euvb_	d1fcya_	d1g2ya_
d1aco_1	d1bdo_	d1clia2	d1div_1	d1e79a1	d1euwa_	d1ffgb_	d1g3p_1
d1aco_2	d1be3k_	d1cmia_	d1div_2	d1e79a2	d1ew4a_	d1fh6a_	d1g4yb_
d1ad2_	d1bf4a_	d1cola_	d1dj0a1	d1e79g_	d1ewfa1	d1fid_	d1g51a2
d1adt_1	d1bfd_2	d1cpo_1	d1dj7a_	d1e79i_	d1ewqa1	d1fjgb_	d1g5ga1
d1adt_2	d1bg1a1	d1cq3a_	d1dj8a_	d1e7la1	d1ewqa3	d1fjgc2	d1g5ha1
d1ae9a_	d1bgf_	d1crua_	d1dk0a_	d1e8ca1	d1ewqa4	d1fjgj_	d1g61a_
d1aep_	d1bhta1	d1crza1	d1dk8a_	d1e8oa_	d1exta1	d1fjgp_	d1g66a_
d1af7_1	d1bjt_	d1crza2	d1dkga1	d1eaic_	d1eyea_	d1fjgr_	d1g6ga_
d1ah7_	d1bkds_	d1csei_	d1dkza_	d1eaqa_	d1eyqa_	d1fjgs_	d1g6sa_
d1aho_	d1bkra_	d1csh_	d1dl2a_	d1eaza_	d1eyva_	d1fjgt_	d1g6xa_
d1aie_	d1ble_	d1ctf_	d1dl5a2	d1eb6a_	d1ez3a_	d1fjja_	d1g71a_
d1ail_	d1bm8_	d1cuk_1	d1dlja3	d1ebdc_	d1ezga_	d1fjla_	d1g72b_
d1ajj_	d1bmta1	d1cuk_2	d1dmga_	d1ecra_	d1ezvc2	d1fjra_	d1g73a_
d1ako_	d1boua_	d1cvra2	d1dmha_	d1ecwa_	d1ezvd2	d1fk5a_	d1g7sa3
d1al01_	d1boub_	d1cxqa_	d1doza_	d1edmb_	d1ezvc2	d1fkma1	d1g8ea_
d1aln_1	d1brwa3	d1cxzb_	d1dpjb_	d1eeja2	d1ezvf_	d1flel_	d1g8la1
d1amm_1	d1bwza1	d1cy5a_	d1dq3a2	d1eexb_	d1ezvg_	d1flma_	d1g8la2
d1amua_	d1bx4a_	d1cy9a_	d1dqaal	d1eexg_	d1ezvh_	d1fltv_	d1g8ma1
d1aoha_	d1bx7_	d1d09b1	d1dqaal	d1eflc_	d1ezvi_	d1fm0d_	d1g8ma2
d1aol_	d1bxoa_	d1d09b2	d1dqaal	d1efnb_	d1f00i1	d1fm0e_	d1g8qa_
d1aop_1	d1bxya_	d1d0ca_	d1dqia_	d1egwa_	d1f0ia1	d1fnta1	d1g9mg_
d1aop_3	d1by2_	d1d0qa_	d1ds1a_	d1ehkc_	d1f0ja_	d1fn9a_	d1g9za_
d1aqt_1	d1byi_	d1d2oa1	d1dt9a3	d1ehs_	d1f0la1	d1fpoa1	d1gci_
d1aqt_2	d1byqa_	d1d3va_	d1dtdb_	d1ei5a1	d1f0la3	d1fpoa2	d1gd8a_

d1gdna_	d1h3fa2	d1hxra_	d1liq4a_	d1jfb_	d1jx4a_	d1kala_	d1kxpd1
d1gdoa_	d1h3ia1	d1hyoa1	d1liq8a2	d1jfla1	d1jyaa_	d1kafa_	d1kyfa1
d1gk8a1	d1h3na2	d1hyoa2	d1liqqa_	d1jgl1a_	d1jyha_	d1kapp1	d1kyfa2
d1gk8a2	d1h4xa_	d1hz4a_	d1liqza_	d1jg5a_	d1jz8a1	d1kb0a2	d1kypa_
d1gk8i_	d1h59b_	d1hzta_	d1lirqa_	d1jh6a_	d1jz8a4	d1kbla2	d1kyqa2
d1gkma_	d1h5wa_	d1i12a_	d1isia_	d1jhda1	d1jzta_	d1kcfal	d1kzqa1
d1gkpa1	d1h6ha_	d1i19a1	d1itxa2	d1jhfb_	d1k04a_	d1kcca_	d1i0ia_
d1gkpa2	d1h6la_	d1i1wa_	d1iuua_	d1jhga_	d1k0ra2	d1keka4	d1i0qa1
d1gkza1	d1h6pa_	d1i27a_	d1iv3a_	d1ji7a_	d1k0ra4	d1keya_	d1i0sa_
d1gli1_	d1h72c1	d1i2ta_	d1ivwa1	d1jida_	d1k1fa_	d1kgsa1	d1i1sa_
d1gm5a1	d1h72c2	d1i31a_	d1ivwa2	d1jiwi_	d1k20a_	d1khca_	d1i2ha_
d1gmia_	d1h7ca_	d1i3ja_	d1iw7e_	d1jixa_	d1k24a_	d1khda1	d1i2pa_
d1gmua1	d1h8ga_	d1i40a_	d1iwga1	d1jj21_	d1k28a1	d1khda2	d1i3ka1
d1gmua2	d1h8la1	d1i4ja_	d1iwga5	d1jj2h_	d1k28a2	d1khia1	d1i3la2
d1gmxa_	d1h99a1	d1i4ma_	d1iwga7	d1jj2k_	d1k28d1	d1kid_	d1i4ia1
d1gnla_	d1h9db_	d1i50a_	d1ix9a1	d1jj2m_	d1k2yx1	d1kjna_	d1i4ia2
d1go3f_	d1hbka_	d1i50j_	d1ix9a2	d1jj2o_	d1k2yx4	d1kjqa1	d1i5ja1
d1go4a_	d1hbna1	d1i50l_	d1ixh_	d1jj2r_	d1k30a_	d1kjqa2	d1i5oa_
d1goia1	d1hbnc_	d1i6pa_	d1ixma_	d1jj2u_	d1k32a2	d1kjqa3	d1i6pa_
d1gp0a_	d1hc7a3	d1i71a_	d1j09a1	d1jj2w_	d1k32a3	d1kkea2	d1i6ra_
d1gpja1	d1hdha_	d1i7qa_	d1j3aa_	d1jj2x_	d1k3ba_	d1kkoa1	d1i7va_
d1gpja3	d1hela_	d1i8aa_	d1j57a_	d1jj2y_	d1k3ia2	d1kkoa2	d1i8ra_
d1gpma3	d1heta1	d1i9ba_	d1j5ua_	d1jk4a_	d1k3ia3	d1klxa_	d1i8wa_
d1gppa_	d1heta2	d1iata_	d1j5ya2	d1jkea_	d1k3xa1	d1km4a_	d1i9la_
d1gpr_	d1hf2a1	d1iaza_	d1j6qa_	d1jkxa_	d1k3xa2	d1kmta_	d1i9na2
d1gpua3	d1hf2a2	d1ib8a1	d1j6za1	d1jl0a_	d1k3xa3	d1kmva_	d1i9va1
d1gqea_	d1hf8a_	d1ib8a2	d1j8ba_	d1jlxal	d1k3ya1	d1knca_	d1i9va2
d1gqia2	d1hfel1	d1icfi_	d1j8ra_	d1jmla_	d1k4cc_	d1knma_	d1lam_1
d1grj_1	d1hfes_	d1iega_	d1j96a_	d1jmma_	d1k4ga_	d1knza_	d1lb2b_
d1gtda_	d1hg7a_	d1iexa2	d1j98a_	d1jmsa1	d1k4ia_	d1ko7a1	d1lb3a_
d1gtea1	d1hi9a_	d1ifqa_	d1j9la_	d1jmx5a	d1k4ta1	d1koia_	d1lb6a_
d1guta_	d1hlwa_	d1ifra_	d1jata_	d1jmxg_	d1k4za_	d1koya_	d1lba_
d1guxa_	d1hnja1	d1ig0a1	d1jb0a_	d1jn3a1	d1k5ca_	d1kp6a_	d1lbu_1
d1gvna_	d1ho1a_	d1ig0a2	d1jb0d_	d1jn3a2	d1k5ja_	d1kpf_	d1lc0a2
d1gvoa_	d1hoe_	d1igqa_	d1jb0f_	d1jnra1	d1k5na2	d1kp1a_	d1lc5a_
d1gwea_	d1hoza_	d1ihna_	d1jb0i_	d1jnra3	d1k5nb_	d1kpsb_	d1ld8a_
d1gxja_	d1hq0a_	d1ihra_	d1jb0j_	d1jo0a_	d1k5oa_	d1kpta_	d1ld8b_
d1gxma_	d1hq1a_	d1iiba_	d1jb0k_	d1jo8a_	d1k6da_	d1kq1a_	d1ldg_2
d1gxra_	d1hqka_	d1ijqa1	d1jb0l_	d1jofa_	d1k6ka_	d1kq3a_	d1ldja2
d1gxua_	d1hqsa_	d1ijva_	d1jb0m_	d1joga_	d1k6ya1	d1kq4a_	d1ldja3
d1gyha_	d1hs6a2	d1ijya_	d1jb0x_	d1jp3a_	d1k75a_	d1kqfb2	d1lfpa_
d1gyxa_	d1htya_	d1ik9a1	d1jb3a_	d1jqia1	d1k7ca_	d1kqfc_	d1lfwa2
d1gzbb1	d1hufa_	d1ikpa2	d1jb9a1	d1jqia2	d1k7ka_	d1kqpa_	d1lg7a_
d1gzsb_	d1hw1a2	d1ikpa3	d1jb9a2	d1jr2a_	d1k87a1	d1kr4a_	d1lj5a_
d1gzta_	d1hw5a2	d1ikta_	d1jbea_	d1jr8a_	d1k87a2	d1ku3a_	d1lkka_
d1h05a_	d1hw7a_	d1im5a_	d1jcha2	d1js8a2	d1k8ke_	d1kuua_	d1lkta_
d1h0xa_	d1hx0a1	d1io0a_	d1jcla_	d1jsda_	d1k8kf_	d1kwfa_	d1lla_1
d1h16a_	d1hx1b_	d1io1a_	d1jd5a_	d1jtgb_	d1k8kg_	d1kwma2	d1lla_2
d1h2rl_	d1hx6a1	d1ipaa2	d1jeya_	d1jvsa1	d1k8ta_	d1kwna_	d1llwa1
d1h2rs_	d1hxn_	d1ipca_	d1jf8a_	d1jw9b_	d1k92a2	d1kx9a_	d1lm5a_



d1lm8v_	d1m5ya1	d1n08a_	d1pda_2	d1qov1_	d1xsoa_	d2ptd_
d1lmia_	d1m6na1	d1n0ua4	d1pdo_	d1qpca_	d1xxaa_	d2pth_
d1lnia_	d1m6na2	d1n1ba2	d1pfka_	d1qpoa1	d1ycqa_	d2pvba_
d1lnsa1	d1m6ya1	d1n1ma1	d1pfo_	d1qpoa2	d1yge_1	d2reb_2
d1lnza1	d1m98a1	d1n2fa_	d1pgs_1	d1qqqa_	d1yge_2	d2sak_
d1lo7a_	d1m9sa2	d1n3la_	d1php_	d1qrea_	d1ytfb_	d2sici_
d1loua_	d1maz_	d1n45a_	d1phza1	d1qrva_	d1ytfc_	d2spca_
d1lpba1	d1mbya_	d1n4ka2	d1pina1	d1qsaa1	d1zeia_	d2tct_2
d1lp1a_	d1mc2a_	d1n55a_	d1pina2	d1qu9a_	d1zfja1	d2tpsa_
d1lq9a_	d1me4a_	d1n62a1	d1pprm1	d1rb9_	d1zfja2	d2uaga1
d1lqpa_	d1mfga_	d1n62a2	d1prea2	d1regx_	d1zin_2	d2uaga2
d1lqta2	d1mgpa_	d1n62b1	d1ptf_	d1rl6a1	d1zmec1	d2uaga3
d1lqvc_	d1mgta1	d1n62b2	d1ptq_	d1rlr_1	d1zymal	d3chbd_
d1lria_	d1mgta2	d1n62c1	d1puc_	d1rmd_2	d256ba_	d3cla_
d1lrra_	d1mhda_	d1n62c2	d1pysb4	d1rss_	d2a0b_	d3eipa_
d1lrza1	d1mi1a1	d1n7oa1	d1pysb6	d1scia_	d2ahja_	d3euga_
d1ls1a1	d1miua1	d1n7oa2	d1qc7a_	d1sfp_	d2ahjb_	d3ezma_
d1lshal	d1miua2	d1nbwa1	d1qcsa2	d1sgpi_	d2arca_	d3fapb_
d1lshb_	d1miwa1	d1nfn_	d1qcza_	d1sknp_	d2bbkh_	d3grs_3
d1ls1a1	d1mixa1	d1ni8a_	d1qd1a1	d1slua_	d2bbkl_	d3lkfa_
d1lt8a_	d1mj4a_	d1nj4a1	d1qd5a_	d1soxa3	d2bopa_	d3lzt_
d1ltza_	d1mjna_	d1nkd_	d1qdda_	d1sty_	d2btva_	d3nul_
d1lu0a_	d1mk0a_	d1nkp_	d1qexa_	d1svb_2	d2cba2_	d3procl
d1luca_	d1mkca_	d1nkza_	d1qf6a2	d1swua_	d2cpl_	d3pyp_
d1lvk_1	d1mkfa_	d1nls_	d1qf6a3	d1tl1da	d2end_	d3seb_2
d1lyva_	d1mkya3	d1nlxa_	d1qf8a_	d1tbge_	d2eng_	d3sil_
d1lzja_	d1ml9a_	d1nox_	d1qfma1	d1tfe_	d2erl_	d4bcl_
d1ml5a1	d1mla_1	d1ntha_	d1qgta_	d1tfr_1	d2hft_1	d4kbpa1
d1ml5a2	d1mla_2	d1o0ua_	d1qgwa_	d1tfr_2	d2igd_	d4mt2_
d1mlca_	d1mlva1	d1o0wa1	d1qh5a_	d1tif_	d2ilk_	d4ubpa_
d1mlcb_	d1mmsa2	d1o13a_	d1qhda1	d1tig_	d2liga_	d4ubpb_
d1mlfa_	d1mmsb_	d1o6sb_	d1qhoa2	d1tl2a_	d2lisa_	d5csma_
d1mlha1	d1mola_	d1o6va2	d1qhva_	d1tpg_2	d2mcm_	d7odca1
d1mlha2	d1mooa_	d1o70a1	d1qj4a_	d1tul_	d2mhr_	d7odca2
d1mlna_	d1mpga2	d1o75a1	d1qjba_	d1tx4a_	d2naca2	d8abp_
d1mlqa_	d1mr1c_	d1o75a3	d1qjpa_	d1lugia_	d2occa_	
d1mlxa1	d1mrj_	d1o7ja_	d1qkra_	d1uroa_	d2occb2	
d1mlxa4	d1msk_	d1o7na2	d1qksa2	d1ush_1	d2occc_	
d1mlxb3	d1msla_	d1o7nb_	d1ql0a_	d1utea_	d2occd_	
d1m22a_	d1mtyg_	d1o8ba1	d1qlac_	d1utg_	d2occe_	
d1m2da_	d1mun_	d1o8ba2	d1qlma_	d1luxy_2	d2occg_	
d1m2oa1	d1muwa_	d1oaca4	d1qmeal	d1vcc_	d2occh_	
d1m2oa2	d1mwpa_	d1oai_	d1qmga1	d1vfya_	d2occi_	
d1m2oa4	d1mwxa2	d1ois_	d1qmha1	d1vhh_	d2occj_	
d1m2oa5	d1mx_1	d1ospo_	d1qn1_1	d1vii_	d2occk_	
d1m40a_	d1mxma2	d1p32a_	d1qn1_2	d1vmoa_	d2occl_	
d1m41a_	d1mxma3	d1p35a_	d1qnxa_	d1vns_	d2occm_	
d1m4za_	d1mxta1	d1pa2a_	d1qopb_	d1wapa_	d2por_	
d1m55a_	d1mxta2	d1pbv_	d1qovh1	d1whi_	d2prgc_	
d1m56d_	d1mzga_	d1pcfa_	d1qovh2	d1who_	d2pspa1	

