

**PEPTIDE SYNTHESIS, PROTEIN FOLDING AND STABILITY**

Thesis submitted for the degree of  
Doctor of Philosophy  
at the University of Leicester

by

Christopher Cyril Bose BSc., ARCS.  
Department of Biochemistry  
University of Leicester

April 2000

UMI Number: U532822

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.



UMI U532822

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



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

## **Abstract**

### Peptide synthesis, protein folding and stability Christopher Cyril Bose

The aim was to evaluate the effects of backbone substituents on the stability of protein domains. The substituents used were a variety of peptidomimetics and uncommon amino acids.

The highly constrained substituents were introduced by total chemical synthesis using solid phase peptide synthesis techniques. Protein G B2 domain analogues containing 6-amino hexanoic acid, bithiazolidine dipeptide (BTD), aminoisobutyric acid, D-proline and N-methyl alanine were made using O-(7-Azahydroxybenzotriaz-1-yl)-1,1,3,3, -tetramethyl uronium hexafluorophosphate (HATU) as a coupling agent for hindered residues.

The stabilities, relative to wildtype, of nine synthetic analogues were analysed by denaturant unfolding, circular dichroism and differential scanning calorimetry. Binding to hA33 Fab was analysed using surface plasmon resonance methods. The wildtype sequence had an affinity for Fab of  $35\mu\text{M}$ . The mutant K15d-P T16N-MeA was  $2\text{kJmol}^{-1}$  more stable than wildtype and had an affinity for Fab of  $280\mu\text{M}$ . The mutant Q37Aib was just as stable as wildtype and had an affinity for Fab of  $250\mu\text{M}$ .

It is now possible to make analogues of protein domains with any amino acid residue, no matter how hindered, anywhere in the sequence. The introduction of restrained turn mimetics perturbs the folded state but does not destroy it. The introduction of uncommon side chains can make domains more stable than the wildtype structure and maintain the ability to bind antibodies to the wildtype. Structure/activity studies on small proteins can now be extended to use any uncommon amino acid in addition to the naturally occurring ones.

## **Acknowledgements**

I would like to thank my supervisor Professor Gordon Roberts for all his help and support. I am grateful to Dr. Lu-Yun Lian for help with the NMR results, Professor Nick Price for the circular dichroism studies, Professor Alan Cooper for the calorimetry results, Mr. Lloyd King for the electrospray mass spectra, and Dr. Lesley Chaplin and Dr. Shauna West for supervising the Biacore measurements. I would like to thank Dr. Denis Bulgin for proofreading. I am especially grateful to Mrs. Janeen Linsley, Ms. Eileen Holmes-Ievers and Mstr. Declan Bose for their patience, belief, help and support.

## Abbreviations

2-D	Two dimensional
3-D	Three dimensional
Acm	Acetamidomethyl
Ahx	6-Aminohexanoic acid
Aib	aminoisobutryl
AIDS	Acquired immunodeficiency syndrome
Aloc	Allyloxycarbonyl
Boc	Tertiary-butoxycarbonyl
BTD	bisthiazolidine dipeptide
BPTI	Basic pancreatic trypsin inhibitor
Cbz	Benzyloxycarbonyl
CD	Circular Dichroism
CZE	Capillary zone electrophoresis
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
Dde	N-1- (4,4-dimethyl-2, 6-dioxocyclohexylidene)
DIEA	Diisopropylethylamine
DIPCDI	Diisopropylcarbodiimide
DMAP	Dimethylaminopyridine
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DSC	Differential Scanning Calorimetry
DTT	Dithiothreitol
EDT	Ethanedithiol
EDTA	Ethylene diamine tetraacetic acid
ES-MS	Electrospray mass spectrometry
Fmoc	9-Fluorenylmethoxycarbonyl
GSH, GSSG	Reduced and oxidized glutathione, respectively

HATU	O- (7-Azahydroxybenzotriaz-1-yl)-1,1,3,3, - tetramethyl uronium hexafluorophosphate
HBTU	O- (Hydroxybenzotriaz-1-yl)-1,1,3,3, -tetra methyl uronium Hexafluorophosphate
HF	Hydrogen fluoride
HMP	para-hydroxymethylphenoxymethyl polystyrene resin
HOAc	Acetic acid
HOAt	7-Azahydroxybenzotriazole
HOBt	Hydroxybenzotriazole
HPLC	High performance liquid chromatography
MALD-TOF	Matrix assisted laser desorption time of flight mass spectroscopy.
MBHA	4-Methyl benzhydrylamine
Mtr	4-Methoxy-2, 3,6-trimethyl-benzene sulphonyl
NMP	N-Methylpyrrolidone
NMR	Nuclear magnetic resonance
NOe	Nuclear Overhauser effect
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuransulphonyl.
PDMS	Plasma desorption mass spectroscopy
Pmc	2,2,5,7,8,-pentamethylchroman-6-sulphonyl
PPI	Prolyl peptide isomerase.
PSA	Picric sulphonic acid
PTH	Phenyl thiohydantoin
PVDF	Poly vinylidene difluoride
SCX	Strong cation exchange
SPPS	Solid phase peptide synthesis
t-Bu	Tertiary butyl
tRNA	transfer ribonucleic acid

TFA	Trifluoroacetic acid.
TIPS	Triisopropylsilane
TMP	2,4,6-trimethylpyridine, collidine
TPPI	Time -proportional phase incrementation
TRIS	Tris (hydroxymethyl) aminoethane
Trt	Trityl
UV	Ultraviolet

### **Amino Acid Residues of Proteins**

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

Abstract.....	i
Acknowledgements .....	ii
Abbreviations.....	iii
Chapter 1: Introduction.....	1
1.1 What is Protein Stability?.....	1
1.2 The Unfolded State.....	3
1.3 Factors affecting protein stability .....	4
1.3.1 Water .....	4
1.3.2 Hydrophobic effects and conformational entropy .....	7
1.3.3 Hydrophobic packing .....	9
1.3.4 Hydrogen bonding .....	10
1.3.5 Helix dipoles.....	11
1.3.6 Disulphides .....	12
1.3.7 Ligand binding.....	13
1.4 Interactions. ....	14
1.5 Protein Folding .....	15
1.6 Modification of the covalent structure of proteins .....	16
1.6.1 Site directed mutagenesis .....	16
1.6.2 Chemically acylated tRNA.....	17
1.6.3 Native chemical ligation.....	18
1.6.4 Subtiligase .....	19
1.6.5 Solid Phase Peptide Synthesis .....	20
1.6.6 Active Esters.....	26
1.6.7 Non-natural residues in proteins.....	28
1.7 Thermodynamic Analysis.....	36
1.7.1 Circular Dichroism (CD).....	36
1.7.2 Differential Scanning Calorimetry (DSC).....	37
1.7.3 Denaturant Unfolding .....	38
1.8 Aim .....	39
Chapter 2: Trypstatin .....	41
2.1 Proteases and their Inhibitors .....	41
2.2 Bovine Pancreatic Trypsin Inhibitor (BPTI) .....	44
2.3 How HIV harms.....	46
2.4 Trypstatin.....	50
2.5 Results .....	52
2.5.1 Synthesis.....	52
2.5.2 Refolding .....	59
2.5.3 Conclusion .....	66
Chapter 3: Protein G .....	68
3.1 Protein G.....	68
3.2 Synthesis.....	75
Wild type B2 domain.....	75
Q37Aib .....	85
K15T16BTD.....	88
K15T16Ahx .....	96
G14K15BTD .....	98
Biotin 0, G14K15BTD .....	100

K15dP T16N-MeA .....	102
G14K15Ahx.....	104
G14K15BTD .....	106
15BTD16 .....	107
15Ahx16 .....	109
Synthesis Discussion .....	111
3.3 Guanidine unfolding.....	113
pH 7.6 .....	116
pH 4.7 .....	126
3.4 Circular Dichroism .....	138
3.5 Calorimetry .....	149
3.6 Binding Studies.....	158
Chapter 4: Discussion.....	169
4.1 Discussion.....	169
Wildtype protein G B2 domain .....	176
15Ahx16 and 15BTD16.....	178
D-Proline N-Me Alanine Analogue and Q37Aib .....	180
G14K15Ahx and G14K15BTD .....	181
K15T16Ahx and K15T16BTD.....	183
4.2 Conclusions .....	185
Chapter 5: Experimental.....	189
5.1 Materials and Methods .....	189
Reagents.....	189
Chain Assembly.....	191
General Deblock for Peptides attached to Acid Labile Handles .....	192
Matrix-assisted laser desorption .....	193
Electrospray .....	194
Circular Dichroism. ....	194
Guanidine denaturation.....	195
Differential Scanning Calorimetry .....	196
NMR .....	197
Binding Studies.....	197
5.2 Experimental.....	199
V3 Loop of gp120.....	199
Disulphide bridge formation.....	200
Trypstatin.....	202
Trypstatin-74 .....	206
Refolding Methods .....	209
Synthesis of Fmoc-L-BTD-OH .....	212
Wild-type Protein G B2 Domain .....	214
K15, T16 BTD B2 Domain .....	216
G14, K15 BTD B2 Domain.....	217
Biotin 0, G14, K15 BTD B2 Domain.....	217
Q37Aib B2 Domain.....	218
K15T16Ahx B2Domain .....	219
G14K15Ahx B2 Domain .....	219
K15dPT16N-MeA B2 Domain.....	220

G14K15BTD B2 Domain .....	220
15BTD16 B2 Domain.....	221
15Ahx16 B2 Domain.....	221
Appendices .....	222
Appendix 1 .....	222
Appendix 2. ....	227
References.....	230

## Chapter 1: Introduction

### 1.1 What is Protein Stability?

The three dimensional structure of a protein is very complicated; yet proteins rapidly and spontaneously fold to a unique structure. Why is it that their net stability is so small, of the order of 20 to 42 kJ mol<sup>-1</sup>?

By protein stability, we mean the net balance of energies that determine whether a protein will be in its native folded conformation, or a denatured state. This is to be distinguished from protein folding, which is the mechanism by which the native state is attained.

There are two kinds of protein stability, chemical and physical. Chemical stability [1] involves loss of structure due to bond cleavage. Examples include low pH hydrolysis of peptide bonds at aspartic acid residues; deamidation of asparagine and glutamine residues; oxidation of cysteine and methionine residues; disulphide interchange; and beta-elimination of cystine residues.

The physical or thermodynamic stability of a protein is the difference in Gibbs free

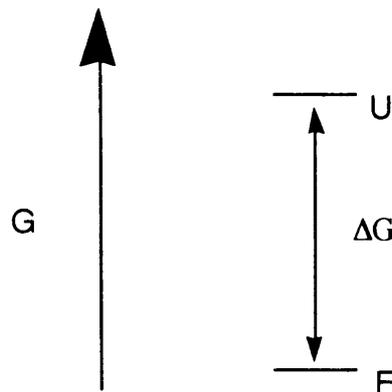


Figure 1. The free energy diagram for protein folding. U is the unfolded state, F the folded state.

energy,  $\Delta G$ , between the folded and the unfolded states as shown in Figure 1. Note that both the free energy of the unfolded state and the native state contribute to the overall energy. The larger and more negative  $\Delta G$ , the more stable the protein.

The free energy may be calculated from the following relationships:

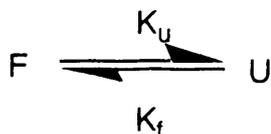
$$K_f = [F]/[U] = (1-x)/x \text{ where } x = \text{fraction unfolded, and } \Delta G = -RT \ln K_f$$

The Gibbs free energy is made up of two terms; enthalpy  $\Delta H$ , and entropy  $\Delta S$ , which are related by the equation:

$$\Delta G = \Delta H - T\Delta S$$

where  $T$  is the temperature in Kelvin.

A discussion of physical or thermodynamic stability of a protein that unfolds and refolds is usually represented by an equation such as:



Where  $F$  is the native protein,  $U$  the denatured state and  $K_u$  is the equilibrium constant for unfolding. As written, it tends to imply that there are single states on both sides. It is important to realise and state explicitly that both the folded and unfolded states are dynamic.  $F$  and  $U$  are heterogeneous states. Proteins are always moving. All the molecules in the same protein have different conformations from all the others. No protein molecule ever exists in the same conformation twice. The folded state must now be considered to be compact, having an average

conformational homogeneity with recognised regions of secondary structure. It has a well-defined hydrogen bond pairing set, many of which are internal. The unfolded state is an ill defined set of states whose population of conformations will include open, irregular, flexible, highly dynamic structures.

The study of protein stability is important because it leads to a basic understanding of the thermodynamics of the folding process. For practical reasons, stability is important in the use of protein drug therapies. An understanding of protein stability may lead to a better understanding of protein function and then to better therapies. In biotechnology, the question of enzyme stability in the field of biocatalysis is economically very important. The stability of proteins in the food processing industry may be of multi-billion pound value.

## **1.2 The Unfolded State**

In the previous section it was noted that the overall stability of a protein depends on both the folded and the unfolded state. Most studies have assumed that the unfolded state of a protein is a random coil, with no residual structure. This could safely be treated as a single state in the thermodynamic description. However, this simple view is now contested [2]. The second state may often not be the unfolded state. Denatured states may have substantial structure, especially secondary structure. Shortle has reviewed the evidence [3]. NMR studies have shown that interactions do occur between residues in the unfolded state and that these interactions are not mirrored in the folded state [4]. Sauer has also found that substitutions of polar residues at a surface position of the lambda repressor led to decreased stability [5]. If the denatured

state involves most residues in fully extended peptide chains i.e. maximal solvent exposure, then substitutions involving solvent exposed residues in the native state will have only small effects. The observed decreased stability implies some, possibly transient structure in the denatured state. Shortle made a large number of staphylococcal nuclease mutations and examined their stability to guanidine [6]. They fell into two categories: those that increased the value of  $m$ , the dependence of  $\Delta G$  on denaturant concentration relative to wild type, and those that caused a decrease in  $m$ . The interpretation was that there are two classes of the denatured protein, those with more or those with less structure. An increase in  $m$  means less structure in the denatured states and *vice versa*. Since the increase in  $m$  suggests an increase in solvent accessible area in the denatured state, the conclusion is that the mutations cause changes in the structure of the denatured state, which implies that it is not fully unfolded.

### **1.3 Factors affecting protein stability**

#### **1.3.1 Water**

Water is important to protein structure and function and is also important to protein stability. All of our structural knowledge about proteins is derived in the context of water as solvent. Proteins are crystallized from water and NMR structures are determined in aqueous media. Enzyme catalysis often depends on water participation. Proteins fold in water, and some proteins bind to it very firmly. Others are loosely bound to water with a structure oriented like ice. Water molecules may also form bridges between the amide groups of adjacent peptide chains, resulting in structures

similar to those of the pleated sheet but with a water molecule in the position of the hydrogen bonds of that configuration.