**Appendix C. Perl script to search  $\chi_1\chi_2$  rotamer at N3 position**

(This script only searches for Glu and Lys rotamers at N3. Other residues were searched in similar way by adding the corresponding lines)

```
#!/usr/bin/perl
$datafile = '/fs/wolf/home/mjfikti2/program/Perl_scripts/Rotamercount/c1c2_N3.out';
open (OUTPUT, ">$datafile");

do{
    $totnumE=0; $totnumgplusgplusE=0; $totnumgplustransE=0; $totnumgplusgminusE=0;
    $totnumtransgplusE=0; $totnumtranstransE=0; $totnumtransgminusE=0;
    $totnumgminusgplusE=0; $totnumgminustransE=0; $totnumgminusgminusE=0;
    $totnumK=0; $totnumgplusgplusK=0; $totnumgplustransK=0; $totnumgplusgminusK=0;
    $totnumtransgplusK=0; $totnumtranstransK=0; $totnumtransgminusK=0;
    $totnumgminusgplusK=0; $totnumgminustransK=0; $totnumgminusgminusK=0;

    do {
        $file = <STDIN>;
        chomp $file;
        if ($file eq 'eof'){
            print OUTPUT "Number of amino acids and c1c2 rotamer distribution in N3\n";
            print OUTPUT " E $totnumE g+g+ $totnumgplusgplusE g+t $totnumgplustransE g+g-
                $totnumgplusgminusE tg+ $totnumtransgplusE tt $totnumtranstransE tg- $totnumtransgminusE g-
                g+ $totnumgminusgplusE g-t $totnumgminustransE g-g- $totnumgminusgminusE;\n";
            print OUTPUT " K $totnumK g+g+ $totnumgplusgplusK g+t $totnumgplustransK g+g-
                $totnumgplusgminusK tg+ $totnumtransgplusK tt $totnumtranstransK tg- $totnumtransgminusK
                g-g+ $totnumgminusgplusK g-t $totnumgminustransK g-g- $totnumgminusgminusK;\n";

            print OUTPUT "TOTAL AMINO ACIDS IN ANY CONFORMATION $totaa\n";
            close (OUTPUT);
            exit;
            next;
        }

        $sst = '.sst';
        $dir = '/home/mjfikti2/program/Perl_scripts/SSTRUC_OUT/';
        $sstfile = $dir.$file.$sst;
        #print "the file being opened is $sstfile\n\n";

        unless ( open(SSTFILE, $sstfile) ) {
            print "sstfile doesn't exist - $sstfile \n\n";
            exit;
        }
        unless ( open(SSTFILE, $sstfile) ) {
            print "sstfile can't be opened - $sstfile \n\n";
            exit;
        }
        @strucfile = <SSTFILE>;
        #now we have sstfiles in array

        close <SSTFILE>;
        $sleng=$strucfile[3];
        $seqlength = substr($sleng,19,4);
        $totaa=$totaa+$seqlength;
        $Kabs = "";
        $fullseq = "";
        $aa = "";
        $ks = "";
        $secount = 0;
        $CHI1="";
        $CHI2="";
    }
}
```

```

while($secount < $seqlength) {
    $seq = $strucfile[$secount+7];
    $aa = substr($seq, 19, 1);
    $ks = substr($seq, 23, 1);
    $chis = $strucfile[$secount + 15 + $seqlength];
    $chi1 = substr($chis, 29, 6);
    $chi2 = substr($chis, 36, 6);
    $CHI1 = $CHI1.$chi1;
    $CHI2 = $CHI2.$chi2;
    $fullseq = $fullseq.$aa;
    $Kabs = $Kabs.$ks;
    $secount = $secount+1;
}

$numE=0; $numgplusgplusE=0; $numgplustransE=0; $numgplusgminusE=0; $numtransgplusE=0;
    $numtranstransE=0; $numtransgminusE=0; $numgminusgplusE=0; $numgminustransE=0;
    $numgminusgminusE=0;
$numK=0; $numgplusgplusK=0; $numgplustransK=0; $numgplusgminusK=0; $numtransgplusK=0;
    $numtranstransK=0; $numtransgminusK=0; $numgminusgplusK=0; $numgminustransK=0;
    $numgminusgminusK=0;

$oldKabs = "";
$oldKabs1 = "";
$oldKabs2 = "";
$n3 = "";
$motif = "";
$c1 = "";
$c2 = "";
$c1rot = "";
$c2rot = "";
$posn = 0;

while ($posn < $seqlength){
    $motif = substr($Kabs, $posn, 4);
    $oldKabs = substr($Kabs, $posn-3, 1);
    $oldKabs1 = substr($Kabs, $posn-2, 1);
    $oldKabs2 = substr($Kabs, $posn-1, 1);
    $n3 = substr($fullseq, $posn, 1);
    $c1 = substr($CHI1, $posn*6, 6);
    $c2 = substr($CHI2, $posn*6, 6);

if    ($c1 == 999.9){
    $c1rot = 'notdef';
    } elsif ((0<$c1) and ($c1<120)) {
    $c1rot = 'g-';
    } elsif ((-120<$c1) and ($c1<0)) {
    $c1rot = 'g+';
    } else {
    $c1rot = 't';
    }

if    ($c2 == 999.9){
    $c2rot = 'notdef';
    } elsif ((0<$c2) and ($c2<120)) {
    $c2rot = 'g-';
    } elsif ((-120<$c2) and ($c2<0)) {
    $c2rot = 'g+';
    } else {
    $c2rot = 't';
    }

#print "$file $motif $n3 $c1rot\n";

if    ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq 'E' and
    $c1rot eq 'g+' and $c2rot eq 'g+') {
    $numE=$numE+1;
    $numgplusgplusE=$numgplusgplusE+1;

```

```

} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'E' and $c1rot eq 'g+' and $c2rot eq 't') {
$numE=$numE+1;
$numgplustransE=$numgplustransE+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'E' and $c1rot eq 'g+' and $c2rot eq 'g-') {
$numE=$numE+1;
$numgplusgminusE=$numgplusgminusE+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'E' and $c1rot eq 't' and $c2rot eq 'g+') {
$numE=$numE+1;
$numtransgplusE=$numtransgplusE+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'E' and $c1rot eq 't' and $c2rot eq 't') {
$numE=$numE+1;
$numtranstransE=$numtranstransE+1;

} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'E' and $c1rot eq 't' and $c2rot eq 'g-') {
$numE=$numE+1;
$numtransgminusE=$numtransgminusE+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'E' and $c1rot eq 'g-' and $c2rot eq 'g+') {
$numE=$numE+1;
$numgminusgplusE=$numgminusgplusE+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'E' and $c1rot eq 'g-' and $c2rot eq 't') {
$numE=$numE+1;
$numgminustransE=$numgminustransE+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'E' and $c1rot eq 'g-' and $c2rot eq 'g-') {
$numE=$numE+1;
$numgminusgminusE=$numgminusgminusE+1;
}

if ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq 'K' and
$c1rot eq 'g+' and $c2rot eq 'g+') {
$numK=$numK+1;
$numgplusgplusK=$numgplusgplusK+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'K' and $c1rot eq 'g+' and $c2rot eq 't') {
$numK=$numK+1;
$numgplustransK=$numgplustransK+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'K' and $c1rot eq 'g+' and $c2rot eq 'g-') {
$numK=$numK+1;
$numgplusgminusK=$numgplusgminusK+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'K' and $c1rot eq 't' and $c2rot eq 'g+') {
$numK=$numK+1;
$numtransgplusK=$numtransgplusK+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'K' and $c1rot eq 't' and $c2rot eq 't') {
$numK=$numK+1;
$numtranstransK=$numtranstransK+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'K' and $c1rot eq 't' and $c2rot eq 'g-') {
$numK=$numK+1;
$numtransgminusK=$numtransgminusK+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'K' and $c1rot eq 'g-' and $c2rot eq 'g+') {
$numK=$numK+1;
$numgminusgplusK=$numgminusgplusK+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
'K' and $c1rot eq 'g-' and $c2rot eq 't') {
$numK=$numK+1;
}

```

```

$numgminustransK=$numgminustransK+1;
} elsif ($motif eq 'HHHH' and $oldKabs ne 'H' and $oldKabs1 eq 'H' and $oldKabs2 eq 'H' and $n3 eq
      'K' and $c1rot eq 'g-' and $c2rot eq 'g-') {
$numK=$numK+1;
$numgminusgminusK=$numgminusgminusK+1;
}
} {
$posn = $posn+1;
}
}
$totnumE=$totnumE+$numE; $totnumgplusgplusE=$totnumgplusgplusE+$numgplusgplusE;
    $totnumgplustransE=$totnumgplustransE+$numgplustransE;
    $totnumgplusgminusE=$totnumgplusgminusE+$numgplusgminusE;
    $totnumtransgplusE=$totnumtransgplusE+$numtransgplusE;
    $totnumtranstransE=$totnumtranstransE+$numtranstransE;
    $totnumtransgminusE=$totnumtransgminusE+$numtransgminusE;
    $totnumgminusgplusE=$totnumgminusgplusE+$numgminusgplusE;
    $totnumgminustransE=$totnumgminustransE+$numgminustransE;
    $totnumgminusgminusE=$totnumgminusgminusE+$numgminusgminusE;

$totnumK=$totnumK+$numK; $totnumgplusgplusK=$totnumgplusgplusK+$numgplusgplusK;
    $totnumgplustransK=$totnumgplustransK+$numgplustransK;
    $totnumgplusgminusK=$totnumgplusgminusK+$numgplusgminusK;
    $totnumtransgplusK=$totnumtransgplusK+$numtransgplusK;
    $totnumtranstransK=$totnumtranstransK+$numtranstransK;
    $totnumtransgminusK=$totnumtransgminusK+$numtransgminusK;
    $totnumgminusgplusK=$totnumgminusgplusK+$numgminusgplusK;
    $totnumgminustransK=$totnumgminustransK+$numgminustransK;
    $totnumgminusgminusK=$totnumgminusgminusK+$numgminusgminusK;

} until ($file eq 'eof');
}

```

## Appendix D. Perl script to search amino acid pairs and their rotamers

### D.1. Probing the EK pair

```
#!/usr/bin/perl
#This program will look for the EXXXK motif\n";
#using non redundant dataset from ASTRAL for protein domains including the rotamers for E and K \n";

    $datafile = '/fs/wolf/home/mjfikti2/program/Perl_scripts/EKE/EK.out';
    open (OUTPUT, ">$datafile");
    $round = 0;

    do {
        $directory = '/home/mjfikti2/program/Perl_scripts/SSTRUC_OUT/';
#print "enter the name of the sstfile:\n\n";

        $file = <STDIN>;
        chomp $file;

if      ($file eq 'eof'){
    print OUTPUT "total number of EK is $totnumEK \n";
    close (OUTPUT);
    exit;
    }

        $sst = '.sst';
#print "filename given is $file\n\n";

        $sstfile = $directory.$file.$sst;
#print "the file you are opening is $sstfile\n\n";

        unless ( open(SSTFILE, $sstfile) ) {
            print "sstfile doesn't exist - $sstfile \n\n";
            exit;
        }
        unless ( open(SSTFILE, $sstfile) ) {
            print "sstfile can't be opened - $sstfile \n\n";
            exit;
        }
        @strucfile = <SSTFILE>;
#now we have sstfile in an array

        close SSTFILE;
        $sleng = $strucfile[3];
        $seqlength = substr($sleng, 19, 4);
        $totaa = $totaa + $seqlength;
        $totnumEK = 0;
        $numEK=0;
        $Kabs = "";
        $fullseq = "";
        $aa = "";
        $ks = "";
        $seccount = 0;
        $CHI1="";
        $CHI2="";
#get the length of the sequence and print it

        while($seccount < $seqlength) {
            $seq = $strucfile[$seccount+7];
            $schis = $strucfile[$seccount + 15 + $seqlength];
            $chi1 = substr($schis, 29, 6);
            $chi2 = substr($schis, 36, 6);
            $aa = substr($seq, 19, 1);
            $ks = substr($seq, 23, 1);
            $CHI1 = $CHI1.$chi1;
```

```

$CHI2 = $CHI2.$chi2;
$fullseq = $fullseq.$aa;
$Kabs = $Kabs.$ks;
$seccount = $seccount+1;
}
$posn=0;

do {
    $motif = substr($fullseq, $posn, 11);
    $four = substr($motif, 3, 1);
    $eight = substr($motif, 7, 1);
    $chi1E = substr($CHI1, (($posn*6)+18), 6);
    $chi2E = substr($CHI2, (($posn*6)+18), 6);
    $chi1K = substr($CHI1, (($posn*6)+42), 6);
    $chi2K = substr($CHI2, (($posn*6)+42), 6);

if    ($chi1F == 999.9){
    $Ec1rot = 'notdef';
    } elsif ((0 < $chi1E) and ($chi1E < 120)) {
    $Ec1rot = 'g-';
    } elsif ((-120 < $chi1E) and ($chi1E < 0)) {
    $Ec1rot = 'g+';
    } else {
    $Ec1rot = 't';
    }

if    ($chi2E == 999.9){
    $Ec2rot = 'notdef';
    } elsif ((0 < $chi2E) and ($chi2E < 120)) {
    $Ec2rot = 'g-';
    } elsif ((-120 < $chi2E) and ($chi2E < 0)) {
    $Ec2rot = 'g+';
    } else {
    $Ec2rot = 't';
    }

if    ($chi1K == 999.9){
    $Kc1rot = 'notdef';
    } elsif ((0 < $chi1K) and ($chi1K < 120)) {
    $Kc1rot = 'g-';
    } elsif ((-120 < $chi1K) and ($chi1K < 0)) {
    $Kc1rot = 'g+';
    } else {
    $Kc1rot = 't';
    }

if    ($chi2K == 999.9){
    $Kc2rot = 'notdef';
    } elsif ((0 < $chi2K) and ($chi2K < 120)) {
    $Kc2rot = 'g-';
    } elsif ((-120 < $chi2K) and ($chi2K < 0)) {
    $Kc2rot = 'g+';
    } else {
    $Kc2rot = 't';
    }

if    ($four eq 'E' and $eight eq 'K') {
    $numEK = $numEK + 1;
    $secstruc = substr($Kabs, $posn, 11);
    #print "helix found at $helpos \n\n";
    #print "old Kabs = $oldKabs \n\n";

    print OUTPUT "EK motif starts $posn file $file E chi1 = $chi1E rotamer $Ec1rot E chi2 = $chi2E
        rotamer $Ec2rot K chi1 = $chi1K rotamer $Kc1rot K chi2 = $chi2K rotamer $Kc2rot
        secondary structure, $secstruc,\n";
    #print OUTPUT "The secondary structure is $secstruc \n";

```

```

$posn = $posn + 1;
} else {
$posn = $posn + 1;
}

} until ( $posn > $seqlength-10);
$totnumEK = $totnumEK + $numEK;
#print OUTPUT "TOTAL number of EK is $totnumEK \n";

} until ($file eq 'eof');
close (OUTPUT);
exit;

```

## D.2. Probing the EKE triplet

```

#!/usr/bin/perl
#This program will look for the EXXXKXXE motif\n";
#using non redundant dataset from ASTRAL for protein domains including the rotamers for E, K and E \n";

```

```

$datafile = '/fs/wolf/home/mjfikti2/program/Perl_scripts/EKE/EKE.out';
open (OUTPUT, ">$datafile");
$round = 0;

do {
$directory = '/home/mjfikti2/program/Perl_scripts/SSTRUC_OUT/';
#print "enter the name of the sstfile:\n\n";

$file = <STDIN>;
chomp $file;

if ($file eq 'eof'){
print OUTPUT "total number of EKE is $totnumEKE \n";
close (OUTPUT);
exit;
}
$sst = '.sst';
#print "filename given is $file\n\n";

$sstfile = $directory.$file.$sst;
#print "the file you are opening is $sstfile\n\n";

unless ( open(SSTFILE, $sstfile) ) {
print "sstfile doesn't exist - $sstfile \n\n";
exit;
}
unless ( open(SSTFILE, $sstfile) ) {
print "sstfile can't be opened - $sstfile \n\n";
exit;
}
@strucfile = <SSTFILE>;

#now we have sstfile in an array

```

```

close SSTFILE;

$sleng = $strucfile[3];
$seqlength = substr($sleng, 19, 4);
$totaa = $totaa + $seqlength;
$totnumEKE = 0;
$numEKE=0;
$Kabs = "";
$fullseq = "";
$aas = "";
$ks = "";
$seccount = 0;

```



```

$CHI1="";
$CHI2="";
#get the length of the sequence and print it

while($seccount < $seqlength) {
$seq = $strucfile[$seccount+7];
$chis = $strucfile[$seccount + 15 + $seqlength];
$chi1 = substr($chis, 29, 6);
$chi2 = substr($chis, 36, 6);
$aas = substr($seq, 19, 1);
$ks = substr($seq, 23, 1);
$CHI1 = $CHI1.$chi1;
$CHI2 = $CHI2.$chi2;
$fullseq = $fullseq.$aas;
$Kabs = $Kabs.$ks;
$seccount = $seccount+1;
}
$posn=0;

do {
$motif = substr($fullseq, $posn, 9);
$one = substr($motif, 0, 1);
$five = substr($motif, 4, 1);
$nine = substr($motif, 8, 1);
$chi1E = substr($CHI1, ($posn*6), 6);
$chi2E = substr($CHI2, ($posn*6), 6);
$chi1K = substr($CHI1, (($posn*6)+24), 6);
$chi2K = substr($CHI2, (($posn*6)+24), 6);
$chi1E = substr($CHI1, (($posn*6)+48), 6);
$chi2E = substr($CHI2, (($posn*6)+48), 6);

if ($chi1E == 999.9){
$Ec1rot = 'notdef';
} elsif ((0 < $chi1E) and ($chi1E < 120)) {
$Ec1rot = 'g-';
} elsif ((-120 < $chi1E) and ($chi1E < 0)) {
$Ec1rot = 'g+';
} else {
$Ec1rot = 't';
}

if ($chi2E == 999.9){
$Ec2rot = 'notdef';
} elsif ((0 < $chi2E) and ($chi2E < 120)) {
$Ec2rot = 'g-';
} elsif ((-120 < $chi2E) and ($chi2E < 0)) {
$Ec2rot = 'g+';
} else {
$Ec2rot = 't';
}

if ($chi1K == 999.9){
$Kc1rot = 'notdef';
} elsif ((0 < $chi1K) and ($chi1K < 120)) {
$Kc1rot = 'g-';
} elsif ((-120 < $chi1K) and ($chi1K < 0)) {
$Kc1rot = 'g+';
} else {
$Kc1rot = 't';
}

if ($chi2K == 999.9){
$Kc2rot = 'notdef';
} elsif ((0 < $chi2K) and ($chi2K < 120)) {
$Kc2rot = 'g-';
} elsif ((-120 < $chi2K) and ($chi2K < 0)) {
$Kc2rot = 'g+';
} else {

```

```

        $Kc2rot = 't';
    }
    if ($chi1E == 999.9){
        $Ec1rot = 'notdef';
    } elsif ((0 < $chi1E) and ($chi1E < 120)) {
        $Ec1rot = 'g-';
    } elsif ((-120 < $chi1E) and ($chi1E < 0)) {
        $Ec1rot = 'g+';
    } else {
        $Ec1rot = 't';
    }
    if ($chi2E == 999.9){
        $Ec2rot = 'notdef';
    } elsif ((0 < $chi2E) and ($chi2E < 120)) {
        $Ec2rot = 'g-';
    } elsif ((-120 < $chi2E) and ($chi2E < 0)) {
        $Ec2rot = 'g+';
    } else {
        $Ec2rot = 't';
    }
}

if ($one eq 'E' and $five eq 'K' and $nine eq 'E'){
    $numEKE = $numEKE + 1;
    $secstruc = substr($Kabs, $posn, 9);
    #print "old Kabs = $oldKabs \n\n";

    print OUTPUT "EKE motif starts $posn file $file E chi1 = $chi1E rotamer $Ec1rot E chi2 = $chi2E
        rotamer $Ec2rot K chi1 = $chi1K rotamer $Kc1rot K chi2 = $chi2K rotamer $Kc2rot E chi1
        = $chi1E $Ec1rot E chi2 = $chi2E rotamer $Ec2rot secondary structure, $secstruc,\n";
    #print OUTPUT "The secondary structure is $secstruc \n";

    $posn = $posn + 1;
    } else {
        $posn = $posn + 1;
    }

    } until ( $posn > $seqlength-8);
    $totnumEKE = $totnumEKE + $numEKE;
    #print OUTPUT "TOTAL number of EKE is $totnumEKE \n";
    } until ($file eq 'eof');

close (OUTPUT);
exit;

```



# Effect of the N3 residue on the stability of the $\alpha$ -helix

TEUKU M. IQBALSIAH AND ANDREW J. DOIG

Department of Biomolecular Sciences, University of Manchester Institute of Science and Technology (UMIST), Manchester M60 1QD, UK

(RECEIVED July 30, 2003; FINAL REVISION September 16, 2003; ACCEPTED September 23, 2003)

## Abstract

N3 is the third position from the N terminus in the  $\alpha$ -helix with helical backbone dihedral angles. All 20 amino acids have been placed in the N3 position of a synthetic helical peptide ( $\text{CH}_3\text{CO}$ -[AAX AAAAKAAAAKAGY]- $\text{NH}_2$ ) and the helix content measured by circular dichroism spectroscopy at 273 K. The dependence of peptide helicity on N3 residue identity has been used to determine a free energy scale by analysis with a modified Lifson-Roig helix coil theory that includes a parameter for the N3 energy ( $n_3$ ). The most stabilizing residues at N3 in rank order are Ala, Glu, Met/Ile, Leu, Lys, Ser, Gln, Thr, Tyr, Phe, Asp, His, and Trp. Free energies for the most destabilizing residues (Cys, Gly, Asn, Arg, and Pro) could not be fitted. The results correlate with N1, N2, and helix interior energies and not at all with N-cap preferences. This completes our work on studying the structural and energetic preferences of the amino acids for the N-terminal positions of the  $\alpha$ -helix. These results can be used to rationally modify protein stability, help design helices, and improve prediction of helix location and stability.

**Keywords:**  $\alpha$ -helix; N3 position; circular dichroism; protein folding; protein stability; helix propensities; helix-coil theory; macrodipole

The  $\alpha$ -helix is the most frequently observed secondary structure in proteins. N1, N2, and N3 are the first three amino acids with helical  $\phi$ ,  $\psi$  backbone dihedral angles at the helix N terminus. They differ from interior positions as their amide NH groups do not participate in backbone-backbone  $i, i+4$  hydrogen bonds within the helix. The presence of these otherwise unsatisfied hydrogen bond donors has profound structural effects, most often satisfied by hydrogen bonds to side chains local in sequence, such as to preceding N-cap side chains.

In native proteins, Ala, Asp, Gln, and Glu have the highest propensity for the N3 position. We have made a thorough study of the structures adopted by side chains at the N-cap, N1, N2, and N3 positions in proteins (Doig et al. 1997; Penel et al. 1999). Asn, Asp, Gln, Glu, and Thr occasionally form  $i, i$  hydrogen bonds to the N3 backbone NH group, although these bonds may be weak, as they are non-linear, particularly for the shorter side chains (Penel et al.

1999; Wan and Milner-White 1999). A second important feature of the N3 position is the capping box, where the N3 side chain accepts a hydrogen bond from the N-cap backbone NH group (Harper and Rose 1993). This is most important for Gln or Glu at N3. Helix termini are highly solvent exposed, and tertiary interactions are rare, so the environment of peptide and protein helices are very similar. The unique structural trends for N3 imply that the free energies for substituting amino acids at this position differ from all other positions in the helix. In this work, we measure these free energies using helical peptides.

$\alpha$ -Helices in aqueous solution adopt a large number of structures, with fully helical, fully coil, and partly helical conformations all populated. For a complete understanding of helix formation and stability, all the factors contributing to this equilibrium need to be assessed thoroughly. These include the helix-forming tendencies of constituent amino acids, capping preferences at the carboxyl and amino termini, and preferences for the N1, N2, and N3 positions. Our helix-coil model N1N2N3 (Sun et al. 2000) distinguishes between each of these by assigning them unique weights. Factors such as side-chain-main-chain hydrogen bonds, electrostatic interactions with the helix dipole, solvent exposure, and conformational entropy are implicitly included

Reprint request to: Andrew J. Doig, Department of Biomolecular Sciences, UMIST, P.O. Box 88, Manchester M60 1QD, UK; e-mail: Andrew.Doig@umist.ac.uk; fax: 44-161-236-0409.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.03341804>.

in the positional preferences. Side-chain interactions between residues spaced  $i, i+3$  and  $i, i+4$  are also of considerable importance to helix stability, although are not included in the N1N2N3 helix/coil model or the peptides studied here.

We recently made a study of the energetics of the N1 and N2 helix position for all 20 amino acid residues (Cochran and Doig 2001; Cochran et al. 2001). Previous studies have measured position-dependent effects on helix content of several amino acid residues (Asp [Huyghues-Despointes et al. 1993], Glu [Scholtz et al. 1993], His [Armstrong and Baldwin 1993], Val, Leu, Met, Ile, Gly, Ser, Thr, Asn, Gln [Petukhov et al. 1998, 1999]), but a complete survey of all 20 residues in the same peptide is needed. The energetic contribution of the N3 residue to helix stability in the controlled environment of a synthetic peptide is of fundamental interest, with practical implications in the development of secondary structure prediction algorithms, peptide, and protein design, and development of site-directed mutagenesis strategies.

## Results and Discussion

### *Design of the N3 peptide sequence*

N3 preferences can be determined by substituting the third residue in an acetylated peptide and measuring the change in helix content. In this study an intrinsically helical peptide was used to measure the energetic contribution (the  $n3$  parameter) of 19 amino acids in the N3 position in the same way that the  $n1$  and  $n2$  parameters have recently been investigated (Cochran and Doig 2001; Cochran et al. 2001). We used an AK-based peptide with a sequence of  $\text{CH}_3\text{CO}-[\text{AAXAAAKAAAKAGY}]-\text{NH}_2$  to study the energetic of the N3 residue in relative isolation from neighboring charges. This peptide is intrinsically helical because of its high Ala content. X is the variable residue. The N terminus is acetylated and the positively charged lysine residues included for solubility are spaced  $i, i+5$  to each other and the substitution site so they do not interact. The Tyr residue is present to give the peptide a UV absorption for determination of concentration. The penultimate Gly ensures that the  $\alpha$ -helix almost always terminates at, or before, this position so that problems arising from a helical Tyr affecting the CD signal are minimized (Chakrabartty et al. 1993).

The acetyl group is a strong N-cap so helix formation will often initiate at the acetyl group. Ala is a poor N-cap and a relatively good N1 and N2, so conformations that initiate with the first two Ala residues at the N-cap, and hence, X at N1 or N2 will have low populations. The third residue (X) will therefore have a high probability of being at the N3 position, increasing the sensitivity of the helix content to the N3 preference.

The stability of the substituted peptides is readily evaluated by measurements of CD intensity at 222 nm, which is proportional to the average helicity of the peptide in solution. The N3 preference can thus be found by finding the value of  $n3$  that gives predicted helix content in agreement with the experiment. No side-chain interactions are present in this sequence. The energy scales of the interior and N-cap of Rohl et al. (1996) as well as the energy scales of N1 (Cochran et al. 2001) and N2 (Cochran and Doig 2001) are used in the calculations, leaving the N3 preference ( $n3$  in the helix-coil theory) values as the only unknown parameter.

The N termini of naturally occurring helices in proteins are very often solvent exposed (Doig et al. 1997). Thus, the AK peptide is a good model for the study of factors that stabilize the  $\alpha$ -helix in naturally occurring proteins.

### *Helix contents of N3 peptides*

The helix content of each peptide was measured by CD at 222 nm at a concentration of 20  $\mu\text{M}$  in 10 mM NaCl, 5 mM sodium phosphate pH 7.0 at 273 K. Low-temperature measurements are made because the helix-coil equilibrium favors the helical conformation at lower temperature so signal intensities are consequently larger. Molar ellipticities for these types of simple peptides are known to be concentration independent (Padmanabhan et al. 1990; Stapley and Doig 1997) indicating that aggregation does not occur. This was confirmed by checking that the helix contents of N3(Glu) and N3(Ile) were invariant between 5–200  $\mu\text{M}$ . The CD signal at 222 nm was converted to  $[\theta]_{222}$  (see Materials and Methods), and the percentage helicity calculated with  $\pm 3\%$  uncertainty as experimental error (Table 1). N3(Cys) was measured in the presence of 1 mM dithiothreitol to keep it reduced.

The peptide containing Val at the N3 position unexpectedly formed a  $\beta$ -sheet structure at a concentration higher than 20  $\mu\text{M}$ , as shown by CD. At lower concentrations this peptide formed a random coil structure (data not shown). As the Val peptide did not form a helix, the data for N3(Val) could not be obtained. We find it remarkable that the peptide with Ile at N3 is monomeric and helical up to 200  $\mu\text{M}$ , while the substitution of this Ile for the smaller and less hydrophobic Val causes it to aggregate into a  $\beta$ -sheet amyloid at only 20  $\mu\text{M}$ . We suspect that the amyloid structure is the most stable form for many, if not all, of the peptides, and it is only a very high activation energy that allows the peptides to form monomeric helices, in agreement with the ideas of Dobson (1999).

The percentage helicities for the series of homologous peptides are shown in Table 1, and range widely from 59% for Ala to 16% for Pro. This large range indicates that the N3 residue plays a significant role in the overall stability of the peptide. However, the equilibrium that exists between the multitudes of partially helical states must be analyzed

**Table 1.** Helix contents of N3 peptides and energetic preferences for the N3 position

N3 residue	$[\theta]_{222}^a$	% Helix <sup>b</sup>	AGADIR % helix prediction <sup>c</sup>	$n3$ value <sup>d</sup> (range $\pm$ 3% helicity)	$n3w$ value	$\Delta G$ for coil-to-N3 transition <sup>e</sup>	$\Delta\Delta G^f$
A	-19,970	59.2	59	0.59 (0.42 to 0.91)	1.00 (0.72 to 1.55)	0.00 (-0.24 to 0.18)	0
E <sup>-</sup>	-17,019	50.4	58	0.64 (0.51 to 0.85)	0.35 (0.27 to 0.46)	0.57 (0.70 to 0.42)	
M	-15,381	45.6	46	0.42 (0.35 to 0.54)	0.28 (0.22 to 0.35)	0.70 (0.81 to 0.57)	0.36
I	-15,428	45.7	48	0.60 (0.48 to 0.76)	0.28 (0.22 to 0.35)	0.70 (0.81 to 0.57)	0.54
L	-15,231	45.1	51	0.28 (0.23 to 0.35)	0.24 (0.20 to 0.30)	0.77 (0.88 to 0.65)	0.36
K	-13,348	40.0	46	0.18 (0.15 to 0.21)	0.18 (0.15 to 0.21)	0.93 (1.03 to 0.84)	
E <sup>0</sup>	-13,478	39.9		<sup>h</sup>			
S	-13,723	40.7	44	0.31 (0.27 to 0.35)	0.13 (0.11 to 0.14)	1.13 (1.21 to 1.07)	0.65
Q	-14,749	43.7	53	0.41 (0.34 to 0.50)	0.12 (0.10 to 0.14)	1.16 (1.26 to 1.05)	0.25
T	-12,800	37.9	43	0.64 (0.56 to 0.67)	0.12 (0.10 to 0.12)	1.17 (1.24 to 1.15)	0.65
Y	-13,434	39.8	44	0.23 (0.19 to 0.26)	0.11 (0.09 to 0.13)	1.20 (1.29 to 1.12)	
D <sup>0</sup>	-11,820	35.0		<sup>h</sup>			
F	-11,595	34.4	39	0.31 (0.28 to 0.32)	0.08 (0.07 to 0.09)	1.35 (1.41 to 1.34)	
D <sup>-</sup>	-14,594	43.2	52	0.20 (0.18 to 0.19)	0.08 (0.07 to 0.08)	1.15 (1.21 to 1.15)	
W	-10,835	32.1	44	0.002 (- to 0.039)	0.001 (0 to 0.001)	4.04 (2.43 to — <sup>g</sup> )	
H <sup>0</sup>	-9645	28.6	36	0.023 (0.042 to — <sup>g</sup> )	0.008 (0 to 0.015)	2.60 (2.27 to — <sup>g</sup> )	
C <sup>0</sup>	-9414	27.9	44	<sup>h</sup>		—	
G	-9309	27.6	39	0 <sup>g</sup>	0	—	0.76
N	-9133	27.1	43	0 <sup>g</sup>	0	—	0.7
R	-7103	21.1	52	0 <sup>g</sup>	0	—	
H	-5636	16.7		<sup>h</sup>		—	
P	-5486	16.3	16	0 <sup>g</sup>	0	—	

<sup>a</sup> The peptide sequence is CH<sub>3</sub>CO-AAXA<sub>4</sub>KA<sub>4</sub>KAGY-NH<sub>2</sub> [ $\theta$ ]<sub>222</sub> units are deg · cm<sup>2</sup> · dmole<sup>-1</sup>.

<sup>b</sup> Calculated as  $(-[\theta]_{222} \times 100)/(-40000 \times (1-2.5/N))$ .

<sup>c</sup> Reference (Muñoz and Serrano 1994).

<sup>d</sup> From fitting with N1N2N3 program.

<sup>e</sup>  $-RT \ln(n3w)$ , in kcal mole<sup>-1</sup>.

<sup>f</sup> From Petukhov et al. 1998, 1999.

<sup>g</sup> Not possible to fit a positive value for  $n3$ .

<sup>h</sup> Not fitted.

using the helix-coil theory. This makes it possible to separate all the possible effects that can arise from making the substitution, because X is at the N-cap, N1, N2, or N3 positions or in a coil state in different conformations. The contribution of the helix dipole is dealt implicitly by the helix coil theory, as its effect is subsumed within the  $n3$  parameter.

#### Determination of $n3$ values and $\Delta G$ from helix contents

The residue  $n3$  values are presented in rank order in Table 1. The order is different when compared with the order of helicity values. This confirms that analysis using the helix-coil theory is essential for understanding the equilibrium. For example, a high helix content in one of these peptides may arise from a strong preference for N-cap, N1, N2, or N3, leading to high populations for that residue to be at each of these positions.

The  $n3$  values were calculated from helicity measurements as described previously for the N1 and N2 positions (Cochran et al. 2001; Cochran and Doig 2001) using the N1N2N3 helix coil algorithm (Sun et al. 2000). In these sequences the only unknown parameter is the N3 preference

( $n3$ ). Initially,  $n3$  for the four sequences with an X residue matching residues of the generic peptide (Ala, Gly, Lys, and Tyr) were solved simultaneously by fitting the equilibrium equation to the observed helicities using the N1N2N3 program (see Materials and Methods). Then,  $n3$  for other residues were found by varying them until the calculated helix content agreed with experiment.

The statistical weight for the N3 coil to helix reaction is  $n3w$ , the product of  $w$ , the intrinsic interior preference value, and  $n3$ , the intrinsic N3 preference value (Sun et al. 2000). Free energies for transfer from the coil to N3 are thus calculated as  $\Delta G = -RT \ln(n3w)$  and the free energy change relative to Ala is given by  $\Delta\Delta G = \Delta G_{Xaa} - \Delta G_{Ala}$ .

$n3$  is an adjustment to the intrinsic helix preference ( $w$ ), and thus shows how the helix preference changes on moving from the helix interior to N3. Table 1 presents the free energies of each amino acid in rank order of decreasing stability. The  $n3$  values for several N3 peptides (C<sup>-</sup>, C<sup>0</sup>, G, H<sup>+</sup>, H<sup>0</sup>, N, P, and R) could not be evaluated as positive numbers, and have been assigned a value of zero. An approximate rank order for their N3 preferences can be seen in the helix contents of the peptides, however. It is not clear why these data could not be fitted. Possibilities include

errors in previously determined parameters, interactions in nonhelical conformations and interactions involving these amino acids that extend beyond the N3 position. The  $n_3$ -values for W, Y, and F are unreliable because aromatic groups perturb the CD signal (Chakrabarty et al. 1993), although this is a minor effect for Phe and Tyr (Andrew et al. 2002). This aromatic effect may well differ at N3 from the helix center, due to the differing conformations between these sites. The N-cap energy of Asp<sup>0</sup>, Cys<sup>0</sup>, Glu<sup>0</sup>, and His<sup>+</sup> have not been measured previously. The  $n_3$  values of these protonated amino acids were therefore omitted, because this requires also knowing their N-cap preferences.

The  $n$ -cap,  $n_1$ ,  $n_2$ , and  $n_3$  parameters represent the probabilities that a particular residue will adopt an N-cap, N1, N2, or N3 conformation, respectively. These parameters are interdependent, so the  $n_3$  values could be perturbed by errors in  $n$ ,  $n_1$ , or  $n_2$  values. This problem is most important when X in the sequence has a high N-cap preference, such as Asp, Asn, and Cys, as conformations with these residues at the N-cap will be more populated. When the  $n$ -cap value for an amino acid is large and of similar magnitude to the acetyl group ( $n$ -cap = 5.9; Doig et al. 1994), it will spend much of its time in a nonhelical N-cap conformation. The large  $n$ -cap values for Asp (6.6), Asn (6.8), Cys (5.4), and Tyr (4.9) tend to shift the equilibrium in favor for the residue X to be at N-cap. Thus, the helices will nucleate with X as the N-cap. In contrast, when X has a low N-cap preference, the acetyl will then be the N-cap, X will be at N3. The conformation when X is the N-cap will therefore be rarely populated and an error in  $n(X)$  will be small.

All amino acids in N3 peptides destabilize the helix relative to Ala. The free energies range from 0 kcal·mole<sup>-1</sup> for N3(Ala) to 4 kcal·mole<sup>-1</sup> for N3(Trp). The highest stabilizing effect of N3(Ala) is not surprising because Ala is close to the center of the sequence where the preference is high. Helix breaking residues at the third position of the peptide have a larger destabilizing effect than at N1, as they are moving toward the center of the sequence. For instance,  $\Delta\Delta G$  for N1(Pro) is 0.6 kcal·mole<sup>-1</sup> and  $\Delta\Delta G$  N1(Gly) is 1.06 kcal·mole<sup>-1</sup>. In this study the  $\Delta\Delta G$  for these residues could not be measured as their  $\Delta G$  are very high.

The AGADIR program (Muñoz and Serrano 1994) predicts the helix contents of monomeric peptides in aqueous solution so work such as this offers a valuable blind test of its accuracy. Table 1 lists predictions of helix contents for all the peptides studied here. Errors range from 0% (Ala and Met) to 31% (Arg). The AGADIR results are all quite similar, apart from the accurate low value for Pro, in contrast to the large variation in helix contents we find when varying the N3 residue.

#### Comparison with other reported $n_3$ values

Free energies for N3 for nonpolar and noncharged polar residues for a series of AR peptides have been reported

(Petukhov et al. 1998, 1999), and are included in Table 1 for comparison. No other  $\Delta\Delta G(N3)$  values have been reported. The free energies measured here are somewhat higher than those of Petukhov et al. This may be due to differences in their peptide sequence and computational methods of Petukhov et al. The peptides studied here are 16 mers, with a wider helicity range, from 59% for N3(Ala) to 16% for N3(Pro), while those of Petukhov et al. range from 28% of N3(Thr) to 47% of N3(Ala). Our data may therefore give more accurate N3 energies.

#### pH Titrations

The helicities of N3(Asp), N3(Cys), N3(Glu), and N3(His) peptides were titrated as a function of pH. pKa values of the ionizable side chain groups were evaluated between approximately pH 2–9 by curve fitting to the Henderson-Hasselbach equation (Fig. 1A–D). The upper range of accessible pH is given by Lys deprotonation, which leads to peptide aggregation.