The three-dimensional structure of the water molecule can be pictured as a tetrahedron with a central oxygen atom and four vectors of high electron probability, as shown in Figure 2.

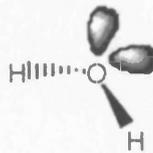


Figure 2. Representation of water molecule showing lone pair orbitals.

The two “legs” in which the hydrogen nuclei are present are called bonding orbitals. All the properties of water are related to the dipole resulting from the greater electronegativity of the oxygen atom compared to the two hydrogen atoms. Opposite the bonding orbitals and directed to the opposite corners of the tetrahedron are two “legs” of negative electrical charge. Known as the lone-pair orbitals, these and the dipole effect are the keys to water’s peculiar behaviour, since they attract the hydrogen nuclei of adjacent water molecules to form weak bonds known as hydrogen bonds. Although these bonds are not especially strong, they orient the water molecules into a specific configuration, and significantly affect the properties of water in its solid and liquid states. They account for the high latent heat of fusion, the high heat capacity, the anomalous freezing point, boiling point, surface tension and viscosity.

These bulk properties result mainly from the local order in water from an extended and dynamic hydrogen-bonded network known as water clusters. Water clusters have been extensively studied both experimentally and theoretically in recent years. At the moment, it is believed that bulk water is composed of mixtures of cyclic trimer to octamer structures. Examples of cyclic hexamers are shown in Figure 3.

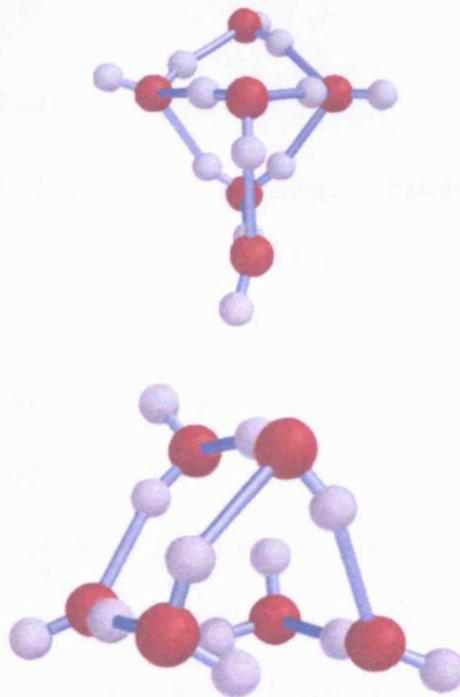


Figure 3. Cyclic water clusters. Oxygen atoms in red; hydrogen atoms in grey; hydrogen bonds in blue.

It is possible to construct a model of water involving isomerization between clusters of polycyclic cubic octamers that can dissociate into two cyclic tetramers. This model reproduces the anomalous heat capacity of water from 273K to 373K.

Hydrogen bonds can occur in a variety of molecules and vary in strength according to the electronegativities of acceptor and donor atoms. Strengths are generally in the range of 8 to 42 kJ mol<sup>-1</sup> at room temperature, but they appear to vary in water alone. Inelastic neutron scattering of ice has shown that two kinds of hydrogen bond exist in the solid phase. A model was proposed in which strong and weak hydrogen bonds in the ratio of 2:1 are randomly distributed through the network. The implications are that a similar bimodal system could exist in the liquid phase and this could explain some of the properties of liquid water.

### 1.3.2 Hydrophobic effects and conformational entropy

Hydrophobic residues are those whose free energy of transfer of their side chains from water to a nonpolar liquid is negative. In aqueous media they are surrounded by a lattice arrangement of water molecules in a cage. One of the simplest models for such a cage is a dodecahedron which is continually being broken down and reformed around the hydrophobic side chain. Each water molecule is positioned to completely satisfy its hydrogen bonding requirements with no polar groups needing to be inside the cage. This has two consequences. Firstly, the ordering of water into cages is entropically costly, giving a large  $T\Delta S$  term. Secondly, the local hydrogen bonded network is stronger than the bulk solvent, because of the ordering of the water molecules into the cage structure, and there is a higher enthalpy of interaction as indicated by the experimental observation that hydrating non-polar groups is an exothermic reaction. Since  $\Delta G = \Delta H - T\Delta S$  and hydrophobic hydration is not a spontaneous reaction, the entropy term must be large and negative. Hydration of non-

polar groups is entropically very costly. If two non-polar groups and their associated water cages are put into Van der Waals contact, two things happen. First, the size of the cage needed to surround both groups together is reduced, so the entropic cost is significantly reduced. Second, energy is gained from the attraction between the groups.

These attractive forces are known as Van der Waals forces, They arise from transient fluctuations in electron density distribution caused by quantum mechanical variations. A transient electric dipole will polarize a similar but opposite effect in an adjacent group so that the two groups attract. The dipole-dipole interaction is short range and is usually quantified by the relation  $F = -C/R^6$ , where  $C$  is a constant.

The hydrophobic groups are driven to form the core in protein structures because it is energetically favourable from an entropic and an enthalpic point of view, with entropy having the greater effect. This is the basis of the transition from the unfolded to the folded state. These terms are opposed by the unfavourable entropy of fixing the molecular structure by restricting the main chain and side chain atoms. However when folding occurs and the water cages are destroyed, the entropy of the bulk solvent increases. This increase in entropy of solvent plus protein offsets the decrease in entropy of the protein alone. This hydrophobic effect is one of the major contributory factors to protein stability. Site directed mutagenesis studies have shown how much of a contribution this makes. Proteins such as T4 lysozyme [7], barnase [8], and staphylococcal nuclease [9] have had one hydrophobic residue replaced with another and the change in free energy  $\Delta\Delta G$  measured. One of the most common approaches is to replace Ile by Val, removing a methylene group.  $\Delta\Delta G$  values for 20 such changes in six proteins show an average contribution to stabilization of 5.5 kJ

mol<sup>-1</sup>. Energy released from packing of the hydrophobic residues is very large. Holes in the internal structure are very destabilizing because there are no Van der Waals contacts and if the space is filled with water molecules instead, a very unstable situation is created. Sauer's group investigated the effects of 15 single and multiple mutations in the core of the N-terminal domain of lambda repressor [10]. Seven mutants with steric changes but no significant volume changes gave  $\Delta\Delta G$  values of 4 kJmol<sup>-1</sup>. Matthews has shown that small hydrophobic compounds will fill cavities of the correct size in proteins and help stabilize the molecule. The L99A / F153A mutant of T4 lysozyme was stabilized by about 1 kJ mol<sup>-1</sup> [11].

### 1.3.3 Hydrophobic packing

When beta sheets form, two chains must be positioned exactly or the sheet will not be stable. Movement by one residue means that the side chains will be on the wrong side of the sheet and in steric conflict with the other strand. It has become clear from Hecht's work [117] that the structures of the turns leading into beta sheets are very important for protein stability. The centres of beta sheets, where H-bonding is most precise, have lower packing densities of the side chains than at the ends of the sheet where the H-bonding lattice is breaking down. Turns leading to helices are not so important [12]. Using an enzyme as a typical example, the overall structure that maximizes protein stability is thought to consist of shells of high and lower density through the molecule [13]. Density is not constant through a protein structure. At the active site itself, the density is low where the local geometry is determined by H-bonding to achieve a lattice effect. Here the orientation of groups is paramount and so

forces a less dense structure. The density is highest in the structure around the active site [14]. Here, hydrophobic effects favour the dense packing. Matthews' work using heat stable mutants of T4 lysozyme showed that only a few residues in the protein are very important for overall stability [7]. Those residues are the ones that show minimal mobility in the crystal structure. This implies that the residues in the most densely packed portion of the molecule contribute most of the stabilization energy. Insertions and deletions in non-critical areas are tolerated well. The protein simply adjusts to tolerate the changed segment.

#### 1.3.4 Hydrogen bonding

Hydrogen bonding is considered to have as much an effect on protein stability as the hydrophobic effect. In a study of RNase T1 mutants in which one to three intramolecular hydrogen bonds were broken, Pace and co-workers estimated that each hydrogen bond contributes  $5.5 \text{ kJmol}^{-1}$  to the protein's stability, after correction for the hydrophobic differences. This was done by multiplying the difference in hydrophobicity between the residues in the wild type and mutant, by the figure for the fraction of side chains that are buried in the wild-type protein. Since 86 intramolecular hydrogen bonds are formed in the folding of RNase T1, the contribution to the stability is close to that made from hydrophobic effects. Similar results were obtained in preparing mutants of barnase and staphylococcal nuclease. There is the potential for Asn to Ala mutants to leave cavities large enough to hold three water molecules. There is then a chance that a hydrogen bond will be formed to one of the water molecules. Tyr to Phe mutants leave a smaller cavity, and Tyr

residues are more likely to be buried. It is more likely that Tyr to Phe mutants will be left with an unpaired hydrogen bond. An analysis of 10 Tyr to Phe mutants similarly gives 1.6 kcal/mol per hydrogen bond. In another study, Sauer used double-mutant cycles in lambda repressor to conclude that the hydrogen bonds between N14 to R17 and D14 to S77 are stabilizing by 3.4 and 6.3 kJ mol<sup>-1</sup> respectively.

BPTI contains four internal water molecules. Berndt prepared the mutant G36S-BPTI in which one of the water molecules is replaced by serine [15]. The mutant was less stable than wild type by 3 kJ mol<sup>-1</sup>.

It is becoming clear that side chain hydrogen bonding plays a role in the stability of  $\alpha$ -helices. Huyghues-Despointes [143] has shown that a hydrogen bond between Gln and Asp side chains spaced four residues apart in an  $\alpha$ -helix contribute 4 kJ mol<sup>-1</sup> to helix stability.

### 1.3.5 Helix dipoles

Since the  $\alpha$ -helix is the most commonly occurring secondary structure in proteins, the stability of the helix is part of the overall stability of the protein. All the hydrogen bonds and peptide groups point in the same direction in an  $\alpha$ -helix. The vector sum of the individual dipoles from the peptide bonds aligned along the axis of the helix is the dipole moment. Since oxygen is more electronegative than nitrogen, a helix is negatively charged at the C-terminal and positively charged at the N-terminal. The dipole moment is equivalent to a bare charge of -1/2 an electron at the C-terminal and +1/2 at the N-terminal end [134]. Placing a positively charged residue near the C-terminal and a negatively charged one near the N-terminal stabilizes the polarity,

helix and finally the protein. Fersht's group first demonstrated this effect. In a study of barnase, [17] he replaced a single histidine residue at either the C-terminal or the N-terminal of each helix. The stabilities of the mutants were measured by urea denaturation. The C-terminal histidine mutants were about 2.5 kJ mol<sup>-1</sup> more stable when the histidine was protonated, while the N-terminal mutants were about 4.6 kJ mol<sup>-1</sup> less stable with the histidine protonated. In addition, Matthews [142] has shown that the thermal stability of T4 lysozyme can be increased by introducing amino acid substitutions designed to interact with helix dipoles. Robinson and Sligar [135] changed E18 at the carboxyl terminal of cytochrome b562, and Q25 at the N-terminal to make E18R, E18K, E18L, Q25E and Q25K mutants. The  $\Delta\Delta G$  were of the order of 2 kJ mol<sup>-1</sup>.

Conversely, reinforcing the charge differences destabilizes the protein. Walter [16] has shown the size of the effect to be 4 kJ mol<sup>-1</sup> in a study of Ribonuclease T1 mutants.

### 1.3.6 Disulphides

If the disulphides of some proteins are broken then capped off chemically, for example by carboxymethylation with iodoacetamide, to avoid introducing a charge, the resulting protein is denatured. Disulphides stabilize structures by constraining the unfolded conformations of the protein. They do this by decreasing the conformational entropy of the unfolded state. Pace has proposed a new relation for predicting the effect of a crosslink on the conformational entropy ( $\Delta S_{\text{conf}}$ ) of a protein:

$$\Delta S_{\text{conf}} = -2.1 - 1.5R \ln(n)$$

where  $n$  is the number of residues in the loop forming the disulphide bond. The relation predicts that by adding loops of 15 and 45 residues the conformational stability of a protein will be increased by  $13 \text{ kJ mol}^{-1}$  and  $17 \text{ kJ mol}^{-1}$  respectively. Attempts to increase the stability of proteins by introducing disulphides have not met with much success. It has proven difficult to stabilize subtilisin BPN' by the introduction of disulphide bonds. Introducing the cysteines within precise distances that must be met for disulphide bond formation and accommodating the bond angles around sulphur atoms leads to strain in the molecule, and formation of the disulphide bond does not offset this. However, in the case of a barnase mutant, Fersht [18] has shown that a disulphide-bridged mutant is more stable than wild type by  $17 \text{ kJ mol}^{-1}$ . The primary aim was not to stabilize the structure but to slow down the denaturation. The disulphide was introduced to link two loops known to pack together early in the folding pathway. Assuming complete reversibility in the folding reaction, this implies that the unfolded state would have the two loops still packed together relative to the unfolded state of the wild type. The mutant unfolds 20 times slower than wildtype, and this is the source of the stability of the mutant.

### 1.3.7 Ligand binding

A general consequence of ligand binding e.g. inhibitors to enzymes, is that the enzyme is stabilized against unfolding and is less flexible. If the ligand is also a protein, it can be assumed that binding also stabilizes it.  $\Delta G (\text{H}_2\text{O})$  of RNase T1 is almost doubled by adding 0.2M sodium hydrogen phosphate. Other salts also work. It

has been shown that this results from the binding of cations at a cation binding site and the binding of one phosphate to an anion-binding site.

#### **1.4 Interactions.**

The factors that affect a protein's stability are not well understood. There is a tendency in the literature to attempt to separate out different contributions under different descriptions. For example, internal hydrogen bonds as opposed to all others; salt bridges; aromatic interactions as opposed to Van der Waal's interactions. This approach of putting different interactions into separate compartments is understandable because it makes thinking about the problem easier. However it may be flawed because the various interactions are really different manifestations of the same energy phenomena. Hydrophobic interactions are a manifestation of the hydrogen bonding properties of water. It is best visualised as a repulsive interaction between non-polar groups and water rather than any sort of attractive force. The same hydrogen bonds that are responsible for the solvation of polar groups and charged side chains are responsible for the repulsive forces. Hydrogen bonds themselves, are just convenient fictions. The thermodynamic contribution of hydrogen bonds to protein stability is very unclear. According to the literature, the strength of a hydrogen bond is typically between 8 and 30 kJmol<sup>-1</sup>. This is very broad for a defined interaction, and is so because liquid water is a very good hydrogen bonding solvent. What is real is the broad background of quantum mechanical fluctuations that are the Van der Waal's forces. These arise from the rapidly shifting dipoles in the electron clouds of molecules that are caused by the effects of Heisenberg's uncertainty principle. A dipole is two charges separated by a certain distance. The uncertainty

principle states that it is impossible to know the exact position and momentum of a particle such as an electron. The result is that the electron cloud in a molecule is smeared in space and dipoles are constantly forming and reforming. Permanent dipoles and charges give rise to more specific and longer range electrostatic interactions. These may be attractive or repulsive. A 'salt bridge' is an example of a close, direct electrostatic energy. Permanent dipoles may also induce dipoles and polarize surrounding groups that may then attract others.