Titration of N3(Asp) and N3(Glu) demonstrate that the neutrally charged species are both destabilizing compared to the negatively charged residues. As the pH rises above the pKa, helix content increases maximally by 8% and 10%, respectively. As previously described, the  $n_3$  values calculation of protonated Asp and Glu were omitted as their N-cap values are unknown. The equation to calculate the free energy for protonation, given by  $\Delta G = -RT \ln[n_3(\text{charged})/n_3(\text{uncharged})]$  could therefore not be calculated.

The pKa of N3(Glu) is  $4.48 \pm 0.11$ , and this value is similar to model compounds (Glu in an N-acetyl acid amide has a pKa of 4.5; Nozaki and Tanford 1967). The unperturbed pKa value for residue Glu at N3 indicates that the side chain is not involved in any strong interactions, yet the helix content is strongly influenced by the ionization state.

The pKa of N3(Asp) of  $3.64 \pm 0.09$  is below those of other reported compounds. In comparison with the pKa value of Asp in N-acetyl acid amide (pKa 4.1; Nozaki and Tanford 1967), N3(Asp) is about 0.5 units lower. The pKa of Asp has also been measured in the coil state of CH<sub>3</sub>CO-[AADAA]-NH<sub>2</sub> and reported as  $3.91 \pm 0.02$  (Huyghues-Despointes et al. 1993). The significant difference between the perturbations of the pKas of Asp and Glu could be attributed to the shorter side chain of Asp, interacting more strongly with the helix dipole.

The pKa of N3(Cys) of  $7.73 \pm 0.12$  differs from other reported pKa values (Miranda 2003). However, the pKa value of Cys in a helical peptide is significantly lower than that in a coil peptide (pKa Cys  $8.69 \pm 0.07$ ; Kortemme and Creighton 1995). The increase in helix content of N3(Cys) at higher pH was not determined because titration beyond pH 8.5 in this peptide begins to deprotonate the side chain of Lys residues leading to peptide aggregation.

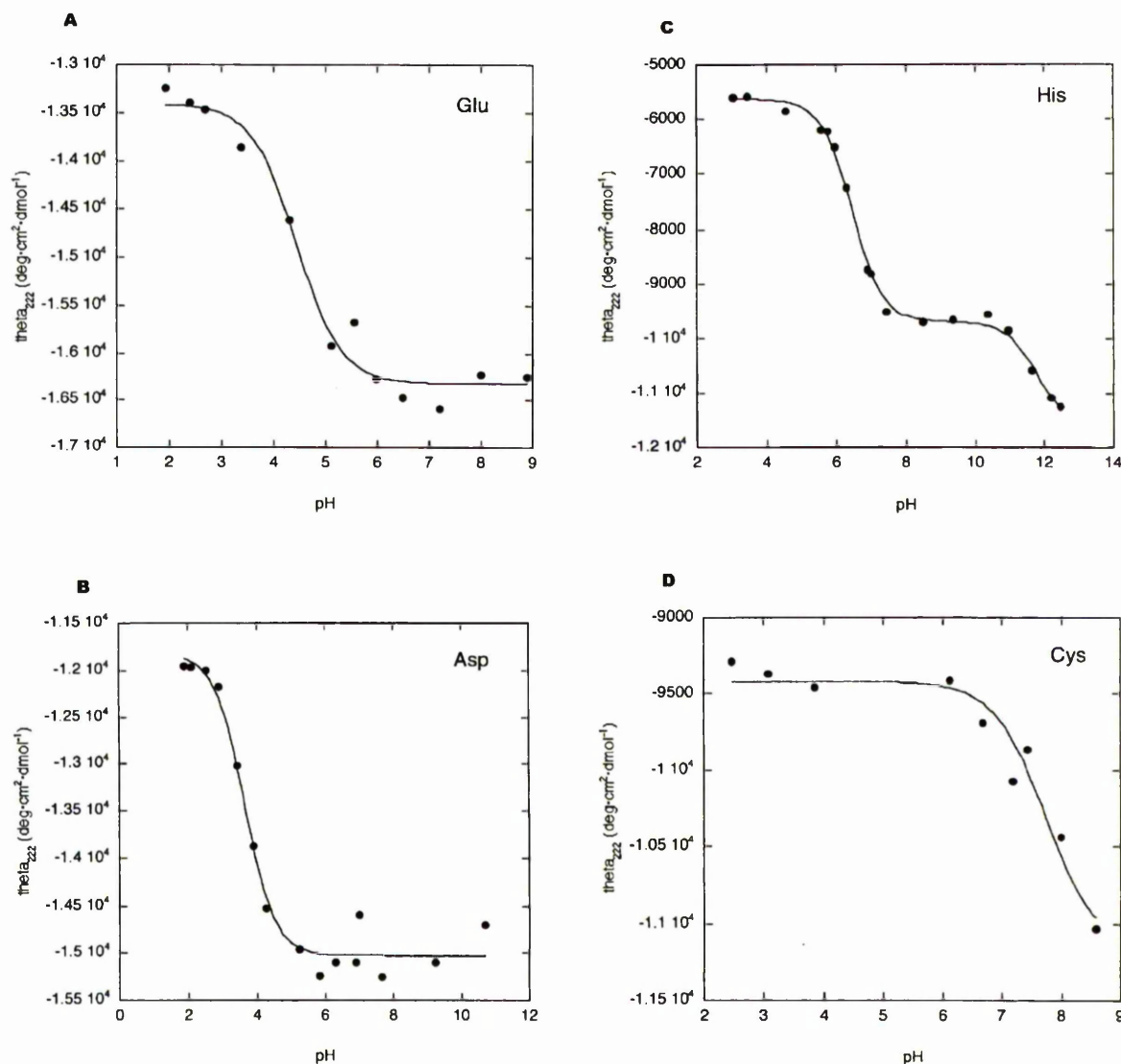


Figure 1. pH titrations for N3 peptides. (A) N3(Glu), (B) N3(Asp), (C) N3(His), (D) N3(Cys).

In general, Asp, Glu, and Cys residues are helix stabilizing at N3 when in the ionized form. The stabilizing contribution might be due to the ability of these residues to form a hydrogen bond to one of unsatisfied amide groups at the N terminus (Penel et al. 1999). Another reasonable contribution to stabilization is the electrostatic interactions of the charged side chains with the positive end of the helix dipole. These two contributions cannot be separated experimentally, however.

The helicity of N3(His) is also pH dependent (Fig. 1D). The uncharged N3(His) at pH 9 is about 9% more helical than the fully protonated species at pH 3. The His<sup>+</sup> side chain destabilizes the helix, possibly as a result of repulsion with the positive end of the helix macrodipole and free

amide groups at the N terminus. The pK<sub>a</sub> for His at N3 is  $6.45 \pm 0.04$ , which is comparable to values commonly found in the range of 6.5 and 7.1 (Kyte 1995). His–dipole interactions have also been measured (Armstrong and Baldwin 1993). His<sup>+</sup> is destabilizing to the helix, and this effect is reduced in 1 M NaCl. The pK<sub>a</sub> of His increases as the His is moved further from the N terminus. This was again interpreted as a charge–dipole effect. The titration above pH 11 is due to Tyr deprotonation.

#### Correlation of N3 energies with N3 propensities

Figure 2 shows the correlation between the N3 energies measured here and the N3 amino acid propensities found by

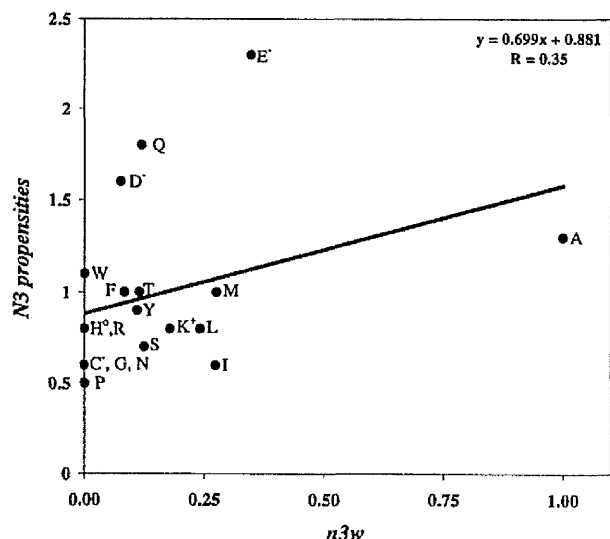


Figure 2. Correlation between N3 propensity ( $P_g N3$ ) and N3 statistical weight ( $n3w$ ).

surveying helices in proteins (Penel et al. 1999). There is a weak correlation observed. Deviations from the line of best fit for Glu, Gln, and Asp show that these are more favored at N3 in proteins than in this peptide. This is due to these side chains forming capping box hydrogen bonds to the backbone NH of the N-cap (Harper and Rose 1993). In the peptides the N-cap is an acetyl that lacks this NH group so this stabilizing hydrogen bond cannot form. In principle, it is possible to design a series of peptides that could form this capping box bond to measure its contribution to helix formation. In practice, however, this is difficult. If the peptides had Gly at their N terminus, they could be studied at high pH, where the N terminus is  $H_2N-CH_2-$ . These peptides would have low helix contents, particularly at their N termini, as Gly is a poor N-cap, making N3 energies inaccurate. They would also need to be studied at high pH where the N-terminal amine is neutral. At high pH the peptides are likely to be insoluble, as Lys side chains would lose their positive charges. In addition, the titration of Tyr would interfere with helix content measurements.

Although propensities are measured relative to all other positions in a protein, N3 energies are measured relative to the coil state. Note that the statistical propensities observed in proteins are not free energies, although they often correlate with them. Surveying protein structures is thus no substitute for experimental measurements of free energies from amino acid substitutions, whether by peptide synthesis or site-directed mutagenesis in proteins.

#### Comparison of N3 preferences with N-cap, N1, N2, and helix interior preferences

Figure 3 shows a plot of the N-cap preference (Rohl et al. 1996) against the N3 preference ( $n3w$ ). There is no corre-

lation, showing that the amino acid preferences for the N-cap position are completely different to N3. Although both are being at the helix N terminus, they adopt different conformations as N-cap and N3 is in nonhelical and helical conformations, respectively.

Figures 4 and 5 compare the N1 and N2 preferences (Cochran et al. 2001; Cochran and Doig 2001) to the N3 preference, respectively. There is a reasonable correlation showing that the forces at both sites are similar. The notable deviations from the linear correlation are Asp<sup>-</sup> and Glu<sup>-</sup> in both comparisons of which have higher N1 and N2 preferences than N3, respectively. A possible reason for this is that Asp<sup>-</sup> and Glu<sup>-</sup> at N1 and N2 can form  $i, i+1$  hydrogen bonds with the adjacent backbone amide group (Penel et al. 1999). This is not possible when they are at N3 because the amide group at N4 is occupied hydrogen bonding to the N-cap C=O.

Figure 6 compares preferences at N3 and helix interior positions. There is a good correlation showing that the N3 and interior environments are similar. Deviations from a linear correlation shown by Arg<sup>+</sup> and Lys<sup>+</sup> can be attributed to charge effects where both residues are disfavored at N3 compared to the helix interior. In general, residues with the ability to hydrogen bond with free amide groups at the helix N terminus (e.g., Asp<sup>0</sup>, Glu, Ser, and Thr) are more stable at N3 than at helix interior.

#### Conclusion

This article completes our work on studying the structural and energetic preferences of the amino acids for the N-terminal positions of the  $\alpha$ -helix. The substitution of amino

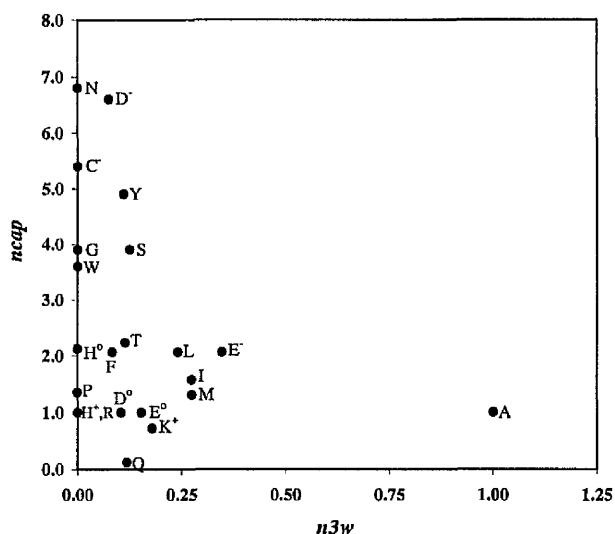
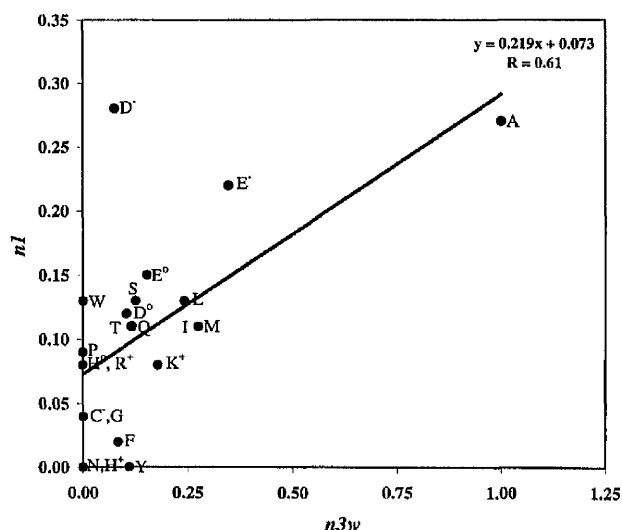


Figure 3. Correlation between N-cap statistical weight ( $n$ ) and N3 statistical weight ( $n3w$ ).





**Figure 4.** Correlation between N1 statistical weight ( $nI$ ) and N3 statistical weight ( $n3w$ ).

acids at the N3 position in an AK peptide has shown the unique structural trends for N3 and yielded the range of intrinsic preferences for residues at this position in the  $\alpha$ -helix. Although the preferences at N3 correlate well to those at the helix interior, the N3 position favors negatively charged residues due to electrostatic interactions with helix dipole. We suggest that this information can be used to rationally modify protein stability, particularly as helix N termini are solvent-exposed and very rarely form tertiary interactions.

## Materials and methods

### Peptide synthesis

Peptides were synthesized by the solid-phase method. C-terminal amides were made using Rink Amide resin (CN Biosciences). N termini were acetylated with 4% v/v acetic anhydride / 5% v/v pyridine dimethylformamide for 30 min at room temperature. Peptides were cleaved from the resin using 95% trifluoroacetic acid, 2.5% triisopropylsilane, 2.5% water. Peptides were purified by  $C_{18}$  reverse-phase HPLC and their molecular weights were checked by ES mass spectrometry. Final purity was checked by analytical  $C_{18}$  HPLC. Solutions of peptides containing Cys were maintained in a reduced state with 1 mM dithiothreitol.

Peptide concentration was determined by measuring tyrosine UV absorbance of aliquots of stock solution dissolved in water or 6 M guanidine hydrochloride using  $\epsilon_{275} = 1390 \text{ M}^{-1} \cdot \text{cm}^{-1}$  or  $\epsilon_{275} = 1450 \text{ M}^{-1} \cdot \text{cm}^{-1}$ , respectively (Brandts and Kaplan 1973). Concentrations of peptides containing Trp were determined in 6M GuHCl, pH 7.0, 298 K using  $\epsilon_{281}(\text{Trp}) = 5690 \text{ M}^{-1} \cdot \text{cm}^{-1} + \epsilon_{281}(\text{Tyr}) = 1250 \text{ M}^{-1} \cdot \text{cm}^{-1}$  (Edelhoc 1967).

### Circular dichroism measurements

Equilibrium CD measurements were made at 222 nm with a Jasco J810 spectropolarimeter at 273 K. Ellipticity was measured in 10

mM NaCl, 5 mM sodium phosphate pH 7.0 in a quartz cell with a 0.5-cm path length. The Cys peptide was measured in the presence of 1 mM dithiothreitol. CD data in mdeg were converted to mean residue ellipticity using  $[\theta]_{222} = \theta / (\text{Molar conc.} \times 16 \text{ residues} \times 10)$ , in  $\text{deg} \cdot \text{cm}^2 \cdot \text{dmole}^{-1}$ . Helix content was calculated as  $[\theta]_{222}(\text{observed}) / [\theta]_{222}(\text{max})$ .  $[\theta]_{222}(\text{max})$  is given by  $-40,000 \cdot (1 - 2.5/n)$ , where  $n$  is the number of amino acids in the peptide (Chen et al. 1974).

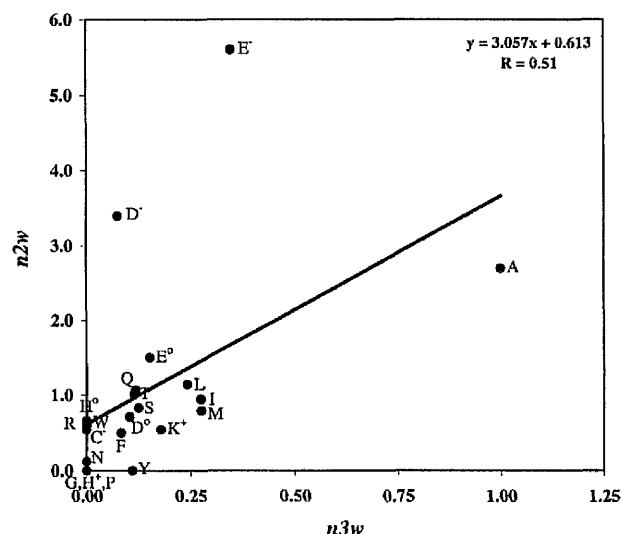
pH titrations were performed in a 3-mL 1.0-cm path length quartz cell at 273 K. The buffer used for pH titrations contained 10 mM NaCl, 1 mM sodium phosphate, 1 mM sodium borate, and 1 mM sodium citrate. The pH meter was calibrated for measurements at 273 K. The pH was adjusted during the titrations with aliquots of HCl or NaOH. Data were only used in the pH range where the equilibrium is reversible. The ellipticity data from the titrations were fitted to a Henderson-Hasselbach equation:

$$[\theta]_{222nm} = [\theta]_{222, \text{high\_pH}} \cdot \left(1 - \frac{1}{1 + 10^{pH - pKa}}\right) + [\theta]_{222, \text{low\_pH}} \cdot \left(\frac{1}{1 + 10^{pH - pKa}}\right)$$

$[\theta]_{222, \text{high\_pH}}$  and  $[\theta]_{222, \text{low\_pH}}$  are the molar ellipticities measured at 222 nm at the titration end points at high and low pH.

### Determination of helix-coil parameters

$n3$ -Values were determined using a statistical mechanical algorithm implemented in the program N1N2N3 (<http://www.bi.umist.ac.uk/users/mjfajdg/n1n2n3.htm>). N1N2N3 implements the modified Lifson-Roig helix-coil theory (Sun et al. 2000) to calculate the helix content of a given peptide sequence. The inputs to the program are the peptide sequences and corresponding experimental helix contents; a library of  $w$ ,  $n\text{-cap}$ ,  $n1$ ,  $n2$ ,  $n3$ ,  $c1$ , and  $c\text{-cap}$  values for each amino acid; an  $n\text{-cap}$  value for acetyl and a  $c\text{-cap}$  value for amide; and a list of parameters to be determined. All  $c$ -values were set to 1; other parameters were taken from Rohl et



**Figure 5.** Correlation between N2 statistical weight ( $n2w$ ) and N3 statistical weight ( $n3w$ ).

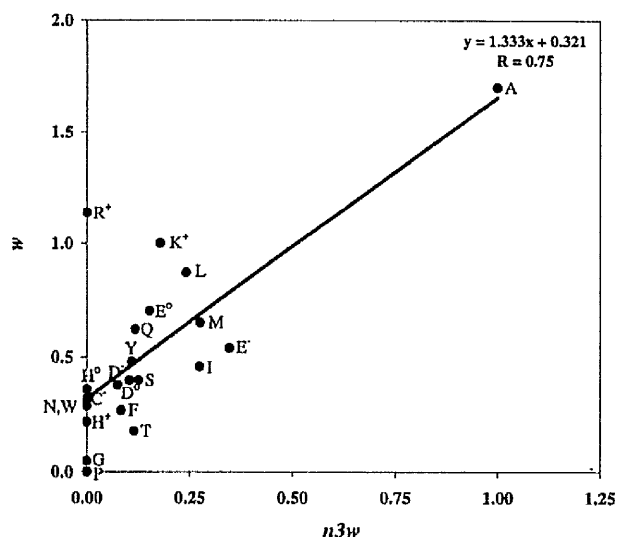


Figure 6. Correlation between helix interior statistical weight ( $w$ ) and N3 statistical weight ( $n3w$ ).

al. (1996), Cochran et al. (2001), or Cochran and Doig (2001). Parameters were varied until they converged on essentially unchanging values.

Results using the AGADIR program (Muñoz and Serrano 1994), which predicts the helix content of any peptide, were obtained from the Web site <http://www.embl-heidelberg.de/cgi/agadir-wrapper.pl>.

## Acknowledgments

The Michael Barber Mass Spectrometry Facility at UMIST is thanked for verifying peptide identity. T.M.I. thanks the Technological and Professional Skills Development Project of the Ministry of National Education of The Republic of Indonesia for a scholarship. We thank the Wellcome Trust (grant no. 057318) for an equipment grant for a CD spectrometer.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## References

- Andrew, C.D., Bhattacharjee, S., Kokkon, N., Hirst, J.D., Jones, G.R., and Doig, A.J. 2002. Stabilising interactions between aromatic and basic side chains in  $\alpha$ -helical peptides and proteins. Tyrosine effects on helix circular dichroism. *J. Am. Chem. Soc.* **124**: 12706–12714.
- Armstrong, K.M. and Baldwin, R.L. 1993. Charged histidine affects  $\alpha$ -helix stability at all positions in the helix by interacting with the backbone charges. *Proc. Natl. Acad. Sci.* **90**: 11337–11340.
- Brandts, J.R. and Kaplan, K.J. 1973. Derivative spectroscopy applied to tyrosyl chromophores. Studies on ribonuclease, lima bean inhibitor, and pancreatic trypsin inhibitor. *Biochemistry* **10**: 470–476.
- Chakrabarty, A., Kortemme, T., Padmanabhan, S., and Baldwin, R.L. 1993. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* **32**: 5560–5565.
- Chen, Y.-H., Yang, J.T., and Chau, K.H. 1974. Determination of the helix and  $\beta$  form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**: 3350–3359.
- Cochran, D.A.E. and Doig, A.J. 2001. Effects of the N2 residue on the stability of the  $\alpha$ -helix for all 20 amino acids. *Protein Sci.* **10**: 1305–1311.
- Cochran, D.A.E., Penel, S., and Doig, A.J. 2001. Contribution of the N1 amino acid residue to the stability of the  $\alpha$ -helix. *Protein Sci.* **10**: 463–470.
- Dobson, C.M. 1999. Protein misfolding, evolution and disease. *Trends Biol. Sci.* **24**: 329–332.
- Doig, A.J., Chakrabarty, A., Klingler, T.M., and Baldwin, R.L. 1994. Determination of free energies of N-capping in  $\alpha$ -helices by modification of the Lifson-Roig helix-coil theory to include N- and C-capping. *Biochemistry* **33**: 3396–3403.
- Doig, A.J., MacArthur, M.W., Stapley, B.J., and Thornton, J.M. 1997. Structures of N-termini of helices in proteins. *Protein Sci.* **6**: 147–155.
- Edelhoch, H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**: 1948–1954.
- Harper, E.T. and Rose, G.D. 1993. Helix stop signals in proteins and peptides: The capping box. *Biochemistry* **32**: 7605–7609.
- Huyghues-Despointes, B.M., Scholtz, J.M., and Baldwin, R.L. 1993. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. *Protein Sci.* **2**: 1604–1611.
- Kortemme, T. and Creighton, T.E. 1995. Ionisation of cysteine residues at the termini of model  $\alpha$ -helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family. *J. Mol. Biol.* **253**: 799–812.
- Kyte, J. 1995. *Structure in protein chemistry*. Garland, New York.
- Miranda, J.J. 2003. Position-dependent interactions between cysteine and the helix dipole. *Protein Sci.* **12**: 73–81.
- Muñoz, V. and Serrano, L. 1994. Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* **1**: 399–409.
- Nozaki, Y. and Tanford, C. 1967. Intrinsic dissociation constants of aspartyl and glutamyl carboxyl groups. *J. Biol. Chem.* **242**: 4731–4735.
- Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T.M., and Baldwin, R.L. 1990. Relative helix-forming tendencies of nonpolar amino acids. *Nature* **344**: 268–270.
- Penel, S., Hughes, E., and Doig, A.J. 1999. Side-chain structures in the first turn of the  $\alpha$ -helix. *J. Mol. Biol.* **287**: 127–143.
- Petukhov, M., Muñoz, V., Yumoto, N., Yoshikawa, S., and Serrano, L. 1998. Position dependence of non-polar amino acid intrinsic helical propensities. *J. Mol. Biol.* **278**: 279–289.
- Petukhov, M., Uegaki, K., Yumoto, N., Yoshikawa, S., and Serrano, L. 1999. Position dependence of amino acid intrinsic helical propensities II: Non-charged polar residues: Ser, Thr, Asn, and Gln. *Protein Sci.* **8**: 2144–2150.
- Rohl, C.A., Chakrabarty, A., and Baldwin, R.L. 1996. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5**: 2623–2637.
- Scholtz, J.M., Qian, H., Robbins, V.H., and Baldwin, R.L. 1993. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry* **32**: 9668–9676.
- Stapley, B.J. and Doig, A.J. 1997. Hydrogen bonding interactions between Glutamine and Asparagine in  $\alpha$ -helical peptides. *J. Mol. Biol.* **272**: 465–473.
- Sun, J.K., Penel, S., and Doig, A.J. 2000. Determination of  $\alpha$ -helix N1 energies by addition of N1, N2 and N3 preferences to helix-coil theory. *Protein Sci.* **9**: 750–754.
- Wan, W.Y. and Milner-White, E.J. 1999. A natural grouping of motifs with an aspartate or asparagine residue forming two hydrogen bonds to residues ahead in sequence: Their occurrence at  $\alpha$ -helical N termini and in other situations. *J. Mol. Biol.* **286**: 1633–1649.

## Pairwise Coupling in an Arg-Phe-Met Triplet Stabilizes $\alpha$ -Helical Peptide via Shared Rotamer Preferences

Teuku M. Iqbalsyah and Andrew J. Doig\*

Faculty of Life Sciences, Jackson's Mill, The University of Manchester, P.O. Box 88, Sackville Street, Manchester M60 1QD, U.K.

Received October 29, 2004; E-mail: andrew.doig@manchester.ac.uk

$\alpha$ -Helices are stabilized by the helix-forming tendencies of constituent amino acids,<sup>1</sup> capping interactions at the amino and carboxyl termini,<sup>2</sup> solvent environment,<sup>3</sup> and side chain interactions.<sup>4–8</sup> Interaction energies, in particular, have been measured for salt bridge,<sup>4</sup> aromatic–basic,<sup>5</sup> hydrogen bond,<sup>6</sup> and hydrophobic<sup>7,8</sup> interactions.

In proteins, side chains interact within a complex network of multiple noncovalent bonds, which may reinforce or weaken each other. The simplest system to investigate these effects is to study triplets of side chains. This has previously been studied as salt bridge interactions in  $\alpha$ -helical peptides. A triplet of charged RER residues spaced  $i, i + 4$  or  $i, i + 3$  stabilizes  $\alpha$ -helical peptides by more than the additive contribution of two single salt bridges.<sup>9</sup> Other triplets have also been studied, for example, EFR and EFE.<sup>6,10</sup> although they do not show significant effects on peptide stability.

Here, we investigate the coupling effect of pairwise interactions. The coupling is stabilizing when the free energy of the pairs present simultaneously is greater than the sum of the individual pairs. Conversely, when the free energy of the simultaneous pairs is less than the sum of the individual parts, the interaction is destabilizing.

We have previously measured the  $\Delta G$  of the interaction between Arg and Phe spaced  $i, i + 4$  in isolated  $\alpha$ -helix peptides. The  $\Delta G$  of the Arg-Phe interaction is  $-0.1$  kcal/mol.<sup>5b</sup> The  $\Delta G$  between Phe and Met interaction in an  $i, i + 4$  spacing has also been measured.<sup>8</sup> In this study, we look at the effects of these interactions on helix stability when present simultaneously via the shared Phe in  $i, i + 4, i + 8$  spacing. The  $i, i + 3$  interactions in a helix are generally weaker, so a coupling effect in  $i, i + 4, i + 7$  or  $i, i + 3, i + 7$  interactions would not be so clear.

Table 1 shows that in both the RF and FM pairs, Phe tends to be in the  $t$  conformation. The preference for the trans rotamer for Phe in helices and in all proteins are 67 and 34%, respectively.<sup>11</sup> This suggests that the two pairs incorporated in a RFM triplet could be stabilizing. The Phe residue in the middle of the triplet will be locked into the  $t$  conformation. RF and FM interactions will, therefore, not both have to pay the entropic cost of restricting the Phe to  $t$ . The stabilizing free energy resulting from the RFM triplet formation may, therefore, be greater than the sum of the individual RF and FM energies. The preference of Met  $\chi_1$   $g^+$  in the FM pair (0.81) is higher than that in the individual Met  $\chi_1$   $g^+$  in helices (0.69).<sup>11</sup> The high flexibility of the  $\chi_3$  rotamer (S–C bond) of Met<sup>12</sup> may also favor binding. This PDB-derived rotamer analysis, however, is based on the assumption that the rotamer distributions determined in the presence of tertiary contacts in proteins are applicable to peptides.

A series of alanine-based peptides was designed to allow the quantitative evaluation of the role of RFM triplets on helix stability (Table 2). Peptides were monomeric throughout the concentration range studied, as shown by the invariance of CD molar ellipticity with concentration.

**Table 1.**  $\chi_1$  Rotamer Populations of RF and FM  $i, i + 4$  Pairs in  $\alpha$ -Helices

residue	$\chi_1$ rotamer populations <sup>a</sup>			total found in $\alpha$ -helices
	$g^+$	$t$	$g^-$	
Arg in Arg-Phe	0.46 $\pm$ 0.06 <sup>b</sup>	0.44 $\pm$ 0.06	0.10 $\pm$ 0.04	71
Phe in Arg-Phe	0.34 $\pm$ 0.06	0.65 $\pm$ 0.06	0.01 $\pm$ 0.01	71
Phe in Phe-Met	0.16 $\pm$ 0.04	0.81 $\pm$ 0.05	0.03 $\pm$ 0.02	69
Met in Phe-Met	0.81 $\pm$ 0.05	0.19 $\pm$ 0.05	0	69

<sup>a</sup> Data from domain search of 1135 nonredundant proteins containing 4778 helices in ASTRAL database (SCOP 1.63 Sequence Resources). Rotamer definitions:  $-120^\circ < \chi_1 < 0^\circ = g^+$ ;  $0^\circ < \chi_1 < 120^\circ = g^-$ ;  $-120^\circ < \chi_1 < 240^\circ = t$ . <sup>b</sup> Errors were calculated as  $\sigma p_i = [p_i(1 - p_i)/N_i]^{1/2}$ , where  $p_i$  is the rotamer populations and  $N_i$  is the total number of pairs in helices.

**Table 2.** Peptides Used to Study Coupling in RFM Triplet

peptide sequence	predicted helicity <sup>a</sup>	predicted helicity <sup>b</sup>	exp helicity <sup>c</sup>
Ac-AAKRAAAFAAAAMKGY-NH <sub>2</sub>	30%	24%	24%
Ac-AAKARAAAFAAAMKGY-NH <sub>2</sub>	33%	27%	32%
Ac-AAKRAAAFAAAAMAKGY-NH <sub>2</sub>	47%	40%	36%
Ac-AAKARAAAFAAAMAKGY-NH <sub>2</sub>	50%	44%	63%

<sup>a</sup> Predicted using the Scint2 program. Values used for side chain  $i, i + 4$  interactions ( $p$  values) for RF and FM were 1.25<sup>b</sup> and 4.08<sup>a</sup>, respectively. Other values of  $n, v, w$ , and  $c$  were taken from ref 1b. <sup>b</sup> Helicity calculated as in 1, but with  $w$  values multiplied by 0.967. <sup>c</sup> Ellipticities  $\theta$  were measured with a Jasco J810 CD spectropolarimeter at 222 nm, 273 K in 5 mM Na/phosphate containing 10 mM NaCl, pH 7.4. CD data (in mdeg) were converted to  $[\theta]_{222} = \theta / (\text{molar concentration} \times 17 \text{ residues} \times 10)$  (in deg·cm<sup>2</sup>·dmol<sup>-1</sup>). Helix content was calculated as  $[\theta]_{222}(\text{observed}) / [\theta]_{222}(\text{max})$ .  $[\theta]_{222}(\text{max})$  is given by  $-40\,000 \times (1 - 2.5/N)$ , where  $N$  is the number of residues. Alternate equations or other proposed parameters give a small ( $\sim 1\%$ ) uncertainty in helical content.

A helicity prediction of the control R5F5M peptide using the  $w$  values (the helix interior preferences) of Rohl<sup>1b</sup> gives a predicted helicity about 6% higher than the experimental result. All  $w$  values of residues in the sequence were then corrected by a factor of 0.967. The discrepancy might be due to the fact that Scint2 does not take into account a possible destabilizing effect of Lys with the helix dipole at the N-terminus.<sup>1j</sup> Helicity disagreements in all peptides could also be due to changes in the ratio of  $\alpha$  to  $3_{10}$  conformations that might affect CD-derived helicity.

Helicity predictions of a series of homologous peptides with Ala replacing Phe show only a small difference for all sequences (50–53% helicity). Helix-coil theory takes account of the effect of moving the Arg and the Met when calculating the side chain interaction free energies.

In R5F4M and R4F5M, the theoretical and experimental helicities are somewhat different. The previous  $p$  value used to calculate the helicity of the FM interaction<sup>8a</sup> was obtained from the peptides: Ac-YGFAKAMA AAKAAAKAA-NH<sub>2</sub>, Ac-YGAAKAAFAKAMA AAKAA-NH<sub>2</sub>, and Ac-YGAAKAAAKAAFAKAM-NH<sub>2</sub>.

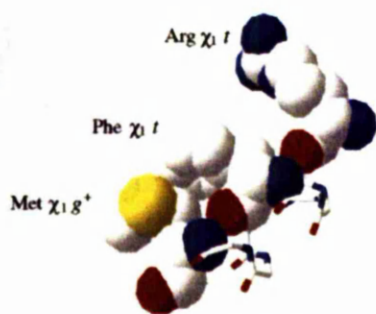


Figure 1. Example of a helical RFM triplet in 1lw3.

There may be additional  $i, i + 3$  side chain interactions between Met and Lys residues in these peptides. This would give an overestimate of the  $p$  value of the FM interaction.

A sequence of Ac-AAAKARAAAFARAAAFKAGY-NH<sub>2</sub> was used to calculate the previous  $p$  value for the RF interaction. The peptide contains two RF pairs. It is not clear why the value is lower than that found in this study.

We, therefore, refitted the  $p$  value for RF and FM interactions. The  $p$  values for each pair were fitted by varying them until the calculated helix contents agreed with experiment. Refitting  $p$  values for RF and FM gives values of 1.70 and 2.94, respectively, equivalent to  $-0.29$  and  $-0.59$  kcal/mol.

The predicted helicity for R4F4M is 44% using the refitted  $p$  values, which is much lower than the experimental result of 63%. This large difference proves qualitatively that there is a large stabilizing effect as a result of pairwise coupling in the R4F4M triplet. To get the predicted helicity to agree with experiment, the  $p$  value of each pair was then increased by a multiplication factor (mf) of 1.99. This increases the statistical weight of all helical residues from Arg to Met simultaneously by (mf)<sup>2</sup>. The  $p$  values for RF and FM in the R4F4M peptide become 3.38 and 5.85, respectively, equivalent to  $-0.96$  and  $-0.66$  kcal/mol.

$\Delta G$  for pairwise coupling was calculated as  $-RT \ln(1.99^2) = -0.75$  kcal/mol ( $-0.60$  to  $-0.91$  kcal/mol), reflecting a strong stabilizing effect in the R4F4M triplet. The energy is nearly as much as the additive energy of the individual pairs. The error in  $p$  was calculated by repeating the fitting procedure using experimental helicities increased or decreased by 3% to account for the experimental error in the measurement of helicity. Alternative parameter sets are available that would give different values, though the qualitative conclusions would not change.

We suggest that stabilizing pairwise coupling is a result of the R4F4M peptide only needing to pay the cost of restricting the Phe residue into a  $t$  conformation once in the triplet, rather than twice when the interactions are separate. The conformational entropy cost of restricting a residue into a helix is given by  $-R \sum p_i \ln p_i$ , where  $p_i$  values are the populations of the three  $\chi_1$  side chain rotamers. From our dataset, we found that these are 0.38, 0.60, and 0.02 for the  $g^+$ ,  $t$ , and  $g^-$  rotamers, respectively, for the Phe residue, giving  $-T\Delta S$  of 0.4 kcal/mol. Conformational entropy can thus account for over one-half of the free energy of pairwise coupling. The remainder ( $\sim 0.35$  kcal/mol) may be strain energy. This is the cost of giving Phe nonoptimal bond angles, dihedral angles, and bond lengths when forming its noncovalent interactions. Rotamer and  $\chi$  strain is of similar magnitude to conformational entropy in opposing protein folding.<sup>13</sup> Loss of conformational entropy in  $\chi_2$  of Phe may also contribute to the coupling free energy.