The complete description of the electrostatics of the protein must also include the solvent water and salts in solution. A thermodynamic description of such interactions is complicated and must include both enthalpy and entropy terms. However, there is little or no correlation between the enthalpy and free energy changes for protein processes. Analyses of protein energetics which lead only to data concerning free energies may only be useful in protein engineering, to improve the thermal stability of a peptide or protein, for example [144].

## **1.5 Protein Folding**

If protein stability is the quantification of the difference between the folded and the unfolded state, then protein folding describes how the protein adopts the native state. Protein folding is not well understood. In 1961, Anfinsen reported that RNaseA could be fully denatured by reducing the disulphide bonds and dissolving in 8M urea, and then fully renatured on diluting the urea and reoxidizing the cysteines. Full catalytic activity was recovered. This experiment is usually taken to mean that all the information necessary to determine the three-dimensional structure is contained in the

amino acid structure. This is incorrect. All the information needed is present in the total system i.e. the protein chain and the solvent. In 1968, Levinthal argued in his “paradox” that there are too many possible conformations for proteins to find the one native structure by random searching of the conformational space. He therefore concluded that there are specific pathways for folding. This concept led to the search for folding intermediates, whether these were on pathway or off pathway. The idea was that if intermediates could be observed, we would learn how nature finds the path to the one native state. These led in turn to protein folding kinetics experiments. These included probing for intermediates in folding which are detectable because they involved proline isomerization as the rate-limiting step. There are many good reviews on protein folding. These include Jaenicke [19], Dill and Shortle [20] Dill and Chan [21], The “new view” of protein folding, which says there are many ways of reaching the native state, is replacing the single pathway concept. This leads to the concept of folding funnels to replace discrete pathways. Dill and Chan review these concepts [21].

## **1.6 Modification of the covalent structure of proteins**

### **1.6.1 Site directed mutagenesis**

This is the well-established method [22] of altering the nucleotide sequence of a gene to change the amino acid sequence of a protein. A synthetic oligonucleotide complementary to the part of the gene to be modified but including the nucleotide changes is hybridized to a suitable plasmid containing the gene to be mutated. A second strand of DNA is enzymatically synthesized from the first strand, using the

oligonucleotide as primer. The hybrid plasmid is used to transform cells and to make homogeneous double-stranded DNA. The latter are selected by sequencing clones, the plasmid DNA extracted and used in an expression system to produce the mutant protein.

### 1.6.2 Chemically acylated tRNA

Schultz [23-26] has described a general biosynthetic method that makes it possible to site-specifically introduce un-natural amino acids into proteins. The codon of the amino acid of interest is substituted by UAG, a so-called “termination” codon. It is called a termination codon because it causes the protein translation apparatus to stop. Addition of a suppressor tRNA that has been misacylated with the desired amino acid to a protein biosynthesis reaction results in specific incorporation of the new amino acid. A suppressor tRNA is a minor tRNA species, which has an anti codon corresponding to one of the termination codons. It inserts amino acid residues where termination normally occurs, though normally with low efficiency. The misacylated tRNA's are accessible by T4 ligase mediated coupling of aminoacylated nucleotides with tRNA's. The method requires efficient procedures for acylation and mild methods of deprotection of the suppressor tRNA's. This involves elaborate protection and deprotection strategies and yields of acylated material can be low. The recent introduction of the 4-pentenoyl protecting group by Hecht [137], with the use of aqueous iodine for its removal, may permit the synthesis of aminacyl-tRNA's such as aspartic acid analogues that have hitherto proved to be problematic.

### 1.6.3 Native chemical ligation

Chemical ligation is the chemoselective reaction of unprotected peptide segments in solution. Developed by Kent [27], it reacts an unprotected synthetic peptide thioester with another unprotected peptide segment containing an amino terminal Cys residue. The intermediate thioester rearranges to give a native amide bond at the ligation site.

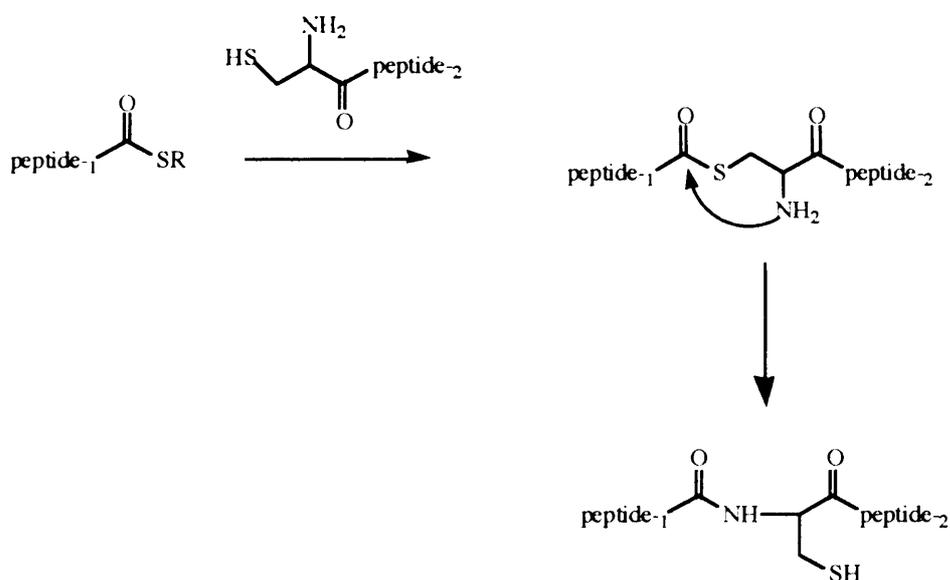


Figure 4. Formation of intermediate thioester and rearrangement to amide bond.

The reaction scheme is shown in Figure 4. It has been used to make a mutant of HIV protease [28], and to make two analogues of barnase. The limitations of the method are that it requires a C-terminal carboxyl activated as a thioester to be available in quantity. It also requires an N-terminal cysteine in the second fragment that may not

be present in the natural protein. Introduction of an un-natural cysteine may also bring other problems like instability and chemical degradation.

#### 1.6.4 Subtiligase

Subtiligase [29,30] is a double mutant of the serine protease subtilisin BPN' and can ligate peptide esters efficiently in aqueous solution. The enzyme attacks the glycolate-phenylalanyl amide of the N-terminally protected fragment to produce the thiol-acyl intermediate. The free N-terminal of the other fragment liberates the free enzyme by attacking the thiol-acyl bond to form the natural amide. The reaction sequence is shown in Figure 5.

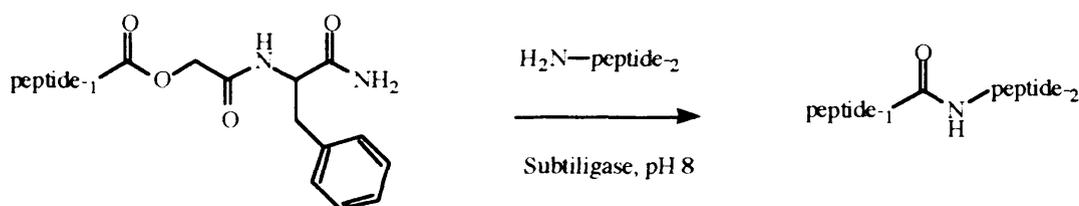


Figure 5. Subtiligase mediated peptide bond formation.

This method was recently used in a synthesis of mutants of the 124 residue RNase A [118]. Fully active RNase was produced in milligram quantities by stepwise ligation of six esterified peptide fragments. Two variants were produced containing an un-natural amino acid 4-fluorohistidine replacing the catalytic histidines at positions 12 and 119.

### 1.6.5 Solid Phase Peptide Synthesis

The principle of solid phase peptide synthesis is outlined in Figure 6. The term “solid phase” is a misnomer as the chemical reactions do not take place in, or on the surfaces of solids which are in heterogeneous contact with solutions, but in gels produced by the penetration of solvent and solute molecules into the polymeric matrix.

An insoluble carrier is first functionalized with a linker molecule called a handle.

Handles are bifunctional spacers that at one end are cleavable protecting groups and at the other contain a functional group, usually a carboxyl group that can be coupled to the insoluble support. A suitably protected amino acid is coupled to the handle. Its N-terminal protection is then removed and the chain elaborated by repetitive coupling and deprotection reactions until the desired sequence has been assembled. The crude peptide results from an acidolytic cleavage from the resin that also removes side-chain protection. The product is then extensively purified and analysed for homogenous material.

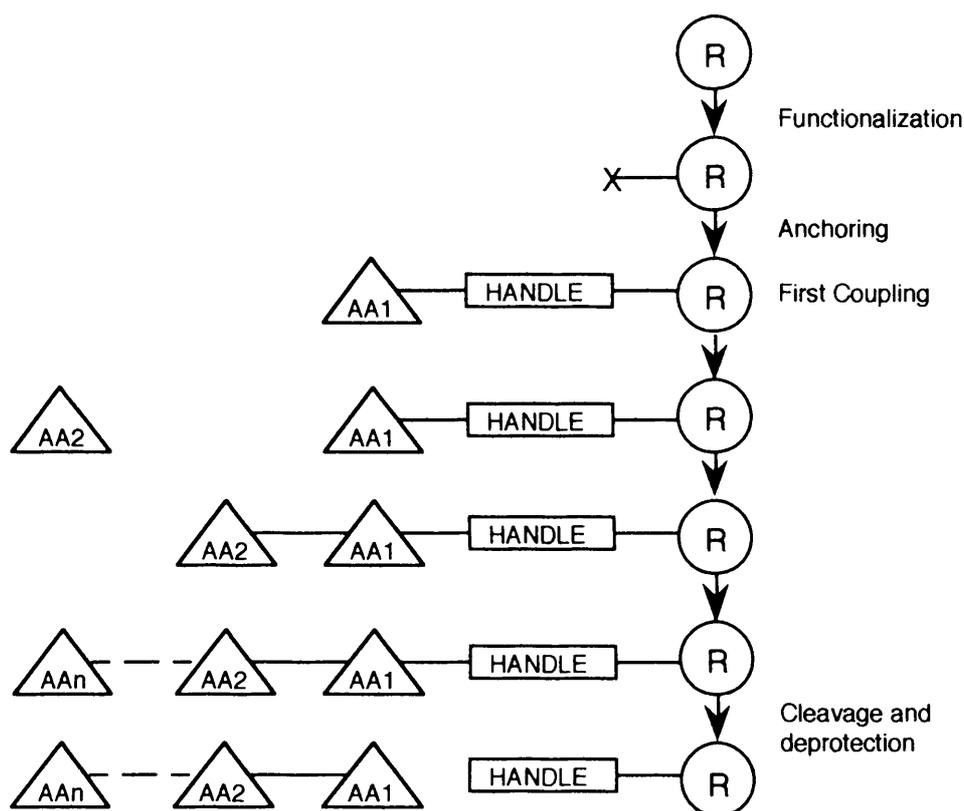


Figure 6. Stepwise assembly of linear peptides. R = insoluble polymeric carrier. AA1 to AAn = protected amino acid residues numbered from the C-terminal. Handle = bifunctional linker molecule used to anchor the chain to the support

Two standard methods exist. The first [31, 32] is based on the original work of Merrifield. The resin is chloromethylated and the target peptide C-terminal amino acid residue, protected with the Boc group, is attached to the resin by nucleophilic displacement. Deprotection with a mild acid, followed by neutralisation, gives a free amino group, to which the second residue may be coupled by means of DCC. The deprotection, neutralisation and coupling steps are then repeated with appropriate Boc amino acids until the desired peptide is fully assembled. Potential reactive side-chains are blocked with protecting groups, derived from benzyl alcohol, that are stable to the mild acid required for removal of the N-terminal protection. At every stage the insoluble resin-peptide is thoroughly washed to remove excess reagents and side

products. Finally, treatment with strong acid removes all the side-chain protection and cleaves the benzyl ester resin peptide link giving the crude peptide.

Coupling of amino acids is mediated by hydroxybenzotriazole esters derived from HBTU. The scheme is summarized in Figure 7.

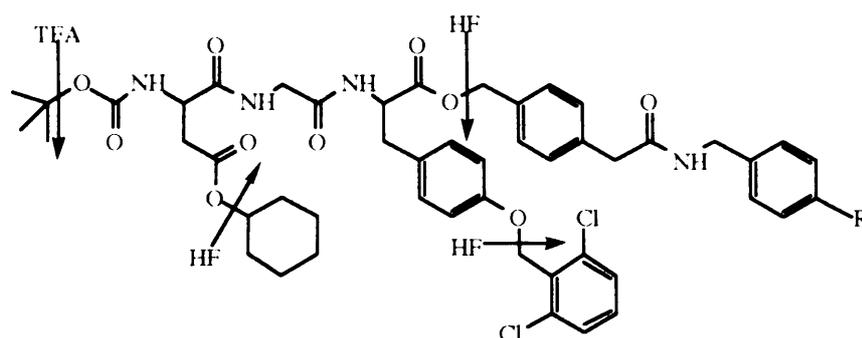


Figure 7. The “Merrifield” protection scheme for solid phase synthesis where R represents the polystyrene backbone.

In 1971, Sheppard reasoned that if the carrier resin was more closely related chemically to the growing peptide (so that both would be equally well solvated in the reaction solvent), there would be better access of reagents to the reactive sites. The standard Sheppard resin consists of a cross-linked polyacrylamide with a number of sarcosine methyl ester side chains.

The latter are extended by treatment with ethylenediamine, addition of a handle 4-hydroxymethylphenoxy acetic acid and a reference amino acid norleucine. The first amino acid is introduced as a symmetrical anhydride.

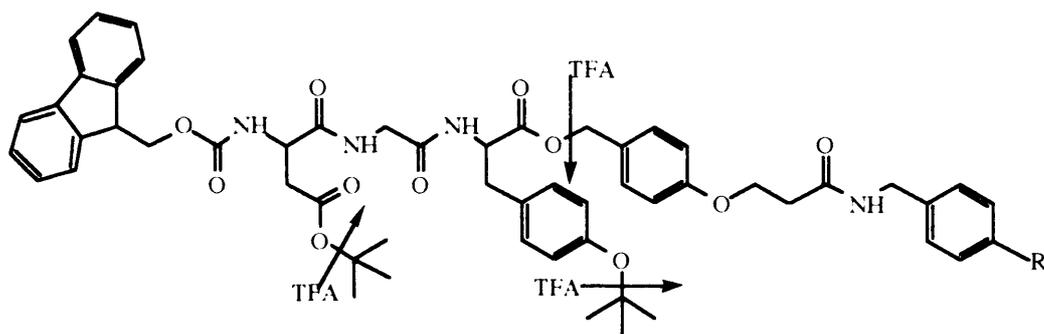


Figure 8. The Fmoc protection scheme for peptide synthesis. R = the resin backbone of polystyrene, polydimethylacrylamide, or polyethylene glycol polystyrene.