We searched crystal structures from the PDB for RFM triplets in helices, finding seven examples. Only structures obtained with the X-ray crystallography method of 2.5 Å resolution or better with

less than 50% homology were considered. The Phe residue is in the  $\chi_1 t$  conformation in all seven structures. Structural analysis of the RFM triplet indicates that the dominant interaction is hydrophobic for both pairs (Figure 1). Adoption of the Phe  $t$  rotamer in helices seems to be essential for forming the Arg and Met simultaneously, most likely through hydrophobic interactions with the face of the planar ring.

The shared rotamer preference effect that rationalizes our results is known to be a general property of proteins, giving cooperativity in folding and increasing protein stability. We have quantified this in a simple system. Any bond which restricts a side chain into a conformation that is favorable for forming additional interactions will show this effect. The effect is substantial, as fixing just a single side chain into its preferred conformation is here worth 0.75 kcal/mol.

**Acknowledgment.** The Michael Barber Mass Spectrometry Facility at UMIST is thanked for verifying peptide identity. T.M.I. thanks the TPSD Project of the Ministry of National Education of The Republic of Indonesia for the scholarship. We thank Neil Errington for helpful assistance in database search.

## References

- (1) (a) Wojcik, J.; Altmann, K. H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121–134. (b) Rohl, C. A.; Chakrabarty, A.; Baldwin, R. L. *Protein Sci.* **1996**, *5*, 2623–2637. (c) Pace, C. N.; Scholtz, J. M. *Biophys. J.* **1998**, *75*, 422–427. (d) Petukhov, M.; Munoz, V.; Yumoto, N.; Yoshikawa, S.; Serrano, L. *J. Mol. Biol.* **1998**, *278*, 279–289. (e) Petukhov, M.; Uegaki, K.; Yumoto, N.; Yoshikawa, S.; Serrano, L. *Protein Sci.* **1999**, *8*, 2144–2150. (f) Cochran, D. A. E.; Doig, A. J. *Protein Sci.* **2001**, *10*, 1305–1311. (g) Cochran, D. A. E.; Penel, S.; Doig, A. J. *Protein Sci.* **2001**, *10*, 463–470. (h) Petukhov, M.; Uegaki, K.; Yumoto, N.; Serrano, L. *Protein Sci.* **2002**, *11*, 766–777. (i) Ermolenko, D. N.; Richardson, J. M.; Makhatadze, G. I. *Protein Sci.* **2003**, *12*, 1169–1176. (j) Iqbalsyah, T. M.; Doig, A. J. *Protein Sci.* **2004**, *13*, 32–39.
- (2) (a) Dasgupta, S.; Bell, J. A. *Int. J. Pept. Protein Res.* **1993**, *41*, 499–511. (b) Forood, B.; Feliciano, E. J.; Nambiar, K. P. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 832–842. (c) Harper, E. T.; Rose, G. D. *Biochemistry* **1993**, *32*, 7605–7609. (d) Zhou, H. X.; Lyu, P. C. C.; Wemmer, D. E.; Kallenbach, N. R. *J. Am. Chem. Soc.* **1994**, *116*, 1139–1140. (e) Aurora, R.; Rose, G. D. *Protein Sci.* **1998**, *7*, 21–38. (f) Viguera, A. R.; Serrano, L. *Protein Sci.* **1999**, *8*, 1733–1742.
- (3) (a) Sonnichsen, F. D.; Van Eyk, J. E.; Hodges, R. S.; Sykes, B. D. *Biochemistry* **1992**, *31*, 8790–8798. (b) Waterhouse, V. D.; Johnson, W. C. *Biochemistry* **1994**, *33*, 2121–2128. (c) Luo, P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413–8421. (d) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13075–13079.
- (4) (a) Stellwagen, E.; Park, S.-H.; Shalongo, W.; Jain, A. *Biopolymers* **1992**, *32*, 1193–1200. (b) Huyghues-Despointes, B. M. P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 1965–1970. (c) Smith, J. S.; Scholtz, J. M. *Biochemistry* **1998**, *37*, 33–40. (d) Luo, R.; David, L.; Hung, H.; Devaney, J.; Gilson, M. K. *J. Phys. Chem. B* **1999**, *103*, 366–380. (e) Taylor, J. W. *Biopolymers* **2002**, *66*, 49–75.
- (5) (a) Fernandez-Recio, J.; Vazquez, A.; Civera, C.; Sevilla, P.; Sancho, J. *J. Mol. Biol.* **1997**, *267*, 184–197. (b) Andrew, C. D.; Bhattacharjee, S.; Kokkon, N.; Hirst, J. D.; Jones, G. R.; Doig, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 12706–12714. (c) Shi, Z.; Olson, C. A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2002**, *124*, 3284–3291. (d) Tsou, L. K.; Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2002**, *124*, 14917–14921.
- (6) (a) Marqusee, S.; Sauer, R. T. *Protein Sci.* **1994**, *3*, 2217–2225. (b) Huyghues-Despointes, B. M.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1995**, *34*, 13267–13271. (c) Stapley, B. J.; Doig, A. J. *J. Mol. Biol.* **1997**, *272*, 465–473. (d) Fernandez-Recio, J.; Romero, A.; Sancho, J. *J. Mol. Biol.* **1999**, *290*, 319–330. (e) Vijayakumar, M.; Qian, H.; Zhou, H. X. *Proteins* **1999**, *34*, 497–507. (f) Shi, Z. S.; Olson, C. A.; Bell, A. J.; Kallenbach, N. R. *Biophys. Chem.* **2002**, *101–102*, 267–279.
- (7) (a) Padmanabhan, S.; Baldwin, R. L. *J. Mol. Biol.* **1994**, *241*, 706–713. (b) Padmanabhan, S.; Jimenez, M. A.; Laurents, D. V.; Rico, M. *Biochemistry* **1998**, *37*, 17318–17330. (c) Andrew, C. D.; Penel, S.; Jones, G. R.; Doig, A. J. *Proteins* **2001**, *45*, 449–455.
- (8) (a) Stapley, B. J.; Rohl, C. A.; Doig, A. J. *Protein Sci.* **1995**, *4*, 2383–2391. (b) Viguera, A. R.; Serrano, L. *Biochemistry* **1995**, *34*, 8771–8779.
- (9) Olson, C. A.; Spek, E. J.; Shi, Z.; Vologodskii, A.; Kallenbach, N. R. *Proteins* **2001**, *44*, 123–132.
- (10) Shi, Z. S.; Olson, C. A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2002**, *124*, 3284–3291.
- (11) Penel, S.; Hughes, E.; Doig, A. J. *J. Mol. Biol.* **1999**, *287*, 127–143.
- (12) Gellman, S. H. *Biochemistry* **1991**, *30*, 6633–6636.
- (13) Penel, S.; Doig, A. J. *J. Mol. Biol.* **2001**, *305*, 961–968.

JA043446E



Articles

# Anticooperativity in a Glu–Lys–Glu Salt Bridge Triplet in an Isolated $\alpha$ -Helical Peptide<sup>†</sup>

Teuku M. Iqbalsyah and Andrew J. Doig\*

Faculty of Life Sciences, Jackson's Mill, The University of Manchester, P.O. Box 88, Sackville Street, Manchester M60 1QD, U.K.

Received May 11, 2005; Revised Manuscript Received June 16, 2005

**ABSTRACT:** Salt bridges between oppositely charged side chains are well-known to stabilize protein structure, though their contributions vary considerably. Here we study Glu–Lys and Lys–Glu salt bridges, formed when the residues are spaced  $i, i + 4$  surface of an isolated  $\alpha$ -helix in aqueous solution. Both are stabilizing by  $-0.60$  and  $-1.02$  kcal/mol, respectively, when the interacting residues are fully charged. When the side chains are spaced  $i, i + 4, i + 8$ , forming a Glu–Lys–Glu triplet, the second salt bridge provides no additional stabilization to the helix. We attribute this to the inability of the central Lys to form two salt bridges simultaneously. Analysis of these salt bridges in protein structures shows that the Lys–Glu interaction is dominant, with the side chains of the Glu–Lys pair far apart.

Proteins are stabilized most commonly by hydrogen bonds, van der Waals packing, and hydrophobic and electrostatic interactions. The amino acid side chains interact within a complex network of multiple noncovalent bonds, which may reinforce or weaken each other. Salt bridge interactions, in particular, are found in  $\sim 60\%$  of protein structures. One-third of the charged amino acids participating in salt bridge formation are found as complex salt bridges (with three or more amino acids), where triplet salt bridges are the most abundant. Approximately 20% of the total salt bridges found in proteins reside in  $\alpha$ -helix secondary structure in either an  $i, i + 3$  or an  $i, i + 4$  spacing (1).

The energetic role of salt bridges in protein structure stability has been found to vary. Solvent-exposed salt bridges contribute only marginally to protein stability (2–4). In contrast, a buried salt bridge stabilizes the native state by up to 5 kcal/mol (5, 6). Theoretical studies of solvation suggest that internal or external salt bridges contribute only marginally to protein stability, because of the high-energy penalty of desolvating ions within the interior or near the surface of a protein, thus diminishing the expected gain in strength of ionic interactions (7). The strength of salt bridges has been found to be determined by several factors: (a) the geometry and distance of the interactions (8), (b) the degree of exposure to solvent (9), and (c) the effect of neighboring residues (10). The frequency of salt bridges in proteins from

thermophilic proteins is high, suggesting an association with increased protein stability (11, 12).

Many attempts have been made to evaluate the strength of multiple salt bridges, particularly in proteins. A coupling interaction of an Asp8–Arg110–Asp12 triplet on the surface of barnase stabilizes the triplet by 0.8 kcal/mol relative to the sum of isolated pairs (4). A salt bridge and a hydrogen bond in an Asp14–Ser77–Arg17 triplet in the  $\lambda$  repressor are stabilizing by 1.5 and 0.8 kcal/mol, respectively (13). Kallenbach and co-workers reported that a GCN4 leucine zipper is stabilized by 1.7 kcal/mol upon introduction of an Arg–Glu–Arg triplet at the helix surface (2). The presence of many additional side chains close to the sites of interest, however, complicates the generalization of the energetic role of the salt bridge network in proteins. For this reason, we have studied a simple salt bridge network in a more isolated environment, i.e., in a short peptide.

The first evaluation of the strength of an engineered complex salt bridge in a peptide was reported by Mayne et al. (14) after studying a multiple-salt bridge involving Glu3, Asp4, and Arg7 in an 11-mer  $\alpha$ -helix. The triplet contributes substantially to the  $\alpha$ -helix stability with a  $\Delta G$  of  $-1.2$  kcal/mol. A triplet of charged Arg–Glu–Arg residues spaced  $i, i + 4, i + 8$  or  $i, i + 3, i + 6$  also stabilizes  $\alpha$ -helical peptides by  $-1.5$  or  $-1.0$  kcal/mol, respectively, which is more than the additive contribution of two single salt bridges (15). A similar stabilizing effect in an Arg–Phe–Met triplet in  $i, i + 4, i + 8$  spacing was also observed, although the triplet is not a complex salt bridge (16). Other non-salt bridge triplets in isolated helical peptides have also been reported, for example, Glu–Phe–Arg (17) and Glu–Phe–Glu (18), although they do not show significant effects on peptide stability.

<sup>†</sup> T.M.I. is supported by a TPSDP scholarship from the Ministry of National Education of the Government of Indonesia.

\* To whom correspondence should be addressed: Faculty of Life Sciences, Jackson's Mill, The University of Manchester, P.O. Box 88, Sackville Street, Manchester M60 1QD, U.K. Phone: +44-(161) 200 4224. Fax: +44(161) 236 0409. E-mail: andrew.doig@manchester.ac.uk.

Most triplet salt bridges studied so far were found to be cooperative; i.e., the free energy of the pairs present simultaneously is greater than the sum of the individual pairs. It is thus interesting to study a triplet salt bridge in an isolated  $\alpha$ -helix having the opposite effect. In this study, we examine Glu–Lys–Glu interactions when present simultaneously in an  $i, i + 4, i + 8$  spacing. The free energies of Glu–Lys and Lys–Glu interactions spaced  $i, i + 4$  in isolated  $\alpha$ -helical peptides have previously been found to be stabilizing by approximately  $-0.4$  kcal/mol in both orientations (19). In a Glu–Lys–Glu triplet, the central Lys side chain cannot point toward both Glu side chains simultaneously. The Glu–Lys–Glu triplet may therefore be less stabilizing than the sum of the two individual interactions, giving rise to anticooperativity. We have recently demonstrated that strong cooperativity in Arg–Phe–Met triplets can be attributed to a shared preference for the trans rotamer in the central Phe. The energetic cost of restricting the Phe residue is only paid once in the triplet, rather than twice when the interactions are separate (16). We consider whether rotamer preferences can also help explain the energetics of the Glu–Lys–Glu triplet.

## MATERIALS AND METHODS

**Peptide Synthesis.** Peptides were synthesized by the solid-phase method on an Applied Biosystems 433A peptide synthesizer on a 0.1 mmol scale using Fmoc (9-fluorenylmethoxycarbonyl) chemistry. C-Terminal amides were made using Rink amide resin (CN Biosciences). N-Termini were acetylated with a 4% (v/v) acetic anhydride/5% (v/v) pyridine dimethylformamide mixture for 30 min at room temperature. Peptides were cleaved from the resin using a mixture of 95% trifluoroacetic acid, 2.5% triisopropylsilane, and 2.5% water.

Peptides were purified by  $C_{18}$  reverse-phase HPLC, and their molecular weights were checked by MALDI mass spectrometry. Final purity was checked by analytical  $C_{18}$  HPLC. Peptide concentration was determined by measuring the tyrosine UV absorbance of aliquots of a stock solution dissolved in water using an  $\epsilon_{275}$  of  $1450 \text{ M}^{-1} \text{ cm}^{-1}$  (20).

**Circular Dichroism Measurements.** The secondary structure of each peptide was checked using circular dichroism between 190 and 250 nm. Equilibrium CD measurements for  $\alpha$ -helical structure were made at 222 nm with a Jasco J810 spectropolarimeter at 273 K. Ellipticity was measured in 10 mM NaCl and 5 mM sodium phosphate at pH 2.9 and 8.2. CD data in millidegrees were converted to mean residue ellipticity using the relation  $[\theta]_{222} = \theta/(\text{molar concentration} \times 18 \text{ residues} \times 10)$ , in degrees square centimeter per decimole. Helix content was calculated as  $[\theta]_{222}(\text{observed})/[\theta]_{222}(\text{max})$ .  $[\theta]_{222}(\text{max})$  is given by  $-40000(1 - 2.5/N)$ , where  $N$  is the number of amino acids in the peptide (21).

pH titrations were performed in a 3 mL 1.0 cm path length quartz cell at 273 K. The buffer used for pH titrations contained 10 mM NaCl, 1 mM sodium phosphate, 1 mM sodium borate, and 1 mM sodium citrate. The pH meter was calibrated for measurements at 273 K. The pH was adjusted during the titrations with aliquots of dilute HCl or NaOH. The volume of HCl or NaOH added was used for concentration correction.

Depending on the number of apparent titratable groups in the curve, the ellipticity data from the titrations were fitted

to a Henderson–Hasselbach equation. For one apparent  $pK_a$ ,  $[\theta]_{222}$  as a function of pH is given by

$$[\theta]_{222} = [\theta]_{222, \text{high-pH}} \left( 1 - \frac{1}{1 + 10^{\text{pH} - pK_a}} \right) + [\theta]_{222, \text{low-pH}} \left( \frac{1}{1 + 10^{\text{pH} - pK_a}} \right)$$

where  $[\theta]_{222, \text{high-pH}}$  and  $[\theta]_{222, \text{low-pH}}$  are the molar ellipticities measured at 222 nm at the titration end points at high and low pH, respectively.

The equation for two different  $pK_a$ 's is given by

$$[\theta]_{222} = [\theta]_{222, \text{mid-pH}} \left( 1 - \frac{1}{1 + 10^{\text{pH} - pK_{a1}}} \right) + [\theta]_{222, \text{low-pH}} \left( \frac{1}{1 + 10^{\text{pH} - pK_{a1}}} \right) + [\theta]_{222, (\text{mid-pH} - \text{high-pH})} \left( 1 - \frac{1}{1 + 10^{\text{pH} - pK_{a2}}} \right)$$

where  $pK_{a1}$  and  $pK_{a2}$  are the  $pK_a$ 's measured for the acid–base equilibrium at low and high pH, respectively.  $[\theta]_{222, \text{high-pH}}$ ,  $[\theta]_{222, \text{mid-pH}}$ , and  $[\theta]_{222, \text{low-pH}}$  are the molar ellipticities measured at 222 nm at the titration end points at high, intermediate, and low pH, respectively.  $[\theta]_{222, \text{mid-pH} - \text{high-pH}}$  is the change in molar ellipticity associated with  $pK_{a2}$ .

**Fitting of Helix–Coil Parameters for Side Chain–Side Chain Interactions ( $p$  values).** A simplified explanation of  $p$  values in helix–coil theory is as follows. Lifson–Roig-based models (22) describe the helix–coil transition as a two-state model on the level of individual amino acids, as either  $h$  (helix) or  $c$  (coil) according to its backbone dihedral angles ( $\psi$  and  $\phi$ ). Each residue in a helical sequence has a nonunity statistical weight based upon its own state and the states of the neighboring residues. The model now considers not only the original nucleating ( $v$ ) and propagating parameters ( $w$ ) but also N-capping ( $n$ ) and C-capping ( $c$ ),  $N1$  ( $n_1$ ),  $N2$  ( $n_2w$ ),  $N3$  ( $n_3w$ ), and side chain–side chain interactions (for review, see ref 23). When an  $i, i + 4$  interaction occurs between the residues at both ends of a helical  $hhhhh$  stretch, the residue in the middle of the quintet is mathematically assigned weight  $pw$ . The  $p$  value shows how the weight for forming the helical stretch is perturbed by the side chain interaction across the quintet. It can then be used to calculate the free energy of interaction as  $\Delta G_{(i,i+4)} = -RT \ln p$  (24). The  $p$  values for the  $i, i + 4$  side chain–side chain interactions were determined using a modified Lifson–Roig-based helix–coil theory, implemented in Scint2 (24). The inputs to the program are the peptide sequences and corresponding experimental helix contents, a library of  $w$ ,  $n$ -cap,  $c$ -cap, and  $v^2$  values for each amino acid, an  $n$ -cap value for acetyl and a  $c$ -cap value for amide (25), and a library of  $i, i + 4$  interactions. The  $p$  values for EK and KE  $i, i + 4$  interactions were calculated from the interaction free energies of Smith and Scholtz (19) as  $\Delta G = -RT \ln p$ . Values used for side chain  $i, i + 4$  interactions for  $E^+K^+$  and  $K^+E^+$  were 2.15 and 2.09, respectively, and for  $E^0K^+$  and  $K^+E^0$  were 1.50 and 1.45, respectively. When the  $p$  values needed refitting, parameters were varied until they converged on essentially unchanging values in agreement with experimental helicities.

Table 1: Peptides Used To Measure EKE Triplet Interactions

peptide	sequence	predicted helicity <sup>a</sup> (%)		experimental helicity (%)	
		parameter values of E <sup>-</sup> and K <sup>+</sup> <sup>b</sup>	parameter values of E <sup>o</sup> and K <sup>+</sup> <sup>c</sup>	pH 8.2	pH 2.9
E5K5E	Ac-AAAEAAAAKAAAAEAKGY-NH <sub>2</sub>	50	57	51	56
E5K4E	Ac-AAAEAAAAKAAAAEAKAGY-NH <sub>2</sub>	56	59	67	65
E4K5E	Ac-AAAEAAAAKAAAAEAKGY-NH <sub>2</sub>	59	61	63	63
E4K4E	Ac-AAAEAAAAKAAAAEAKAGY-NH <sub>2</sub>	64	63	60	61

<sup>a</sup> Predicted using Scint2. Parameter values of  $n$ ,  $v$ ,  $w$ , and  $c$  were taken from ref 25. <sup>b</sup>  $p$  values for E<sup>-</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>-</sup> were 2.15 and 2.09, respectively (19). <sup>c</sup>  $p$  values for E<sup>o</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>o</sup> were 1.50 and 1.45, respectively (19).

**Database Search.** A survey was conducted on a data set obtained from a set of crystal structures in the ASTRAL database (SCOP 1.63 Sequence Resources) (26), which is based on domains rather than entire structures. Each domain in the data set represents one superfamily and is the member of that superfamily with the best SPACI score. The SPACI score is a measure of the reliability and precision of a crystallographically determined structure in a PDB file. The data set contained 1135 chains with a total of 195 875 amino acid residues, and there was less than 20% sequence identity between any pair. These were processed using the DSSP algorithm to give secondary structure information as implemented in SSTRUC (27).

First, the data set was probed for helices of at least five helical residues (*hhhhh*) for pairwise  $i$ ,  $i + 4$  EK and KE interactions. EK and KE were categorized as pairs in coil when they, and the three intervening residues, were not all in helical conformations. A Glu-Lys-Glu helical triplet requires at least nine consecutive helical residues for  $i$ ,  $i + 4$ ,  $i + 8$  interactions to be present. The rotamer  $\chi_1$  and  $\chi_2$  preferences for Glu and Lys residues were then determined. Definitions of  $\chi_1$  and  $\chi_2$  rotamer angles for an  $sp^3$ - $sp^3$  bond are as follows: gauche<sup>+</sup> ( $g^+$ )  $\chi_1 = -60 \pm 60^\circ$ , trans ( $t$ )  $\chi_1 = 180 \pm 60^\circ$ , and gauche<sup>-</sup> ( $g^-$ )  $\chi_1 = 60 \pm 60^\circ$ .

The frequencies and errors were further calculated according to the following equations (28):

$$p_i = n_i/N$$

where  $n_i$  is the number of occurrences in  $N$  observations

$$\sigma p_i = [p_i(1 - p_i)/N]^{0.5}$$

where  $\sigma p_i$  is the standard deviation in  $p_i$ .

## RESULTS

**Peptide Design.** A series of intrinsically helical alanine-based peptides with different spacing of the interacting residues was designed to allow the quantitative evaluation of the role of Glu-Lys-Glu triplets in helix stability (Table 1). E5K5E was the control peptide in the EKE series. The Lys-Glu  $i$ ,  $i + 4$  interaction was achieved by swapping Ala at N5 with Glu at N4. Similarly, the Glu-Lys  $i$ ,  $i + 4$  interaction was achieved by changing AEAK to EAKA. An Ala residue was placed between Glu and Lys to avoid any close-range charge effects. The  $i$ ,  $i + 2$  spacing ensures the residues are on opposite faces of the helix. Shorter helices that could be made by moving Glu toward the N-terminus were prevented because Glu has unique preferences to be at N1 (29), N2 (30), or N3 (31). The helicity would be perturbed by Glu interacting with free amide NH groups at N1, N2,

and N3, as well as with the helix macro dipole at the N-terminus.

Gly is a helix breaker and was placed between Tyr and the rest of the sequence to prevent Tyr perturbation on the circular dichroism signal (32). Tyr was introduced at the C-terminus to allow the concentrations of the peptides to be determined. The N- and C-termini were blocked with an acetyl group and an amide group, respectively, minimizing the destabilizing interactions that may occur with the charged termini. This also creates an extra hydrogen bond at each terminus. Acetyl groups also substantially stabilize the helix (33).

All peptides were monomeric between 5 and 100  $\mu$ M as shown by invariance of CD molar ellipticity with concentration (data not shown). E4K4E, however, showed a change in secondary structure from entirely helical to a mixture of helix and  $\beta$ -sheet when left for a prolonged time (>48 h). This was observed by running analytical HPLC followed by MS analysis on the peptide after incubation for 48 h. HPLC analysis showed a tailing in the pure E4K4E peak, but both species gave identical masses when checked by MALDI. The CD spectra in range of 190–250 nm showed a conformational shift from  $\alpha$ -helix to  $\beta$ -sheet (data not shown). Data on the E4K4E peptide were thus collected solely on freshly dissolved pure lyophilized peptide. The reason behind the delayed aggregation could be that the E4K4E peptide contains residues assisting solubilization (two Glu and three Lys residues) distributed on only one face of the helix. Two, if not three, of the residues are also involved in side chain interactions.

**Refitting  $p$  Values.** The Glu-Lys pair has previously been studied for different spacings and orientations in isolated  $\alpha$ -helices (19, 34). The stabilizing effect of EK and KE did not vary significantly with orientation. We found dissimilar results, however, with Lys  $i$  and Glu  $i + 4$  more stabilizing than Glu  $i$  and Lys  $i + 4$ . The ellipticities of all EKE peptides were invariant and within error at three different salt concentrations, 0.01, 1.0, and 2.5 M NaCl (data not shown). This suggests that the interaction at high salt is not just a simple electrostatic interaction as the major component of the interaction between Glu and Lys cannot be screened by external salts. Hydrogen bonding thus provides the major part of the stabilizing effect in the EK and KE pairs, as proposed previously by Scholtz et al. (34).

The helicities of the EKE series peptides were measured at a range of pH values (Table 1). The terms E and B in ExKyE and BxKyB stand for Glu<sup>-</sup> and Glu<sup>o</sup>, respectively. The control peptides E5K5E and B5K5B show an excellent agreement between predicted and experimental helicity, indicating that errors in previous parameter values used in

Table 2: Energetics of Interaction in EKE Peptides

interaction	refitted $p$ value ( $\pm 3\%$ helicity error)	$\Delta G(i, i+4)$ (kcal/mol) <sup>a</sup>	multiplication factor (mf) ( $\pm 3\%$ helicity error)	refitted $p$ values $\times$ mf in triplet ( $\pm 3\%$ helicity error)	$\Delta G$ anticooperativity (kcal/mol) <sup>b</sup>
KE	6.6 (4.4–11.7)	–1.02 (–0.81 to –1.33)		2.6 (1.5–5.2)	
EK	3.0 (2.2–4.5)	–0.60 (–0.43 to –0.81)		1.2 (0.8–2.0)	
EKE		–1.62 <sup>c</sup> (–1.24 to –2.15)	0.39 (0.34–0.45)		1.02 (0.87–1.16)
KB <sup>d</sup>	2.5 (1.9–3.6)	–0.50 (–0.35 to –0.69)		1.6 (1.1–2.6)	
BK <sup>d</sup>	1.7 (1.3–2.4)	–0.29 (–0.13 to –0.47)		1.1 (0.7–1.8)	
BKB <sup>d</sup>		–0.79 <sup>c</sup> (–0.48 to –1.16)	0.65 (0.57–0.74)		0.47 (0.33–0.60)

<sup>a</sup> Calculated as  $-RT \ln p$  using refitted  $p$  values in column 2. <sup>b</sup> Calculated as  $-RT \ln(mf)^2$ . <sup>c</sup> Additive  $\Delta G$  of the individual pairs. <sup>d</sup> B represents Glu<sup>o</sup>.

the prediction for residues in the sequences are low and that the helix–coil theory that is used is accurate. The controls thus provide an independent system for measuring the preferred interactions. There are, however, disagreements between predicted and experimental helicities when  $i, i + 4$  interactions are present, particularly for the KE–KB orientation. The sequences of peptides previously used to obtain the  $p$  values for EK–BK and KE–KB interactions are Ac-AAQAAEAQAKAAQAAY-NH<sub>2</sub> and Ac-AAQAQAAQAE-AAQAAY-NH<sub>2</sub>, respectively (19). The Gln residues near the helix termini in these peptides may interact with Glu and Lys in an  $i, i + 3$  fashion that could give errors in the estimation of the interaction energies. Our peptides do not contain these  $i, i + 3$  interactions. In addition, Gln could form a significant interaction at the N3 position with the peptide backbone (31).

The original  $p$  values of the E<sup>–</sup>K<sup>+</sup> (2.15), K<sup>+</sup>E<sup>–</sup> (2.09), E<sup>o</sup>K<sup>+</sup> (1.50), and K<sup>+</sup>E<sup>o</sup> (1.45)  $i, i + 4$  interactions were thus refitted individually by varying them until the calculated helicity agreed with experiment. Refitting  $p$  values for E<sup>–</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>–</sup> gives values of 3.0 and 6.6, respectively, equivalent to –0.60 and –1.02 kcal/mol, respectively, giving an additive  $\Delta G$  of –1.62 kcal/mol. Refitting  $p$  values for E<sup>o</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>o</sup> gives values of 1.7 and 2.5, respectively, corresponding to –0.29 and –0.50 kcal/mol, respectively, giving an additive  $\Delta G$  of –0.79 kcal/mol (Table 2).

Using the refitted values, the predicted helicities are 74 and 69% for the corresponding E4K4E and B4K4B peptides, respectively. These predictions are significantly higher than the experimental helicities of 60 and 61% for E4K4E and B4K4B, respectively. These large differences prove qualitatively that there is a large destabilizing effect as a result of pairwise coupling in the E4K4E and B4K4B peptides. This is clear from the raw data, where there is remarkably a decrease in helix content in E4K4E compared to E5K4E and E4K5E, despite the introduction of an additional salt bridge.

For the predicted helicity to agree with experiment in the E4K4E–B4K4B peptides, the  $p$  value of each pair was decreased by multiplication factors (mf) of 0.39 and 0.65 for E4K4E and B4K4B, respectively (Figure 1). This decreases the statistical weight of all conformations within nine consecutive helical residues from Glu<sub>N-term</sub> to Glu<sub>C-term</sub> simultaneously by  $(mf)^2$ . The  $p$  values for E<sup>–</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>–</sup> in the E4K4E peptide become 1.2 and 2.6, respectively, equivalent to –0.09 and –0.51 kcal/mol, respectively. The  $p$  values for E<sup>o</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>o</sup> in the B4K4B peptide became 1.1 and 1.6, respectively, equivalent to –0.05 and –0.26 kcal/mol, respectively.

The free energy resulting from the EK and KE coupling was calculated as  $-RT \ln(mf)^2$ , giving  $\Delta G$  values of 1.0 kcal/

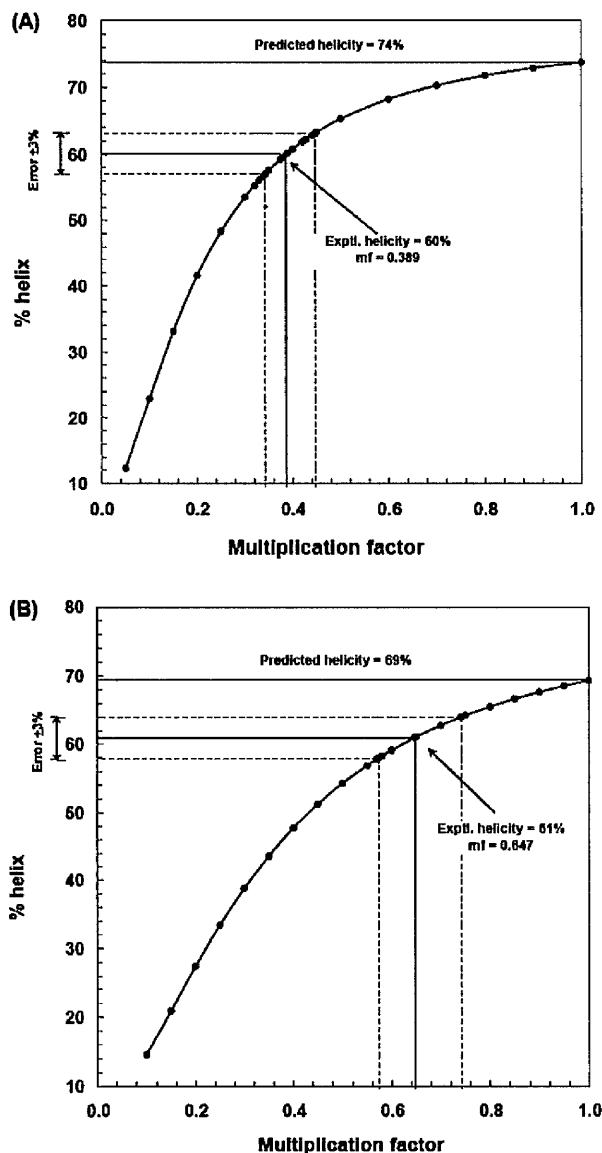


FIGURE 1: Adjustment of  $p$  values for pairs when present simultaneously in the triplet so that predicted helicity agrees with experiment: (A) E<sup>–</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>–</sup> pairs in the EKE triplet and (B) E<sup>o</sup>K<sup>+</sup> and K<sup>+</sup>E<sup>o</sup> pairs in the BKB triplet.

mol (0.9–1.2 kcal/mol) and 0.5 kcal/mol (0.3–0.6 kcal/mol) for E4K4E and B4K4B, respectively. These values show an anticooperativity effect in the triplet as they are positive, making the  $\Delta G$  in the triplets smaller than the sum of the energies of the individual interactions. The error in  $p$  was calculated by repeating the fitting procedure using experi-



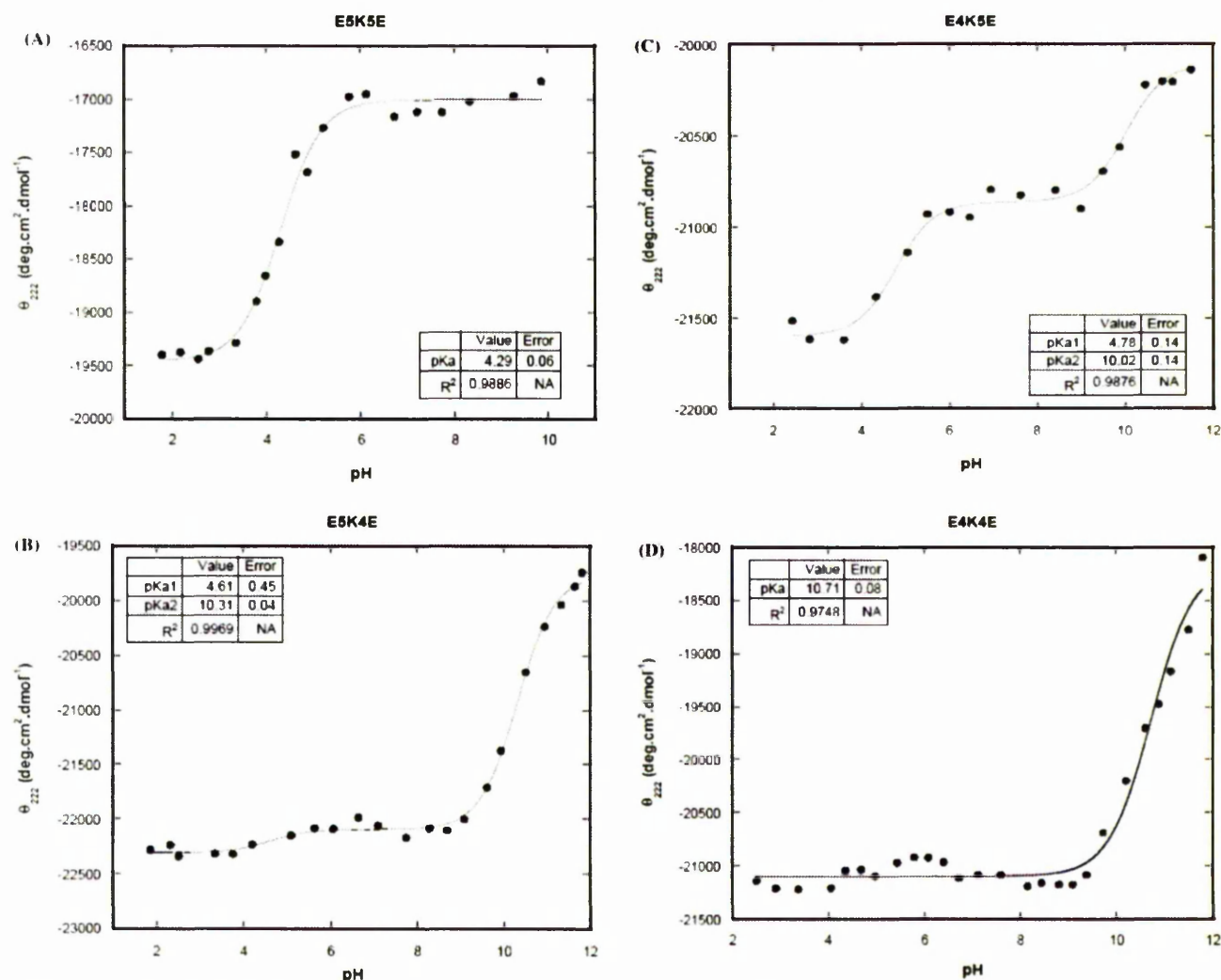


FIGURE 2: pH titration of EKE peptides: (A) E5K5E, (B) E4K5E, (C) E5K4E, and (D) E4K4E.

mental helicities increased or decreased by 3% to account for experimental error in the measurements of helicity.

**pH Titrations of EKE Peptides.** The helicities of the EKE peptides were followed as a function of pH using far-UV CD (Figure 2).  $pK_a$  values for the ionizable side chain groups were evaluated between approximately pH 2 and 11 by curve fitting the CD data to the Henderson–Hasselbach equations (see Materials and Methods). The upper range of accessible pH is given by Lys deprotonation, which leads to peptide aggregation.

The  $pK_a$  values of the two Glu residues in E5K5E cannot be separated. Only one observable transition across the pH range is present as the  $pK_a$  values of the two Glu residues are similar. The  $pK_a$  of  $4.29 \pm 0.06$  is comparable to the  $pK_a$  value of Glu of 4.5 in a model compound *N*-acetyl acid amide (35). The unperturbed  $pK_a$  value for Glu residues is reasonable, as the side chains are not involved in any interactions, yet the helix content is strongly influenced by the ionization state. Protonation of Glu residues increases the helix content of E5K5E by  $\sim 7\%$ . Previous work has found that helix propagation parameters of a single Glu in the helix interior give a difference in energy between protonated and unprotonated Glu of 0.15 kcal/mol (25, 34).