The N-terminal Fmoc group, first introduced by Carpino [33, 34] is removed by 20% piperidine in DMF. The chain elongation is continued by active ester coupling of suitably protected Fmoc amino acids. The side-chain protection scheme adopted is one of total orthogonality to the base labile N-terminal protection. Acid labile protecting groups such as Boc, t-butyl ethers and esters are used throughout. A summary of the orthogonal protection scheme is shown in Figure 8 above. At the end of the synthesis, the peptide, with a free N-terminal is removed from the resin and all the protecting groups removed by treatment with TFA and an appropriate scavenger mix to minimise side reactions. The crude peptide preparation is then subjected to a purification regime that may include gel-filtration, ion-exchange high performance liquid chromatography (HPLC), and reverse phase HPLC. Capillary zone electrophoresis (CZE) and laser desorption mass spectrometry (LD-MS) are used as orthogonal means of monitoring the purification protocol. The final homogeneous material is then analysed by amino acid analysis to confirm the number and type of residues, N-terminal Edman degradation to confirm its sequence, and electrospray mass spectrometry (ES-MS) to confirm the identity and purity of the desired

sequence. The reader is referred to several good books on the practice of Fmoc based peptide synthesis [119].

To date, most protein syntheses on solid phase have used the Merrifield Boc methodology. In this approach the differentiation between the temporary N-terminal amino protection on the one hand, and the permanent side chain protection and peptide-resin link on the other hand is dependent on the much greater acid-lability of Boc groups than benzyl ester groups. The difference is sufficient for syntheses involving a modest number of steps, but the slight side-chain deprotection and some loss of peptide from the resin leads to purification problems in the synthesis of long peptides. The Fmoc polystyrene approach has an orthogonal protecting group strategy with the temporary N-terminal protecting group removed by base and the side-chain and resin link by acid. It was thought that this would allow a more routine synthesis of long peptides.

The presence of tryptophan remains a problem to peptide chemistry [35]. The indole ring of tryptophan does not interfere with peptide bond formation, and is usually left unprotected. However, it can be modified by the products from the deprotection of arginine, 4-methoxy-2,3,6-trimethylbenzenesulphonyl (Mtr), 2,2,5,7,8-pentamethylchroman-6-sulphonyl (Pmc) and 2,2,4,6,7-pentamethyldihydrobenzofuransulphonyl (Pbf) residues, by t-butyl cations, and by re-attachment of the peptide to the resin support. This has been a major shortcoming in Fmoc SPPS. The introduction of Fmoc-Trp(Boc)-OH (Figure 134) has largely overcome these problems, especially in syntheses involving many arginines.

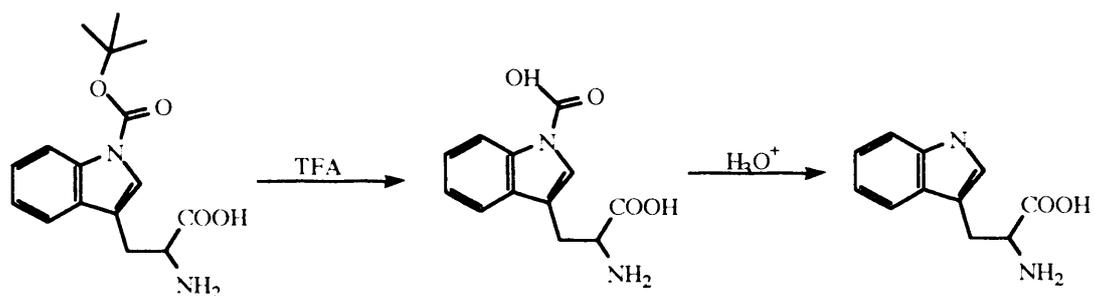


Figure 134. Deprotection of Trp (Boc) with TFA.

During normal deprotection with TFA, the t-butyl group is removed, leaving the indole with an N-carboxy group, thereby preventing alkylation, peptide re-attachment or sulphonation. Tryptophan is regenerated by decarboxylation during the lyophilisation procedure. Trialkylsilanes are very effective scavengers for highly stabilised cations like trityl. However they can cause reduction of the indole ring to indoline. It has been shown that Trp (Boc) greatly reduces the tendency for this to happen.

The other main problem is tryptophan oxidation, which is not reversible. Common oxidation products of the indolyl ring are oxyindolyl and kynureninyl derivatives. The latter derivative results in opening of the five-membered ring of tryptophan and shifting the UV absorbance maximum to shorter wavelengths. Tryptophan oxidation can result from the conditions under which the peptide was handled. The oxidation can be minimized by conducting the deblock under nitrogen.

### 1.6.6 Active Esters.

1-Hydroxybenzotriazole[121] (Figure 9), either as an additive in carbodiimide mediated couplings or combined in to a stand alone reagent such as BOP [122] or HBTU [123] (Figure 10) has become widely adopted for peptide coupling reactions where loss of chirality and minimization of side product formation are a priority.

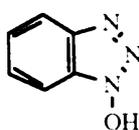


Figure 9. 1-Hydroxybenzotriazole.

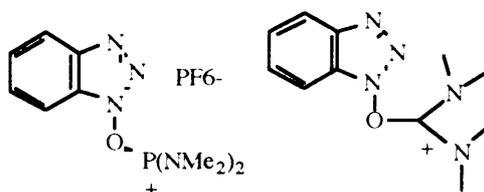


Figure10. BOP and HBTU.

There is a safety reservation about the BOP reagent, because the product generated by its use is the highly toxic hexamethylphosphoramide. For this reason BOP has been largely superseded by the closely related PyBOP [124] (Figure 11).

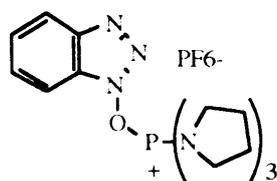


Figure 11. PyBOP.

More recently, Carpino has introduced 1-hydroxy-7-azabenzotriazole [**125**] (Figure 12) as a more efficient additive, which speeds up coupling reactions, reduces loss of chirality, and provides a colour change at the reaction endpoint.

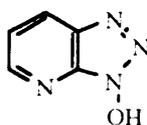


Figure 12. 1-hydroxy-7-azabenzotriazole

The corresponding uronium derivative [**127**] has been shown to give faster, cleaner reactions than HBTU in peptide syntheses. The structure had been assigned as shown in Figure 13 as analogous to HBTU.

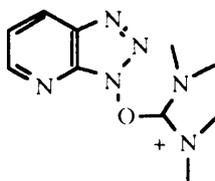


Figure13. HATU.



mutants assessed. The mutant enzyme containing Aib was approximately 1K more stable than the wild type. The other mutations were less stable.

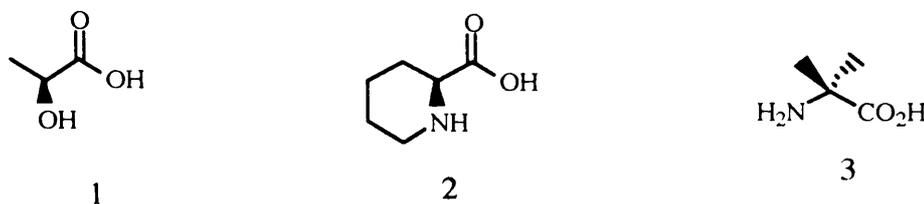


Figure 15. Representations showing stereochemistry of lactic acid(1), pipercolic acid(2) and aminoisobutyric acid(3).

Other residues incorporated were secondary amines such as

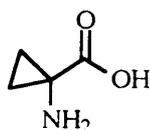


Figure 16. Aminocyclopropanecarboxylic acid.

the Aib analogue amino cyclopropane carboxylic acid shown in Figure 16.

Both of the mutants containing these residues were just as stable as the wild-type protein. The sterically hindered amino acid aminoisobutyric acid (Aib) shown in Figure 17, is the most extensively investigated residue [37].



Figure 17. Aminoisobutyric acid (Aib)

This is a common residue in peptides derived from microbial sources. The replacement of the proton on the  $\alpha$ -carbon atom of alanine with methyl severely restricts the possible rotations about the N-C $\alpha$  and the C $\alpha$ -C' bonds. The torsion angles about these bonds are designated phi and psi. Figure 18 shows the dihedral angle space that 90% of the residues in proteins occupy.

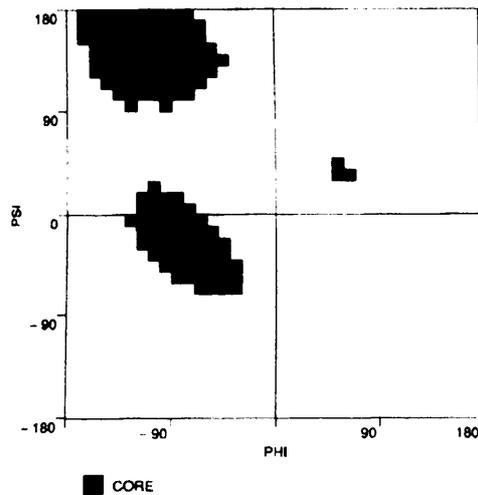


Figure 18. The dihedral angle space that 90% of the residues in proteins occupy.

The allowable phi and psi angles for the Aib residue occur in two regions near  $-57^\circ$ ,  $-47^\circ$  and  $+57^\circ$ ,  $+47^\circ$ . These two regions correspond to a helix. The phi, psi map for Aib (Figure 19) shows that this residue tends to occupy a different dihedral angle space.

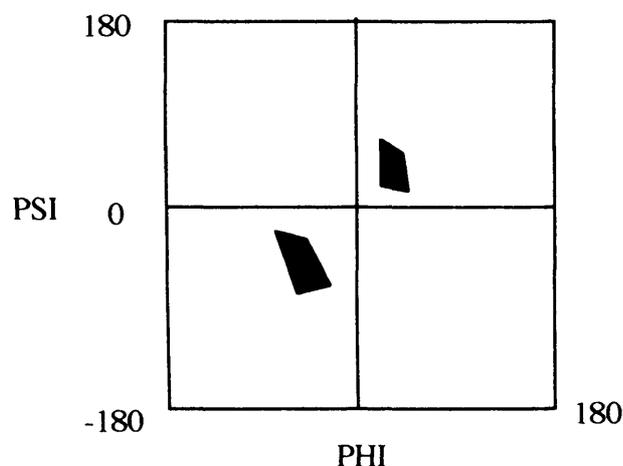


Figure 19. The phi, psi map for Aib

The stabilization of helical peptide conformation by incorporation of Aib residues is a recognised and well-established phenomenon [130]. However, the steric bulk of an  $\alpha, \alpha$ -disubstituted carboxylic acid makes activation and amide bond formation very difficult. The single reported attempt was the introduction of Aib into position 82 of T4 lysozyme [26]. Ala82 is a surface residue at a break between two helices. There have been no reported instances of the replacement of helical residues in a protein with Aib by any method.

#### Dehydro residues

Dehydro residues may restrict backbone conformations. There is evidence that the introduction of these residues induces  $\beta$ -bends in peptide structures [131]. They are found in nature, dehydroalanine, shown in Figure 20, where R is Me, is derived from the dehydration of

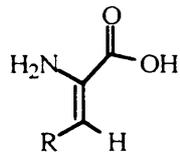


Figure20. Dehydroalanine.

serine residues. The presence of these residues has been shown to impart enhanced resistance to enzymatic degradation.

## D-amino acids

D-amino acids provide an attractive means of stabilizing conformations by chemical synthesis because the precursors for peptide synthesis are readily available and the procedures for incorporation identical to the natural isomer. All-D analogues of HIV protease [38] and rubredoxin have been described. D-amino acids have failed to be incorporated into proteins by the suppressor t-RNA methods so there are few instances of their use in investigations of protein stability.

They are important because they adopt conformations that are generally found in  $\beta$ -turns. They have also been used extensively to increase the resistance of peptides to proteolytic degradation. In 1995, Chalmers and Marshall [39] used molecular dynamics simulation to suggest that D-Pro-NMe Ala, (Figure 21) may be effective as a  $\beta$ -turn stabilizer.

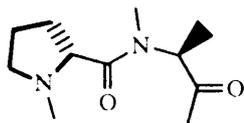


Figure21. D-proline N-Methyl Alanine dipeptide residue.

## Glycine homologues

The introduction of glycine homologues might be expected to introduce flexibility. Little information is available on their effects in peptides, but 6-aminohexanoic acid (Figure22) has been used as a flexible linker between two helical peptides [132].



Figure22. 6-aminohexanoic acid.

Its use in proteins has not been reported. Insertion and deletion mutants have also been used to alter protein structures [133]. They can be viewed as modifications of the spacing between amino acid residues or as mutations targeted at the peptide backbone, because they add or remove atoms. They have been used as probes of the overall fold of domains and are mainly useful in establishing the tolerance of protein structures to changes.

## Peptidomimetics

Peptidomimetics are chemical entities that attempt to mimic structural motifs in peptides and proteins. Kahn has designed and made a mimetic from an antibody complementarity-determining region [40]. For example the

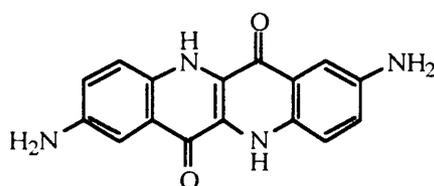


Figure 23. A peptidomimetic used to mimic  $\beta$ -sheet.

diacylaminoepindoleidione, Figure 23, has been used as a template for  $\beta$ -sheet formation. There is no amino acid backbone, but the structure mimics a protein structural motif.

Peptidomimetics have proved important in altering the effect and selectivity of biologically active peptides. There has been relatively little work in the use of peptide mimetics to replace structural elements in proteins. In 1993, Kent reported on the use of the BTD mimetic (Figure 24)

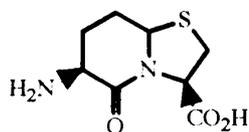


Figure 24. Bis thiazoline dipeptide.

to replace two residues in a turn in native HIV protease [41]. The mutant enzyme was fully active and had the substrate specificity as the native molecule and showed enhanced resistance to thermal inactivation. This demonstrated that the precise geometry, in the turn, was not critical for activity and that replacement of the native sequence with a more rigid structure can lead to enhanced stability.

Turns and loops are essential conformational components of protein structure, but their roles in determining the structure and stability of proteins are not well understood. Some turns appear to simply connect elements of secondary structure, as in HIV protease. In others, the identity of the specific residue at a particular position may be important.