Only one Glu transition was observed in the E4K5E and E5K4E peptides. The  $pK_a$  values for E4K5E and E5K4E are  $4.78 \pm 0.14$  and  $4.61 \pm 0.45$ , respectively. It was expected that one of the Glu residues would have had a different  $pK_a$  value as it interacts with Lys. The difference in ellipticities upon protonation is very small, making it difficult to fit the data for calculation of the  $pK_a$  values, however. When the interaction is present, the destabilizing effect of having a charged Glu in the helix is weakened, which is exhibited by the low ellipticity difference between charged and uncharged Glu. The  $pK_a$  values in E4K5E and E5K4E surprisingly appear to be shifted to higher values than in E5K5E, though the experimental errors in the  $pK_a$  values make this conclusion uncertain. A decrease in Glu  $pK_a$  would be expected within a salt bridge, as the nearby positive charge would favor Glu<sup>−</sup>. The transition above pH 10 can be attributed to protonation of Tyr.

The titration curve of the E4K4E peptide may have a weak transition between pH 4 and 6, similar to that observed in Figure 2B, though this is difficult to identify as the putative change in  $\theta_{222}$  is very small. Repeating the titration three times did not convincingly show that this transition was real. Data fitting when using equations for two or three apparent  $pK_a$  values gave large errors of up to two pH units in the

Table 3:  $\chi_1$  Rotamer Population of Glu and Lys in  $\alpha$ -Helices

residue	rotamer distribution <sup>a</sup>			no. of observations
	$g^-$	$t$	$g^+$	
Glu	$0.04 \pm 0.003^b$	$0.39 \pm 0.01$	$0.57 \pm 0.01$	5947
Lys	$0.03 \pm 0.003$	$0.44 \pm 0.01$	$0.53 \pm 0.01$	4409
Glu in Glu–Lys pair <sup>c</sup>	$0.05 \pm 0.01$	$0.47 \pm 0.02$	$0.48 \pm 0.02$	468
Lys in Glu–Lys pair <sup>c</sup>	$0.02 \pm 0.01$	$0.37 \pm 0.02$	$0.61 \pm 0.02$	468
Lys in Lys–Glu pair <sup>c</sup>	$0.02 \pm 0.01$	$0.62 \pm 0.02$	$0.36 \pm 0.02$	411
Glu in Lys–Glu pair <sup>c</sup>	$0.03 \pm 0.01$	$0.27 \pm 0.02$	$0.70 \pm 0.02$	411

<sup>a</sup> Data were from a domain search of 1135 nonredundant proteins containing 4778 helices with a total of 65 472 residues in the ASTRAL database (SCOP 1.63 Sequence Resources). Rotamer definitions:  $g^+$  when  $-120^\circ < \chi_1 < 0^\circ$ ,  $g^-$  when  $0^\circ < \chi_1 < 120^\circ$ , and  $t$  when  $-120^\circ < \chi_1 < 240^\circ$ . <sup>b</sup> Errors were calculated as  $\sigma p_i = [p_i(1 - p_i)/N_i]^{1/2}$ , where  $p_i$  is the rotamer population and  $N_i$  is the total number of observations. <sup>c</sup> The pairs were observed in *hhhhh* conformations.

$pK_a$  values. The data were therefore fitted using the equation for one apparent  $pK_a$ . In general, the  $\theta_{222}$  does not significantly vary from pH 2 to 9, showing that the helix content of the peptide does not change when either Glu side chain is protonated. This is due to a decrease in helix content when Glu is deprotonated, as  $Glu^-$  has a weaker helix preference than  $Glu^0$ , which is counteracted by stronger salt bridges with  $Glu^-$  in the helix. These two effects are close to equal and opposite in sign, giving no net change in  $\theta_{222}$ .

**Rotamer Distributions of Glu and Lys Residues.** We have previously demonstrated that rotamer preference was helpful in explaining cooperativity in an RFM triplet in an  $\alpha$ -helical peptide (16). This triplet only needs to pay the cost of restricting the central Phe residue into a trans conformation once, rather than twice when the interactions are separate. We therefore examined the  $\chi_1$  rotamer preferences of Glu and Lys. This PDB-derived rotamer analysis is based on the assumption that the rotamer distributions determined in the presence of tertiary contacts in proteins can be applied to peptides. We only analyze the  $\chi_1$  rotamer as  $\chi_2$  is predominantly in the  $t$  rotamer in all cases.

**$\chi_1$  Rotamer Preference of Glu and Lys.** As summarized in Table 3, the Glu  $\chi_1$  population in an EK pair is spread almost equally between  $g^+$  (48%) and  $t$  (47%). The Lys  $\chi_1$  tends to be in  $g^+$  (61%). In the KE pair, the Lys  $\chi_1$  favors the  $t$  rotamer (61%). The Glu  $\chi_1$  is overwhelmingly in  $g^+$ . Individual Glu and Lys rotamer preferences in the helix were included for comparison.

**$\chi_1$  Rotamer Preference of Glu and Lys in the EKE Triplet.** We found 58 helices containing the EKE triplet spaced  $i, i + 4, i + 8$  in our data set (Table 4). This number is not large, making quantitative conclusions about rotamer preferences in the triplet difficult. Almost half the  $\chi_1$  rotamer combinations were  $g^+, t, g^+$  for the corresponding Glu, Lys, and Glu residues in the EKE triplet, while one-fifth were in a  $g^+, g^+, g^+$  combination. This suggests that the N-terminal Glu in the EK pair switches from equally  $\chi_1 g^+$  (48%) and  $t$  (47%) to mostly  $g^+$  when in the EKE triplet. The Lys  $\chi_1$  also switches from mostly  $g^+$  (61%) to  $t$  in the triplet. The Glu and Lys  $\chi_1$  preference in the KE pair seems to be retained in the triplet. The majority of Lys  $\chi_1$  in the EKE triplet adopt the  $t$  conformation, as for Lys in the KE pair. The preference of Glu  $\chi_1 g^+$  in the EKE triplet is also comparable to that of Glu  $\chi_1 g^+$  in the KE pair. Overall, this suggests that within the EKE triplet, the KE interaction is present, while the EK interaction is not. The central Lys cannot adopt the preferred rotamers for both salt bridges simultaneously.

Table 4:  $\chi_1$  Rotamer Distribution of the Interacting Residues in the Glu–Lys–Glu Triplet in  $i, i + 4, i + 8$  Spacing

$\chi_1$ rotamer combination <sup>a</sup>			total no. observed in the <i>hhhhhhhhh</i> conformation	%
Glu	Lys	Glu		
$g^+$	$g^-$	$g^+$	1	2
$g^-$	$g^+$	$g^-$	1	2
$g^+$	$g^+$	$g^+$	12	21
$t$	$g^+$	$t$	5	9
$g^+$	not defined	$g^+$	1	2
$g^-$	$t$	$g^-$	2	3
$g^+$	$t$	$g^+$	27	47
$t$	$t$	$t$	9	16
total			58	100

<sup>a</sup> Data were from a domain search of 1135 nonredundant proteins containing 4778 helices with a total of 65 472 residues in the ASTRAL database (SCOP 1.63 Sequence Resources). Rotamer definitions:  $g^+$  when  $-120^\circ < \chi_1 < 0^\circ$ ,  $g^-$  when  $0^\circ < \chi_1 < 120^\circ$ , and  $t$  when  $-120^\circ < \chi_1 < 240^\circ$ .

Our rotamer analysis contradicts the previous finding that the geometry of the interactions between acidic and basic residues is very similar in simple and complex salt bridges (7). The authors proposed that adding one residue to a simple interaction shows a minor change in geometry, in terms of atomic distance, partly because half of the salt bridge interaction is already constrained in a geometry that satisfies a complex salt bridge. The KE pair is indeed constrained, but the rotamer preferences in EK pairs change when in the triplet. Examples of triplet interactions are shown in Figure 3. Figure 3C shows a typical example of an EKE triplet, showing how the KE interaction dominates, confirming the rotamer analysis above, and that essentially only a single salt bridge is present.

## DISCUSSION

The salt bridges in the  $\alpha$ -helical peptides studied here are solvent-exposed and in isolation from interactions with neighboring residues. The EKE series peptides are a useful model for studying solvent-exposed salt bridges as they simplify the energetic role of the interacting pairs. The E4K5E peptide has a lower helix content than E5K4E, suggesting that the EK interaction is weaker than that of KE. This is confirmed by helix–coil theory analysis, which gives free energies of  $-1.02$  and  $-0.60$  kcal/mol for KE and EK, respectively, and by structural analysis of triplets. When the EKE triplet is present, KE dominates so the Lys predominantly adopts the rotamer that points it toward the  $i + 4$  Glu (Figure 3C). Previously reported  $\Delta G$  values show that the stabilizing free energies for EK and KE pairs are fairly

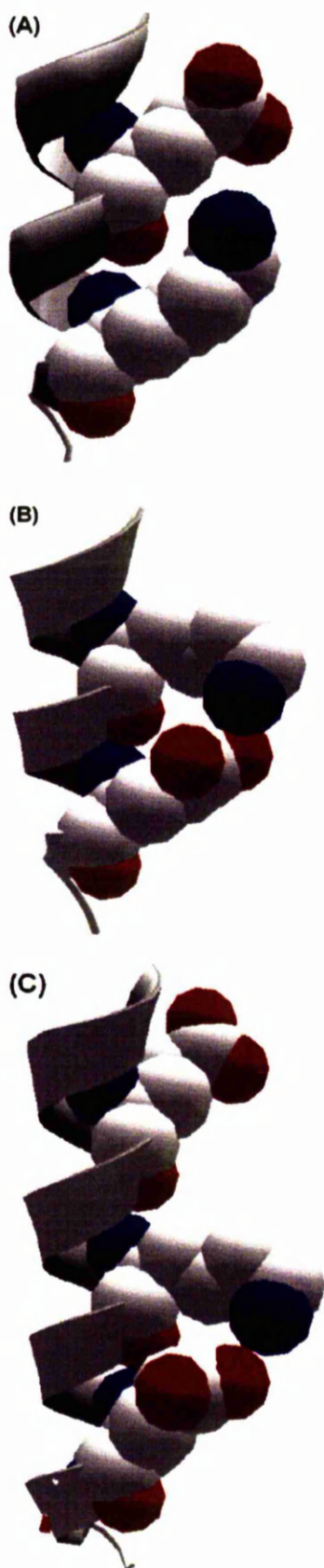


FIGURE 3: Examples of structure: (A) EK interaction between Glu62 and Lys66 in PDB entry 1A48, (B) KE interaction between Lys175 and Glu179 in PDB entry 1AOP, and (C) EKE interaction involving Glu110, Lys114, and Glu118 in PDB entry 1L2P.

similar (19, 34), contradicting the findings in this study. It appears that in the E4K4E peptide the Lys can interact with only one of the Glu residues. Thus, when we compare E4K4E to E5K4E, the addition of the second Glu-Lys salt bridge has little stabilizing effect. In fact, there is a slight decrease in stability that can be attributed to moving a Glu further from the N-terminus of the peptide, weakening any helix-dipole interactions. Glu is a strongly favored residue at the N1, N2, and N3 positions, which it can populate if the helix frays. The lower helicity in E4K4E compared to E4K5E is more difficult to explain. We suggest that when this change is made, the Lys moves to form a KE interaction, rather than the weaker EK in E4K5E. In addition, the Glu is moved further from the C-terminus, where it will form a destabilizing interaction with the helix dipole. The combination of these three effects, which are difficult to separate, results in a small decrease in helix content.

### CONCLUSIONS

Variations in the strength of salt bridges have been noted before and attributed to solvation effects. On the surface of a helical peptide, changes in solvation are minor, however. We have shown that the addition of a salt bridge can be worth essentially nothing. We attribute this to the inability of the central Lys to form two interactions simultaneously. Even in the simplest systems, noncovalent interactions are nonadditive and can be cooperative or anticooperative, depending on how compatible the conformational preferences of the interactions are. Rotamer analysis is useful in rationalizing these effects.

### ACKNOWLEDGMENT

The Michael Barber Mass Spectrometry Facility at The University of Manchester is thanked for verifying peptide identity. We thank Neil Errington for helpful assistance with database searches.

### REFERENCES

1. Musafia, B., Buchner, V., and Arad, D. (1995) Complex salt bridges in proteins: Statistical analysis of structure and function, *J. Mol. Biol.* **254**, 761–770.
2. Spek, E. J., Bui, A. H., Lu, M., and Kallenbach, N. R. (1998) Surface salt bridges stabilize the GCN4 leucine zipper, *Protein Sci.* **7**, 2431–2437.
3. Daopin, S., Sauer, U., Nicholson, H., and Matthews, B. W. (1991) Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis, *Biochemistry* **30**, 7142–7153.
4. Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) Strength and co-operativity of contributions of surface salt bridges to protein stability, *J. Mol. Biol.* **216**, 1031–1044.
5. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) pH-induced denaturation of proteins: A single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme, *Biochemistry* **29**, 2403–2408.
6. Tissot, A. C., Vuilleumier, S., and Fersht, A. R. (1996) Importance of two buried salt bridges in the stability and folding pathway of Barnase, *Biochemistry* **35**, 6786–6794.
7. Hendsch, Z. S., and Tidor, B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis, *Protein Sci.* **3**, 211–226.
8. Kumar, S., and Nussinov, R. (2002) Relationship between ion pair geometries and electrostatic strengths in proteins, *Biophys. J.* **83**, 1595–1612.
9. Takano, K., Tsuchimori, K., Yamagata, Y., and Yutani, K. (2000) Contribution of salt bridges near the surface of a protein to the conformational stability, *Biochemistry* **39**, 12375–12381.



10. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. F., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: Guidelines for protein engineering, *J. Mol. Biol.* 327, 1135–1148.
11. Elcock, A. H. (1998) The stability of salt bridges at high temperatures: Implications for hyperthermophilic proteins, *J. Mol. Biol.* 284, 489–502.
12. Kumar, S., Ma, B. Y., Tsai, C. J., and Nussinov, R. (2000) Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers, *Proteins* 38, 368–383.
13. Marqusee, S., and Sauer, R. T. (1994) Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in  $\lambda$  repressor, *Protein Sci.* 3, 2217–2225.
14. Mayne, L., Englander, S. W., Qiu, R., Yang, J. X., Gong, Y. X., Spek, E. J., and Kallenbach, N. R. (1998) Stabilizing effect of a multiple salt bridge in a pre-nucleated peptide, *J. Am. Chem. Soc.* 120, 10643–10645.
15. Olson, C. A., Spek, E. J., Shi, Z. S., Vologodskii, A., and Kallenbach, N. R. (2001) Cooperative helix stabilization by complex Arg-Glu salt bridges, *Proteins* 44, 123–132.
16. Iqbalsyah, T. M., and Doig, A. J. (2005) Pairwise coupling in an Arg-Phe-Met triplet stabilizes  $\alpha$ -helical peptide via shared rotamer preferences, *J. Am. Chem. Soc.* 127, 5002–5003.
17. Shi, Z. S., Olson, C. A., and Kallenbach, N. R. (2002) Cation- $\pi$  interaction in model  $\alpha$ -helical peptides, *J. Am. Chem. Soc.* 124, 3284–3291.
18. Shi, Z., Olson, C. A., Bell, A. J., and Kallenbach, N. R. (2002) Non-classical helix-stabilizing interactions: C-H $\cdots$ O H-bonding between Phe and Glu side chains in  $\alpha$ -helical peptides, *Biophys. Chem.* 101–102, 267–279.
19. Smith, J. S., and Scholtz, J. M. (1998) Energetics of polar side-chain interactions in helical peptides: Salt effects on ion pairs and hydrogen bonds, *Biochemistry* 37, 33–40.
20. Brandts, J. R., and Kaplan, K. J. (1973) Derivative spectroscopy applied to tyrosyl chromophores. Studies on ribonuclease, lima bean inhibitor, and pancreatic trypsin inhibitor, *Biochemistry* 12, 470–476.
21. Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) Determination of helix and  $\beta$ -form of proteins in aqueous solution by circular-dichroism, *Biochemistry* 13, 3350–3359.
22. Lifson, S. (1961) Theory of Helix-Coil Transition in Polypeptides, *J. Chem. Phys.* 34, 1963–1975.
23. Doig, A. J. (2002) Recent advances in helix-coil theory, *Biophys. Chem.* 101, 281–293.
24. Stapley, B. J., Rohl, C. A., and Doig, A. J. (1995) Addition of side-chain interactions to modified Lifson-Roig helix-coil theory: Application to energetics of phenylalanine-methionine interactions, *Protein Sci.* 4, 2383–2391.
25. Rohl, C. A., Chakrabarty, A., and Baldwin, R. L. (1996) Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol, *Protein Sci.* 5, 2623–2637.
26. Brenner, S. E., Koehl, P., and Levitt, R. (2000) The ASTRAL compendium for protein structure and sequence analysis, *Nucleic Acids Res.* 28, 254–256.
27. Smith, D. (1989) *SSTRUC: A program to calculate a secondary structural summary*, Department of Crystallography, Birkbeck College, University of London, London.
28. Penel, S., Hughes, E., and Doig, A. J. (1999) Side-chain structures in the first turn of the  $\alpha$ -helix, *J. Mol. Biol.* 287, 127–143.
29. Cochran, D. A. E., Penel, S., and Doig, A. J. (2001) Effect of the N1 residue on the stability of the  $\alpha$ -helix for all 20 amino acids, *Protein Sci.* 10, 463–470.
30. Cochran, D. A. E., and Doig, A. J. (2001) Effect of the N2 residue on the stability of the  $\alpha$ -helix for all 20 amino acids, *Protein Sci.* 10, 1305–1311.
31. Iqbalsyah, T. M., and Doig, A. J. (2004) Effect of the N3 residue on the stability of the  $\alpha$ -helix, *Protein Sci.* 13, 32–39.
32. Chakrabarty, A., Kortemme, T., Padmanabhan, S., and Baldwin, R. L. (1993) Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities, *Biochemistry* 32, 5560–5565.
33. Doig, A. J., Chakrabarty, A., Klingler, T. M., and Baldwin, R. L. (1994) Determination of free energies of N-capping in  $\alpha$ -helices by modification of the Lifson-Roig helix-coil theory to include N- and C-capping, *Biochemistry* 33, 3396–3403.
34. Scholtz, J. M., Qian, H., Robbins, V. H., and Baldwin, R. L. (1993) The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide, *Biochemistry* 32, 9668–9676.
35. Nozaki, Y., and Tanford, C. (1967) Intrinsic dissociation constants of aspartyl and glutamyl carboxyl groups, *J. Biol. Chem.* 242, 4731–4735.

BI0508690