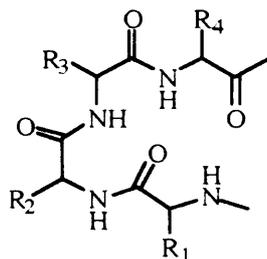


Figure 25. A  $\beta$ -turn representation showing the four residues in the turn.

A typical  $\beta$ -turn is shown in Figure 25. It is formed from four amino acids and is stabilized by a hydrogen bond between the carbonyl group of the first amino acid and the NH group of the fourth.

## 1.7 Thermodynamic Analysis

### 1.7.1 Circular Dichroism (CD).

Circular dichroism refers to the differential absorption of the left and right circularly polarized components of plane-polarized light. This is known as the ellipticity. This will occur when a chromophore is optically active and CD is a measure of the chirality of the sample. Far-UV CD is the most common method used to determine secondary structure, and near-UV CD is used to determine tertiary structure. It is common to use the ellipticity at 222nm as a measure of the secondary structure of a protein. At this wavelength, the major contribution is from the helix, although  $\beta$ -sheet structure also makes a significant contribution.

The principal information available from a CD experiment includes secondary structure, tertiary structure and stability information. CD can yield only “overall”

descriptions of structure in contrast to NMR and X-ray crystallography. However, the main advantage of CD is that it is a much less demanding technique than the other two in terms of time and amount of sample required. CD also offers the possibility of examining the rate of structural changes. It is important to realize however, that there may be significant contributions from aromatic groups at 222nm. This can make quantitative analysis difficult, if the positive contributions from the aromatic groups offset the negative ellipticity of the protein and only a very weak signal is obtained. The size of the CD signal can be used to monitor the unfolding transition of a protein brought on by denaturants such as guanidine and urea. For this type of work, samples of proteins should be homogeneous. Significant errors can arise in determining the molar dichroism values from uncertainties in determining the protein concentration.

#### 1.7.2 Differential Scanning Calorimetry (DSC).

This is a method of measuring energy changes directly as a protein solution is heated at constant rate in a calorimeter cell alongside an identical cell containing the same buffer. Differences in heat energy uptake between the sample and reference cells required to maintain equal temperatures correspond to differences in apparent heat capacity. The shape and area of the transition curve contain thermodynamic information about the unfolding process. The area under the curve per unit of protein is the calorimetric enthalpy  $\Delta H_{cal}$ . The mid-point of the transition is the point at which 50% of the protein is unfolded which, for a simple two-state system, is the temperature at which the Gibbs free energy is zero.

DSC transitions have been measured on a range of small, monomeric globular proteins and so this method is an essential measurement technique for the derivation of thermodynamic information that is independent of the mechanism of unfolding.

### 1.7.3 Denaturant Unfolding

Guanidine, whose structure is shown in Figure 133 below,

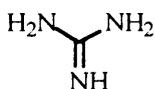


Figure 133. The structure of guanidine.

and urea are the two denaturants, most commonly used for quantitative studies of protein stability. The denaturants bind to sites on the protein and disrupt all the energies that are required for keeping the protein in its native state. The progress of this unfolding may be followed by a variety of physical techniques. The most common is the intrinsic fluorescence of tryptophan, usually situated in the core of proteins and becoming exposed to solvent as the unfolding proceeds. This can result in a four to eight fold change in fluorescence signal for the residue. Because of this, and because of its great sensitivity, precision and ease of data collection, fluorescence is the preferred method for monitoring denaturant unfolding.

## 1.8 Aim

The first objective of this work was to establish a synthetic methodology capable of reliably making small protein domains up to 65 residues in size, containing one or more highly hindered residues.

A model system such as a protease inhibitor or other suitable protein domain that was known to fold well *in vitro* would be selected. Changes would be made in well-defined regions of the structure such as loops, helices and sheets to either constrain structural elements or make them more flexible.

Insertion analogues incorporating the mimetic structures would be made first. The insertion analogues are important to make because they show how well the positions chosen for mimetic substitution can tolerate changes. For stability studies, it is important that the analogue structures are a perturbation of the wildtype structures or no direct comparison of stabilities can be made.

Two approaches would be taken in assessing the effect of backbone substitution on stability and binding activity.

The first would be to substitute a residue in a helix in an attempt to stabilize the helix over wild type. Aib would be substituted for a residue in the helix. This would also be a good test of the synthetic methodology because, to date, no Aib molecule had been inserted into the middle of a protein by chemical synthesis.

The second would be to substitute residues in a loop with mimetics that would loosen and constrain the turn. It has been suggested that turns between segments of secondary structure in proteins play an important part in determining stability.

DeGrado [136] has suggested methods of evolving thermodynamically more stable turns. Genetic methods were used to replace residues in the turns of the Protein G B1 Domain sequence. DeGrado found that it was possible to select mutants containing loops that stabilize the protein by about 12 to 20 kJ mol<sup>-1</sup>. In 1991, Balaram reported the use of 6-aminohexanoic acid as a flexible linker to join two helical peptides [132]. 6-Aminohexanoic acid would be used to replace appropriate residues in the chosen model. The turn would be constrained with a dipeptide mimetic BTD and D-Pro-NMe Ala. This would also be a good test of the synthesis because it involves the sterically hindered coupling of an N-methylated amino acid to proline.

The resulting protein analogues would be assessed for their stability with respect to the wildtype sequence using a variety of physical techniques such as circular dichroism, denaturant unfolding and differential scanning calorimetry.

The binding ability of the protein analogues would also be compared to the wildtype protein by measuring an association constant with a protease, if a protease inhibitor is the model, or an antibody to the wildtype structure.

## Chapter 2: Trypstatin

### 2.1 Proteases and their Inhibitors

Proteases belong to a class of protein that cleave peptide bonds in other proteins.

There are four major types of proteases distinguished by the main functional group in their active site: serine, thiol, carboxyl and metallo. The catalytic mechanisms for all four are different but they share a common property of going through a tetrahedral (due to nucleophile addition) intermediate around the normally trigonal carbonyl carbon of the peptide bond [42]. For serine and thiol proteases the nucleophile is the serine hydroxyl or cysteine thiol at the enzyme active site that reacts to give the acyl-enzyme intermediate which then reacts with water to give the cleaved product; in carboxyl and metallo proteases the nucleophile is a water molecule.

Serine proteases are among the most important physiologically. They split peptide and ester bonds and are specific for the amino acid residue that precedes the bond to be broken. A binding pocket next to the active site determines their specificity. The active site contains a serine, a histidine and a buried aspartate linked by hydrogen bonds. The aspartate is ionized and polarizes the histidine, which in turn polarizes the serine hydroxyl, and makes its oxygen strongly nucleophilic. Serine proteases are known to play a crucial role in the normal functioning of biological systems. They are involved in food digestion, blood coagulation, and fibrinolysis, blood pressure regulation, fertilization, and complement immune reactions. Some proteases release peptide hormones and neuromodulators from inactive precursors or degrade message transmitting peptides, thus beginning or ending a variety of biological responses. Blood clotting is started by a cascade of enzymatic reactions that ends with the

proteolytic cleavage and activation of prothrombin to thrombin [43]. Thrombin is a serine protease that cleaves four arginine-glycine bonds in soluble fibrinogen to insoluble fibrin [44]. The pancreatic serine proteases are active in the digestive tract. Trypsin cleaves the bonds following basic residues such as arginine and lysine. It also activates chymotrypsinogen to chymotrypsin. The latter cleaves bonds following large residues such as phenylalanine, leucine, tryptophan and tyrosine.

There are a great number of potential peptide or protein substrates for proteases in an organism and the process must be regulated. The first level of regulation is usually the cleavage specificity of most proteases. Then the activities are kept in check by inhibitors that fit to the protease very tightly and mimic the transition states of their substrates. For example, the thrombin inhibitor hirudin [45, 46] is the most powerful natural anticoagulant known [47]. It is a polypeptide of 65 residues, which exhibits its anticoagulant properties by binding to thrombin with an association constant of  $10^{10} \text{ M}^{-1}$  and so preventing the cleavage of fibrinogen [44]. Enzymatic inhibition may also be important in several pathological situations. A protease/protease inhibitor imbalance has been proposed as a model for the development of the disease pulmonary emphysema, which is characterized by a decrease in physiological lung function. It is suggested that either an increase in the release of proteases or a decrease in protective serum inhibitors can lead to emphysema [48, 49]. The final stages involve the destruction of lung connective tissue. A similar model may apply to rheumatoid arthritis [50, 51].

The vast majority of inhibitors are directed towards serine proteases. According to the Homologous Structure Alignment Database [52], those inhibitors whose structure is known, can be grouped into five different families on the basis of sequence

homology, location of binding site or location of disulphide bridges and three-dimensional structure.

---

Serine Protease Inhibitor Family	Number of members
Bowman-Birk	5
Kazal [53]	6
Kunitz	10
Potato 1	3
Serpin [54]	8

---

The Kunitz type inhibitors [55] are characterised by a size of approximately 60 residues with six cysteines in three disulphide bridges. They have basic isoelectric points and a broad spectrum of activity towards serine proteases. The best known of this type is basic pancreatic trypsin inhibitor (BPTI) [56]. Recently, a new inhibitor of trypsin called trypstatin [57] purified from rat mast cells has been isolated and found to have 35% sequence homology with BPTI. The average sequence identity of the ten members of the Kunitz family is 38% and trypstatin has been assigned to the Kunitz family on this basis.

Understanding how these inhibitors work is important for two reasons. These proteins are ideally suited for general studies of protein conformation or folding because they are easy to make and purify in the quantities for physical experiments. Detailed knowledge of the structure and reactivity of these inhibitors is needed for a more complete understanding of the control functions they exert in a variety of physiological proteolytic processes. This may lead to the use of synthetic inhibitors as therapeutic compounds in the regulation of certain disease states.

## **2.2 Bovine Pancreatic Trypsin Inhibitor (BPTI)**

BPTI is a small protein of 58 residues folded into a single domain. The folded conformation is stabilized by three disulphide bonds. Cys 14 and cys 38 are shown in blue at the top of Figure 26. The other bridges are formed between cys 5 and 55 and cys 30 and 51. It is found in many tissues throughout the body and inhibits many serine protease proteins such as trypsin, kallikrein, and chymotrypsin with association constants of  $10^{14} \text{ M}^{-1}$ ,  $10^{10} \text{ M}^{-1}$  and  $10^8 \text{ M}^{-1}$ . It is a member of the Kunitz family of inhibitors.

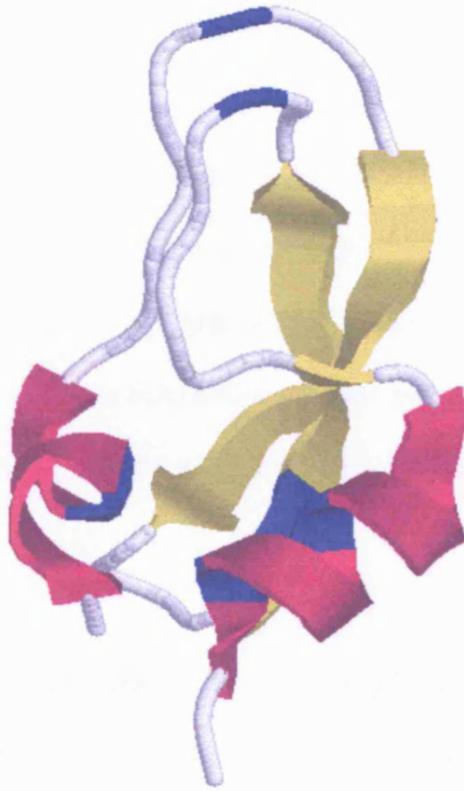


Figure 26. RasMol generated cartoon of the solution NMR generated structure of BPTI (Brookhaven file: 1pti.pdb).

It is one of the most thoroughly studied of all proteins. Its three dimensional structure has been determined by both X-ray crystallography and high resolution NMR spectroscopy [58]. It has also been used for mutational analysis to determine the roles of individual amino acids in determining the folding pathway [59, 60, 61, 62]. It has been cloned and expressed in several systems and together with some of its analogues have been synthesized chemically [63,64]. Studies on thermodynamic stability,

molecular dynamics calculations [65], biosynthesis [66], processing and evolution and prospective therapeutic use have also been performed.

Its disulphide-coupled folding pathway has been studied by Creighton in 1987 [69] and re-examined by Weissman and Kim in 1991 [56]. This pathway remains one of the most detailed descriptions of a protein folding reaction to date. The important feature of the pathway is that the distribution of intermediates is distinctly non-random, and all the well-populated folding intermediates contain only native disulphide bonds.

In 1992 Barany's group synthesized BPTI chemically [63] using a stepwise Fmoc solid-phase peptide synthesis on a polyethylene glycol-polystyrene graft support with p-alkoxybenzyl ester anchoring to the gel. The crude peptide was refolded by oxidation of a dilute solution of the protein at pH 8.7 in the presence of oxidized glutathione. Overall yields of homogeneous proteins were 2 to 4%. Purified synthetic BPTI with the native sequence was indistinguishable from natural material by the analytical and biophysical criteria applied, including circular dichroism (CD) spectra and inhibition of trypsin action.

### **2.3 How HIV harms**

HIV spreads from person to person usually through sexual intercourse, direct exposure to contaminated blood, or direct mother to foetus contact [70][71]. The virus invades CD4 cells [72] and T-lymphocytes, replicates inside them and spreads

to other cells. It gains access to the interior by binding to CD4 itself and to another molecule, a “co-receptor” on the cell surface. The binding enables HIV to fuse with the cell membrane and release its contents into the cytoplasm. The contents are various enzymes and two strands of RNA that carry the entire HIV genome. One of the enzymes, reverse transcriptase [73], copies the HIV RNA into double stranded DNA. A second enzyme, integrase [138] helps to splice the HIV DNA permanently into a chromosome in the host cell. Another enzyme, HIV protease, which is only active as a symmetrical dimer [74], is central to viral infectivity. It cleaves an inactive fusion protein to give reverse transcriptase, and integrase. When a T cell that harbours this integrated HIV DNA becomes activated, the cell replicates and begins to produce new copies of the viral proteins. These proteins anchor to the cytoplasm through the N-terminal tetradecanoic acid portion. The polyproteins aggregate, the adjacent protease monomers dimerize to form active enzyme and cleave themselves off. A cascade of viral proteins is created and in a process similar to crystallisation new virion particles bud from the cell and infect other cells.

The therapeutic strategy that emerges from this data is to stop viral growth by blocking viral replication within cells.

---

<b>Stage</b>	<b>Possible intervention</b>
Target cell binding	Soluble CD4; anti CD4
Fusion to target cell	Antibodies or drugs that block gp41
Viral entry	Drugs or antibodies that block viral entry
Transcription of RNA to DNA	Reverse transcriptase inhibitors e.g. AZT
Transcription and translation	Rev or tat inhibitors; TAR inhibitors
Translation	Antisense constructs against regulatory HIV genes
Gag-Pol cleavage	HIV protease inhibitors

This means inhibiting either reverse transcriptase or the HIV protease [75]. Two drug classes inhibit reverse transcriptase and stop genetic integration [139]. The first class, the nucleoside analogues resemble the natural building blocks of DNA. When reverse transcriptase tries to use one of these to replicate DNA the growth of the strand is stopped. The drug prevents completion of the strand. This class includes AZT the first anti-HIV drug. The second class is the non-nucleoside inhibitor [75]. The protease inhibitors block the active site of HIV protease [76] and so prevent it from cleaving new HIV proteins.

While these strategies were in the early stages of development, workers at Celltech observed cleavage of recombinant gp120 during its production in Chinese hamster

ovary cells. The hypothesis was that post-binding events on the T-cell surface might be necessary for infection [77]. It had been shown that the disulphide bridged V3 loop of surface glycoprotein gp120 was the principal neutralizing determinant of HIV-1 [78]. It was known that antibodies that bind the loop could block HIV infection and cell-cell fusion without preventing the binding of gp120 to CD4 [78,79]. This implied that the V3 loop was involved in the fusion reaction. These antibodies also inhibited syncytia formation, a result of the fusion of CD4 positive cells with infected cells expressing gp120. Mutations in the V3 loop gave rise to non-infective viruses that still could bind CD4. The loop was also cleaved by a variety of proteases [80] including thrombin and trypsin. Consistent with this, HIV-1 infection was reported to be inhibited by trypstatin, [80] a Kunitz type protease inhibitor with limited homology in its active site to a region (GPGRF) at the crown of the V3 loop that is relatively conserved in all HIV isolates [81]. Additionally, antibodies to trypsin, a cell surface serine protease of trypsin-like specificity, inhibited HIV-1 syncytia formation [82]. It had been reported that during the production of recombinant gp 120 the protein was cleaved specifically at a single site between Arg315 and Ala316 (GPGR AFVT) in the V3 loop by a protease from Chinese hamster ovary cells used in the expression system [77]. This suggested that cleavage of the V3 loop might be involved in the fusion reaction. A precedent for this is the proteolysis of the ecotropic Moloney murine leukemia retrovirus envelope that occurs during viral entry.

## 2.4 Trypstatin

The name trypstatin was given to a strong inhibitor of mast cell tryptase, a serine protease produced by rat peritoneal mast cells, first isolated and characterised by Kido in 1988 [59]. It also inhibits chymase and trypsin with association constants of  $10^8 \text{ M}^{-1}$ . The primary sequence of rat trypstatin was determined in 1988 by Kido. It is 75% homologous to the second inhibitory domain of HI30 and some authors believed that rat trypstatin might be a fragment released from rat ITI (Inter alpha trypsin inhibitor) [83]. At this time there was no characterisation of mRNA and no cDNA clone was available. From the amino acid data, trypstatin was a small protein of sixty-one residues, containing three disulphide bonds. It had limited homology in its active site to a region at the crown of the V3 loop (GPGRAF in the loop and GPCRAF in trypstatin). It was reported to inhibit syncytia formation in HIV-I infected cells [80]. In addition, antibodies to tryptase, a cell surface protease, inhibited HIV-1 induced syncytia formation at  $10^{-6} \text{ M}$  [82]. It was proposed that an unknown T-cell surface protease, similar to rat tryptase [84] plays a role in successful HIV infection of T-cells by cleaving the neutralising epitope. The hypothesis was that trypstatin would inhibit this enzyme, block cleavage of the loop and stop cell fusion. The hypothesis is illustrated in Figure 27.

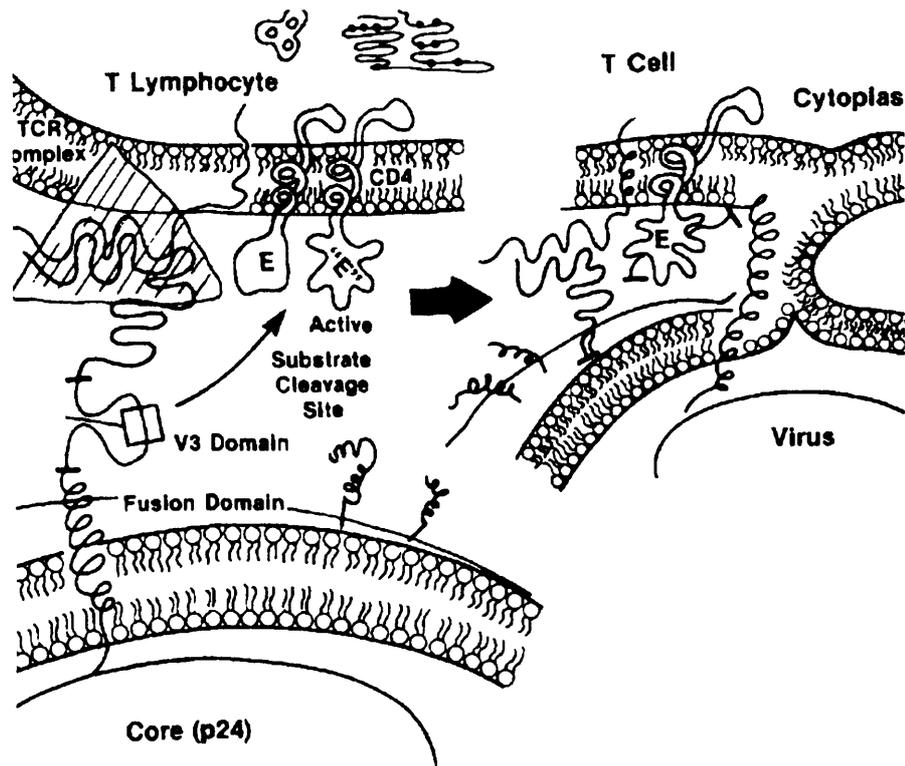


Figure 27. Hypothesis on the possible role of The V3 loop and trypstatin in HIV infection.

The proposed cell surface protease E cleaves a site in the V3 loop. The change in conformation of the gp120 facilitates viral entry to the T cell. Trypstatin was proposed as the inhibitor of E.

Trypstatin was taken to be a Kunitz type inhibitor, by sequence homology, and so is a member of the BPTI family. In the original publication, 6  $\mu\text{g}$  of protein was isolated from the peritoneal cells of 400 rats. Any structural analysis of the peptide would require milligram quantities of material, so isolation of natural material was not a viable approach. The V3 loop and trypstatin would be synthesized chemically because of the precedent set by the synthesis and refolding of chemically synthesized BPTI.

## 2.5 Results

### 2.5.1 Synthesis

The sequence of the V3 loop is: CTRPN NNTRK SIRIQ RGPGR AFVTI GKIGN MRQAHC. It contains one disulphide bridge, six arginine residues and has nine positively charged residues. The linear protected V3 loop sequence was assembled by Fmoc SPPS [85] with couplings mediated by DIPCDI and HOBt in DMF. Triple couplings of each amino acid were necessary to achieve enough homogeneous material for structural studies. Compatible side-chain protection was provided by t-Bu derivatives for serine, tyrosine and threonine, Trt for cysteine, asparagine, glutamine and histidine residues, Boc for lysine and the Pmc protecting group for arginine. Removal of more than four arginine protecting groups per sequence usually leads to side reactions resulting from the increased time of the removal of protecting groups leading to unwanted modification of other residues in the sequence. Under acid conditions, these derivatives give rise to stable carbonium ions that can alkylate tyrosine and cysteine residues irreversibly. Addition of thiophenol, thioanisole and dimethyl sulphide to the deprotection mixture ensured that the side reactions were kept to a minimum. It was decided to synthesize the polypeptide chemically to incorporate the disulphide bond as a separate step and to establish methodology for the synthesis of analogues. Cys (Trt) was chosen for this synthesis because it would be removed under the same conditions as the t-Bu group and the free thiol would not be modified if kept under acid conditions. When assembly and removal of the protecting groups had been successfully accomplished, the disulphide bond was established by oxidation with oxygenated basic buffer of the free cysteine thiols. The

polypeptide was purified using gel-filtration, ion exchange, and reverse phase chromatography with analytical monitoring by capillary zone electrophoresis (CZE). The yield was 28mg of homogeneous material. Figure 28 shows a CZE of the material

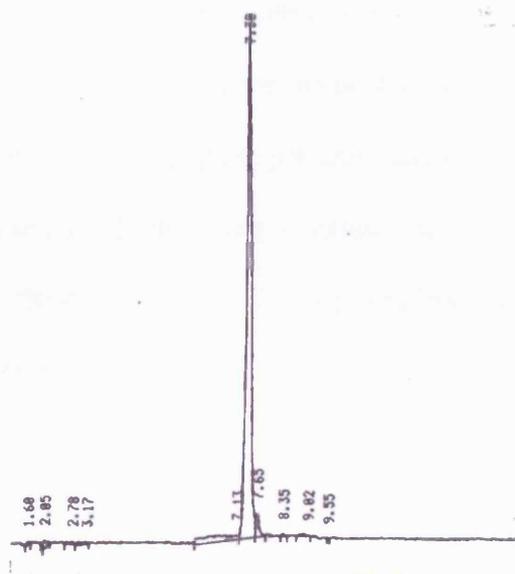


Figure 28. CZE of chemically synthesized V3loop peptide.

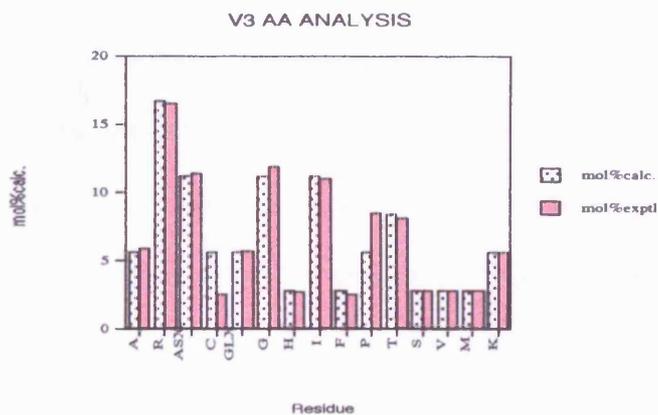


Figure 29. Amino acid analysis of V3loop peptide.

Figure 29 shows the amino acid analysis on the peptide. The value for cysteine is low because it is partially destroyed under the conditions of the analysis.

The concentration of free thiols in the loop was derived from the reaction of a known aliquot of the loop with 4,4'-dithiodipyridine at pH6 for twenty minutes. The V3 loop concentration was calculated using the net peptide weight from the amino acid analysis to give the actual mass of peptide per unit volume. Absorbance measurements were taken at 324nm, using an extinction coefficient of 4,4'-dithiodipyridine as  $19800\text{M}^{-1}\text{cm}^{-1}$ . In the V3 loop solution at 1.5mM there was 0.0029 mM free thiols viz. < 0.2%.

The sequence of trypstatin is

IAACN LPIVQ GPCRA FAELL AFDAQ QGKCI QFIYG GCKGN NKNFY  
SEPKC KWYCG VPGDG Y

It is very basic with an isoelectric point at pH 10.8. It contains one tryptophan residue and six cysteines. Trypstatin was assembled on an Applied Biosystems 430A automated peptide synthesizer, using the standard FMOC/t-Butyl/HBTU strategy. It was decided not to attempt to differentially protect the cysteines and try to form the disulphides sequentially. Trypstatin has 35% sequence homology with BPTI, a known compact structure. Many in vitro experiments have established that BPTI can be made to fold spontaneously from a denatured configuration [86]. It was reasoned

that under the right oxidative conditions the peptide would fold into its native conformation.

At the end of sixty synthetic cycles 1.4g of protected trypstatin resin was obtained. After deprotection and cleavage from the resin, done under nitrogen and in the presence of dimethyl sulphide to reduce tryptophan modification, 371 mg of crude lyophilised product was obtained (47% of theoretical). Analytical HPLC revealed one major peak as shown in Figure 30.

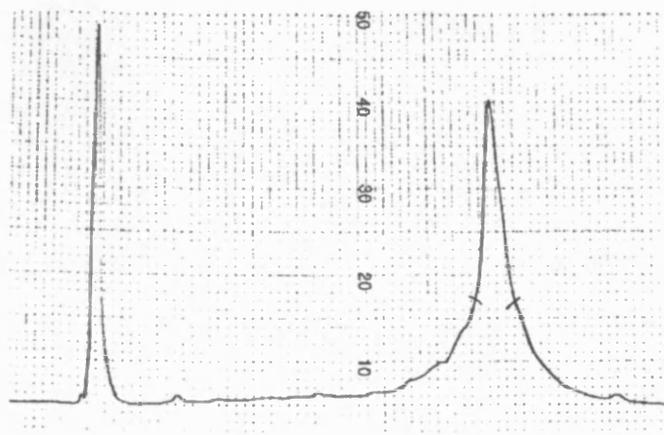


Figure 30. 10 mg of crude reduced trypstatin. Chromatography conditions: Dynamax C8 column, 225 nm, 20 - 50% acetonitrile in 0.1% TFA over 30 min at 10 ml min<sup>-1</sup>.

After thirteen preparative runs and lyophilisation of the main peak, the yield was 264 mg (22% overall, 71% from crude). An analytical HPLC of this material is shown in Figure 31.

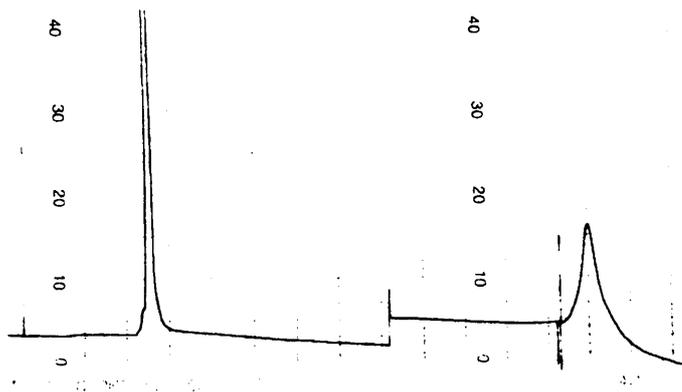


Figure 31. Reduced trypstatin after purification. Chromatography conditions:  
Dynamax C8 column, 225 nm, 20 - 50% acetonitrile in 0.1% TFA over 30 min at 10 ml/min

Figure 32 below shows the matrix-assisted laser desorption mass spectrum of reduced trypstatin. The sample was prepared by adding a solution of the peptide to a solution of the matrix material, a-cyano-4-hydroxycinnamic acid, and pulsed by a nitrogen laser. The masses measured by LD-TOF are generally within the 0.02% of those calculated from the amino acid sequences. In this case, the measured value is 6618 compared with the calculated value of 6610.4.

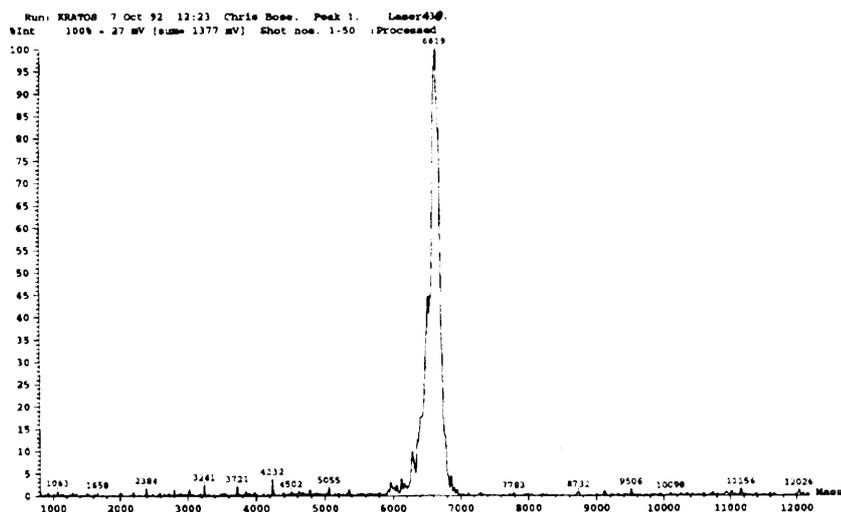


Figure 32. Matrix-assisted laser desorption mass spectrum of reduced trypstatin

Figure 33 shows the electrospray mass spectrum of synthetic trypstatin. The measured value is 6610.4, in agreement with the calculated value for the reduced peptide. The accuracy of the result confirms that the formation of the oxidized species should be able to be confirmed by electrospray mass spectrometry.

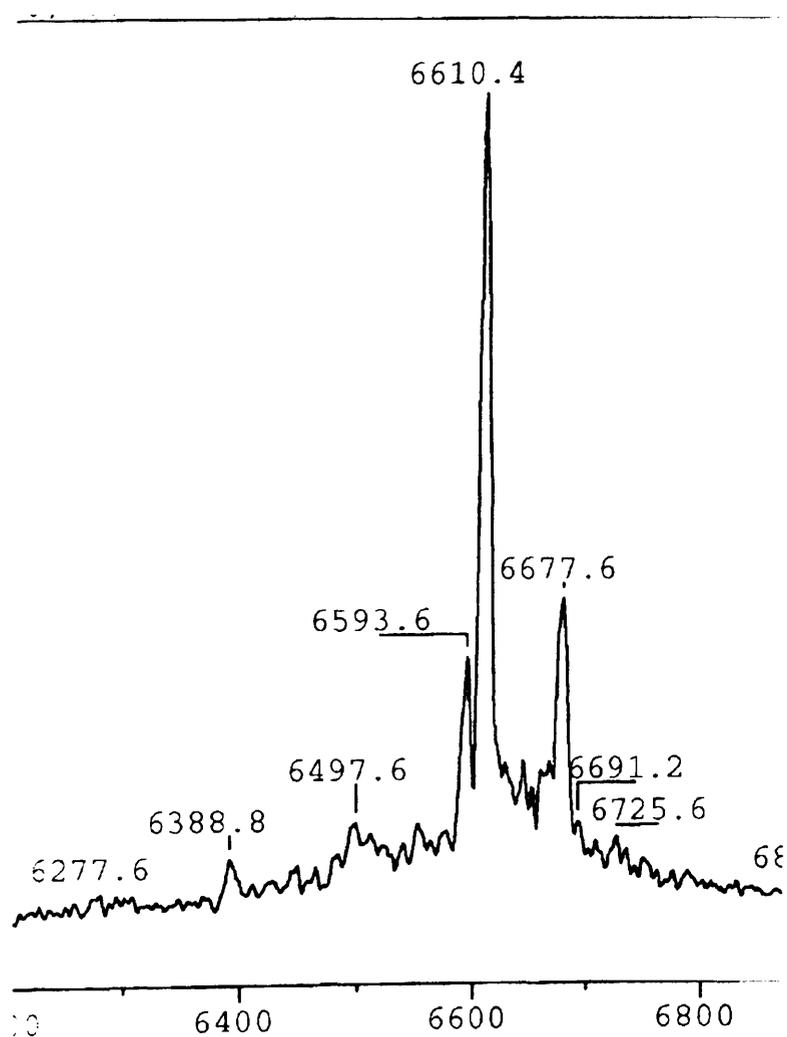


Figure 33. Electrospray mass spectrum of synthetic trypstatin

## 2.5.2 Refolding

The following methods from the literature to fold synthetic peptides or to refold denatured natural proteins were attempted. Air oxidation is the method of choice for single disulphide proteins. This procedure worked for the V3 loop described above. When applied to reduced trypstatin, the solution became very cloudy and all the reduced trypstatin precipitated, as lyophilisation of the residue on filtration recovered 95% of starting material.

Human transforming growth factor – alpha had recently been synthesized chemically by Kent [87]. The synthetic peptide was refolded in guanidine buffer. The same conditions were tried with reduced trypstatin. No oxidized material was recovered after chromatography and there was a considerable amount of precipitation.

Echistatin [88], a polypeptide from the venom of the saw-scaled viper, *Echis carinatus*, containing 49 amino acids and 4 disulphide bridges is a potent inhibitor of platelet aggregation. In the final step, air oxidation of the octahydroderivative was found to be optimal at pH 8. The synthetic product was shown to be physically and biologically indistinguishable from native material. Reduction of purified synthetic echistatin to octahydroechistatin with dithiothreitol followed by air oxidation regenerated homogeneous echistatin in quantitative yield. This highly specific refolding strongly suggests that the linear sequence of octahydroechistatin contains all of the information that is required for the proper folding of the peptide. These conditions were applied to trypstatin at a protein concentration of 0.02 mg/ml. HPLC

analysis gave a very broad diffuse peak indicative of aggregated material. Material losses were consistently high when this type of HPLC profile was generated.

Interleukin-8 is a 72-residue protein with two disulphide bridges that had recently been synthesized chemically on solid phase [89,90]. For refolding of the synthetic peptide, the reduced peptide was dissolved at a concentration of 10 $\mu$ M in a buffer containing 1M guanidine, 0.05M TRIS, GSH at 1mM, GSSG at 0.1 mM, pH 8.5, and stirred for three hours at RT. These conditions were applied to reduced trypstatin. HPLC analysis gave a very broad diffuse peak indicative of aggregated material.

Clark-Lewis has synthesized interleukin-3 chemically using solid phase techniques [91]. HPLC now showed disappearance of the reduced species, and appearance of many peaks with retention times less than that of the reduced material, but no indication of any main peak.

The synthetic material did not fold under “ Creighton conditions” [92] a buffer of 100 mM TRIS, 200 mM KCl, 1 mM EDTA, 10 mM GSH and 1 mM GSSG pH 8.8. These conditions were applied to trypstatin at a protein concentration of 1 $\mu$ M. The mixture was stirred overnight at room temperature and a precipitate developed. HPLC analysis of the supernatant showed only material that co-ran with a marker of reduced material.

The HIV protease is a 99-residue aspartyl protease that is essential for viral maturation. It exists as a dimer. The peptide was synthesized chemically [93] and

produced as a homogeneous protein by RP-HPLC but was enzymatically inactive. It was denatured in 8M urea and then refolded by dilution into a sodium acetate buffer pH 5.5 containing glycerol, ethylene glycol and Nonidet P-40. These conditions were applied to reduced trypstatin. HPLC analysis showed no change in the profile from reduced material after 18 hours incubation.

The Bowman-Birk inhibitor is a 71-residue serine protease inhibitor. Its two subdomains, which contain 7 disulphide bonds, have distinct trypsin and chymotrypsin inhibitory properties. In the study by Flecker, on variants of the inhibitor, refolding conditions described for the parent protein were unsuccessful with the analogues. It was demonstrated that refolding in the presence of trypsin-Sepharose [94] as a template gave a ten-fold increase of correctly folded material than in the absence of trypsin. A trypsin-Sepharose column was made and trypstatin diluted into 0.1M triethylamine/0.3M NaCl (pH7.8) and carefully added with gentle swirling to the trypsin-sepharose matrix, pre-equilibrated in the same buffer. The column was washed with equilibration buffer, then eluted with 0.25M KCl/0.01M HCl (pH 1.5). The eluted fractions were analysed by HPLC. No protein material was returned under the conditions described.

Bovine pancreatic trypsin inhibitor is a 58-residue protein containing three disulphide bridges and has been widely used as a model for addressing folding issues. The first chemical synthesis was published in 1992 [65]. The authors used different conditions from others reported for folding and oxidation of the synthetic material. There was substantial precipitation and a broad HPLC peak of non-native oxidized structures

when glutathione redox buffers were used. Oxidation of the reduced protein, final concentration  $14\mu\text{M}$  was effected in the presence of oxidized glutathione (10 equiv.). These conditions were applied to the reduced tryptostatin, final concentration  $1.5\mu\text{M}$ . HPLC showed a broad unresolved peak characteristic of aggregated soluble material. No folded material was isolated.

The following method is as described by Cleary [95]. The powder (20mg) was suspended in 20 ml of 25mM Tris-HCl/ 6M guanidine hydrochloride, pH 8.0, with shaking for 10 hours. This solution was diluted 5-fold, in three incremental additions, with 25 mM Tris-HCl/ 1.25 mM each of oxidised and reduced glutathione, pH 8.0, with 2 hour intervals between additions. The supernatant was then acidified with acetic acid, gelfiltered on Sephadex G-25, and the eluant lyophilised. HPLC showed a broad unresolved peak characteristic of aggregated soluble material.

Human granulocyte colony stimulating factor [96] is a 175-residue polypeptide with a free cysteine and two intramolecular disulphide bonds. It is one of the hemopoietic growth factors, which plays an important role in stimulating functional activation of blood cells. In the production of this molecule by recombinant means, copper sulphate was used in the oxidation step. Copper ions and other trace metal catalyze air oxidation of proteins because of their ability to accelerate thiol oxidations. These conditions were applied to the reduced tryptostatin, final concentration  $1.5\mu\text{M}$ , with a copper sulphate concentration of  $40\mu\text{M}$ . HPLC showed a broad unresolved peak characteristic of aggregated soluble material. No folded material was isolated.

## DMSO oxidation

All of the refolding conditions reported used basic pH to affect complete oxidation. It was thought that since trypstatin was reported to be a basic hydrophobic protein with an isoelectric point at pH 10.8 that the high levels of precipitation observed may in part be due to the maintenance of the protein solutions at high basic pH's for considerable lengths of time. Tam [146] had demonstrated a selective and efficient method for disulphide bond formation in basic peptides by dimethyl sulphoxide. The oxidations were found to proceed between pH 3-8 with particular suitability for basic and hydrophobic peptides. In all the experiments, although precipitation was not observed, RP-HPLC profiles showed the presence of aggregated material.

In all of the above methods, varying concentrations of the reduced peptide were tried; syringe pump addition at different rates was tried; monitoring of the refolding reaction was tried; different basic pH's were tried.

## Use of a C-terminal analogue to assist folding.

Most prokaryotic and eukaryotic proteases are synthesized as inactive precursors that are normally activated to the functional state by cleavage of an N-terminal extension, the pro region. Recent in vitro studies have shown that the pro region can also be necessary for the correct refolding of the protease. Subtilisin E, a basic serine protease of 275 residues is produced from a pre-pro protein. Deletion of the pro sequence region yields mature but inactive subtilisin. The 77 amino acid pro

sequence region must precede the mature protein to allow the transition to active enzyme [97]. Alpha -Lytic protease [98], an extracellular serine protease, is made as a pre-pro protein. Engineered proteins lacking the 166 amino acid pro region were enzymatically inactive. In fact, the pro region does not require a physical linkage to activate the protease domain. Independent expression of the pro and protease domains produced active enzyme *in vivo*. Carboxypeptidase Y from yeast is made as a precursor of 532 amino acids residues containing a signal peptide of 20 residues and a pro region of 91 residues. The pro region is necessary for the correct and efficient folding of the enzyme *in vitro* and *in vivo*. This was demonstrated by the fact that the proenzyme, denatured in 6M guanidinium chloride, is renatured efficiently after dilution of the denaturant, whereas the mature enzyme regains little activity in the same procedure [99, 100, 101, 102]. *In vivo*, BPTI is synthesized as a peptide with a 13 residue amino-terminal pro region. This pro region contains cysteine residue 10 residues from the N-terminal of the mature protein. Weissman and Kim found that, *in vitro*, both the rate of formation and the yield of properly folded material increased substantially when pro-BPTI was used. They found that the cysteine was necessary for this effect. Replacement by an alanine abolishes the region's ability to facilitate folding. Further, by adding a cysteine to mature BPTI by a ser-gly polymeric repeat also increased both the rate and yield of productive folding. Interestingly, the yield was greater with the synthetic C-terminal region [103].

An analogue of trypstatin containing such a C-terminal region, called trypstatin-74 was made and purified to see if this molecule would fold into a native conformation, using the extra cysteine as an intra-molecular disulphide-shuffling reagent. The

sequence of trypstatin-74 is IAACN LPIVQ GPCRA FAELL AFDAQ QGKCI  
QFIYG GCKGN NNKFY SEPKC KWYCG VPGDG YSGGS GGSGG CSGG-OH.  
This material was made in an analogous way to the parent compound. The analytical  
reverse-phase HPLC is shown in Figure 34. The main peak is characteristically broad  
for this length of polypeptide on a shallow gradient.

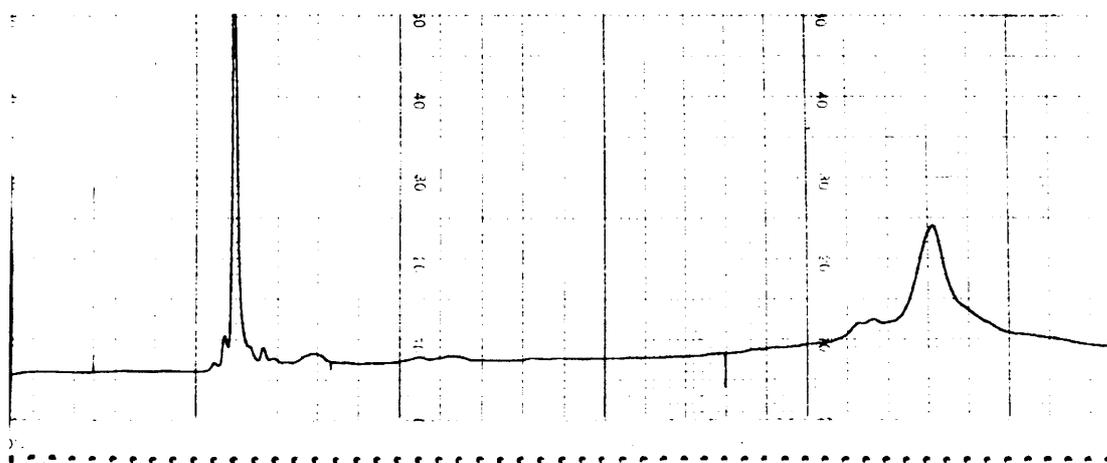


Figure 34. Analytical reverse-phase HPLC of trypstatin-74.

The main peak was collected and analysed by CZE and electrospray-MS.

Figure 35 shows the electrospray mass spectrum. The calculated molecular weight for  
trypstatin-74 is 7517. The experimentally determined molecular weight is 7520.

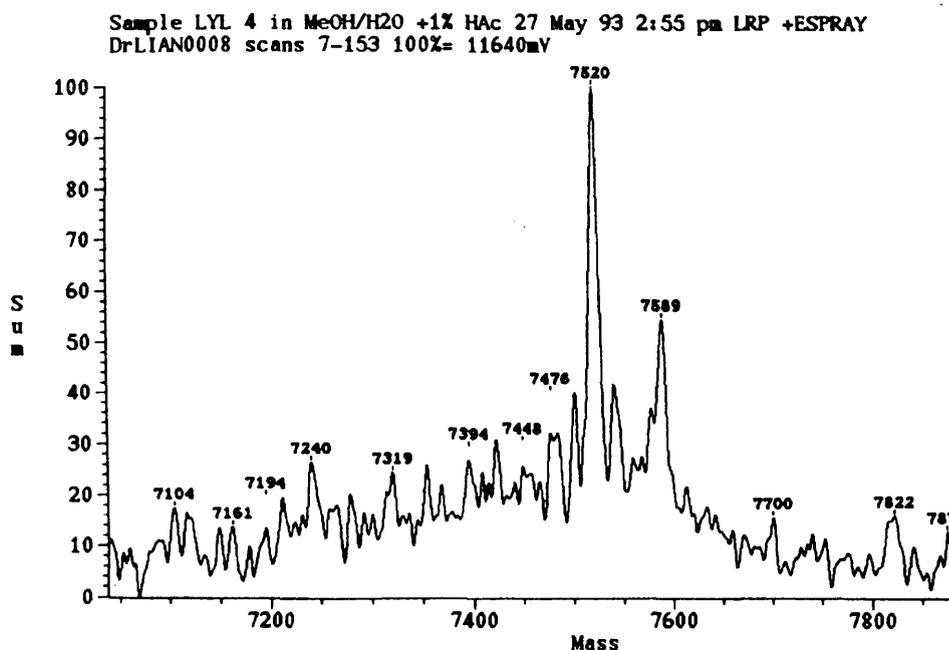


Figure 35. Electrospray mass spectrum of trypstatin-74.

The ES-MS of trypstatin and trypstatin-74 show traces of material with a mass difference of +67 and 69 respectively. This is probably due to  $\beta$ -piperidide formation from the aspartimide formation from Asp (Ot-Bu) followed by aminolysis with piperidine. This side product formation has been reported in the synthesis of aspartyl containing peptides. The trypstatin-74 was to be used under experimental conditions where BPTI was known to refold to active material.

### 2.5.3 Conclusion

The disulphide bridged V3 loop of gp120 was made by solid phase peptide synthesis. This molecule had not been made before chemically. Some preliminary NMR experiments were completed but no full structural determination was made because

the use of the structure was dependent on the successful refolding of the trypstatin moiety and its subsequent solution structure.

Reduced trypstatin was synthesized chemically by solid phase methods. The reason the crude trypstatin would not fold appears to be that since it is a fragment of a larger protein the extra sequence is required for the refolding *in vitro*. The amino acid sequence derived from protein sequencing data is not sufficient evidence of a new protein. The mRNA should be identified and a cDNA made to establish the peptide sequence as a separate molecule. The experimental evidence should have been reassessed when it was not possible to produce  $\mu\text{g}$  quantities of folded protein from the refolding experiments. As soon as no trypsin inhibitory behaviour could be seen in the results of the standard refolding procedures for serine protease inhibitors, doubts about the validity of the sequence should have been raised. More time and effort should have been spent on completing the molecular biology experiments before substantial efforts were put into the biochemistry. If that was not possible, greater efforts should have been made to isolate the natural 'trypstatin' from the mast cells to use as a marker for the synthesis should have been pursued. However a complete literature search for folding methods has enabled a comprehensive summary of protein refolding techniques to be compiled, applied and compared.

This work was abandoned at this stage after all attempts to refold the synthetic material failed.

## Chapter 3: Protein G

### 3.1 Protein G

Protein structures are the product of evolution that probably optimized them for their function and not necessarily for their stability. It should therefore be possible to modify a protein to improve its stability, although perhaps with the risk of compromising some function of the protein.

Many species of pathogenic bacteria have proteins on their surface that bind to immunoglobulin and these interactions may help the organism evade the host's defences. One of these proteins is protein G, which is expressed on the cell surface of *Streptococcus* [104]. It is a multidomain protein of approximately 600 amino acid residues that has high affinity for IgG from many mammalian sources.

Protein G B2 domain was chosen as a model to investigate the effect of chemically synthesized mutations on stability, binding to IgG and its solution structure. It has been cloned and expressed in *Escherichia coli*, its crystal structure determined at high resolution both alone and complexed to Fab, and its solution structure determined by NMR. It undergoes reversible unfolding, induced by guanidinium hydrochloride and heat, which fits to a two state process at equilibrium. The B1 and B2 domains have been characterised thermodynamically [105,106]. It is a monomeric, single domain protein without disulphide bridges, which is accessible by chemical synthesis [107]. The IgG binding regions have been located in the carboxyl terminal of protein G in three domains of 55 amino acids separated by short linker regions. Each of the domains has been expressed in *Escherichia coli*, shown to fold independently and bind to IgG with high affinity. Their structures have been determined in solution by

two-dimensional nuclear magnetic resonance [108, 109] and by X-ray crystallography [110, 111, 112].

The basic folding pattern of the highly homologous domains is a four-stranded  $\beta$ -sheet of the antiparallel-parallel-antiparallel type spanned by an  $\alpha$ -helix. The domains use two distinct regions to bind to  $F_C$  and to Fab, with  $K_D$  of 10 nM and 1  $\mu$ M respectively. Microcalorimetry of B1 and B2 domains show a melting temperature of 87°C at pH 5.4 and completely reversible thermal denaturation. [105,106]. A ribbon drawing of the three-dimensional structure of protein G B2 domain is shown in Figure 36. The N-terminal leucine is shown in green.



Figure 36. A ribbon drawing of the three-dimensional structure of protein G B2 domain. The N-terminal leucine is in green.

The three-dimensional structure of protein G B1 domain was determined in solution by NMR by Gronenborn and Clore in 1991 [113], and that of Domain II in 1991 and 1992 by Lian and Roberts [108]. The crystal structure of the B3 domain in a complex with the Fab portion of an IgG1 was determined in 1992 by Derrick and Wigley [100, 101]. Figure 37 shows an edge to edge interaction between the  $\beta$ -sheet of the domain and that of the CH1 domain of the antibody fragment.

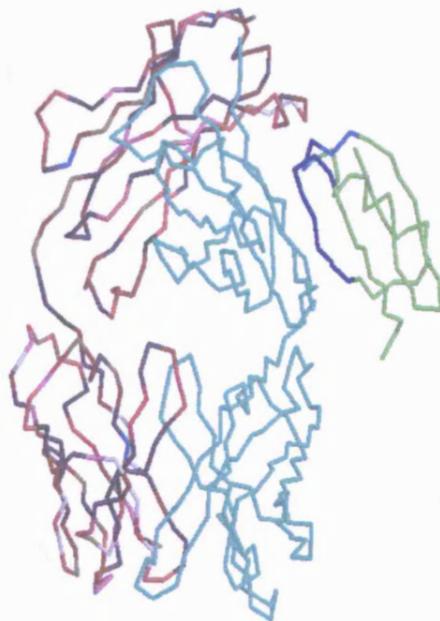


Figure 37. RasMol drawn representation of protein G B3 domain complexed to Fab fragment of IgG1. The light chain is multicoloured; the heavy chain is green

It was suggested that the mode of binding observed in the Fab complex was a non-specific interaction due to crystal packing. However, NMR studies, using differences in chemical shift and linewidths on binding to Fab of a protein G domain, have

identified two regions of the molecule that are involved in Fab binding [114]. The first region is the turn between the first and second strands of the  $\beta$ -sheet and six residues of the second  $\beta$ -strand and the second is the flexible loop region following the helix (residues 41 to 45).

These regions are shown in Figure 38 and are entirely consistent with the crystal structure. W48 and T58 were also affected by the binding to Fab. It was concluded that the crystal structure does correspond to the structure of the complex in solution and is not an artefact of the crystal packing.



Figure 38. RasMol drawn representation of protein G B2 domain with the NMR derived binding residues in blue. The N-terminal leucine is shown in green

B1 and B2 domains have also been analysed by differential scanning calorimetry [116]. They have high denaturation temperatures but the state functions show that they are only marginally stable at room temperature. This is because the stability profile ( $\Delta G$  vs. temperature) is flat and shallow due to the small  $\Delta S$  and  $\Delta C_p$  for unfolding.

Hydrogen-deuterium exchange experiments [115,116] have also been used to measure exchange rates for all the main-chain amide protons in domains B1 and B2 of protein G. These studies provide complementary information to calorimetry. Instead of a macroscopic stability parameter  $\Delta G$ , NH exchange rates provide a very sensitive measure of local stability differences between domains B1 and B2. In both domains the  $\beta 2$  strand is the least stable of the  $\beta$ -sheet. Local stability differences are also evident near the  $\beta 1$ - $\beta 2$  turn. Knowledge of the local stability differences in the  $\beta 1$ - $\beta 2$  region provides potential approaches for designing new stability analogues for protein G. For non-H-bonded NHs, exchange rates can be compared with their calculated intrinsic rates. Most non-H-bonded NHs have exchange rates which are <15 times slower than their intrinsic rates. However, exchange rates for Lys15, Thr16 and Leu17 in the  $\beta 1$ - $\beta 2$  turn are five to  $\geq 24$  times faster than their intrinsic rates. From the X-ray structure, the three NHs of Lys15, Thr16 and Leu17 are pointing in the same direction and may generate a localised electrostatic field. This may increase the local hydroxide ion concentration, which would contribute to the increased exchange rate.

### 3.2 Synthesis

The second domain of protein G has 64 amino acid residues.

NH<sub>2</sub> –LTPAV TTYKL VINGK TLKGE TTTEA VDAAT AEKVF KQTAN DNGVD GEWTY  
DDATK TFTVT EKPE-OH.

Wild type B2 domain

Protein G B2 Domain (6-64) was made using standard Fmoc/t-butyl solid-phase peptide synthetic methodology, with the base labile Fmoc group for temporary N-terminal protection and pre-loaded Fmoc-Glu(Ot-Bu)-p-hydroxymethylphenoxymethyl polystyrene resin. Tryptophan was protected on the side-chain with BOC to reduce side reactions during the 59 steps. A double coupling procedure using an eight-fold excess of O-(Hydroxybenzotriaz-1-yl)-1,1,3,3,-tetramethyl uronium hexafluorophosphate (Figure 39) activated amino acids in the presence of diisopropylethylamine was used.

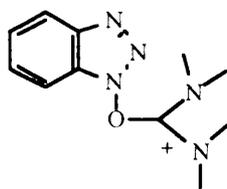


Figure 39. The coupling agent O- (Hydroxybenzotriaz-1-yl)-1,1,3,3, -tetra methyl uronium hexafluorophosphate

Total deprotection of the side chains and cleavage from the resin were performed with trifluoroacetic acid in the presence of water and triisopropylsilane as cation scavengers. After precipitation with ether, the peptide was dissolved in 6M guanidine, 0.1M TRIS, 2 mM EDTA pH 7.2 to break up any aggregated and misfolded material. The solutions were then slowly diluted into 20 mM TRIS pH 7.4 (1 L) by syringe pump to promote the correct folding. When methods were being tried in the attempted refolding of tryptostatin, slow dilution of guanidine solutions into a buffer seemed to be applicable across a broad spectrum of protein and peptide types.

The crude peptide was purified by anion exchange HPLC as shown in Figure 40.

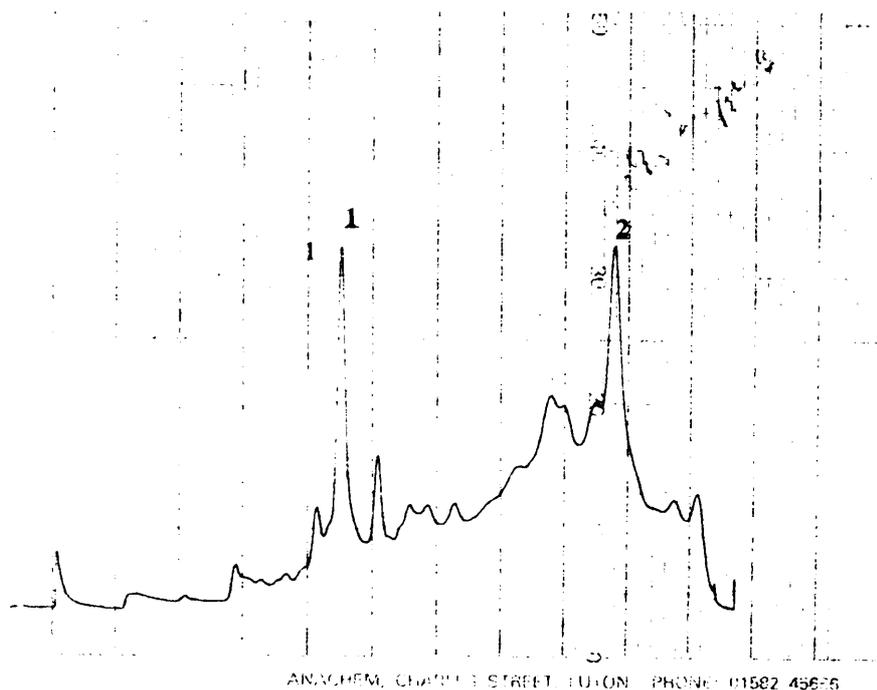


Figure 40. Anion exchange profile of the crude wild-type Protein G B2 domain. Synchropak AX100 column. 0 to 40% 20mMTRIS pH7.4 1M NaCl 8ml/min, 225nm detection.

The two fractions were purified by reverse phase HPLC and the main eluted fraction lyophilised. Fraction 1 contained material whose amino acid analysis was in good agreement with theory. Fraction 2 contained material that was deficient in three lysines. This result indicated poor couplings at lysine residues in general and for subsequent syntheses the activation time for lysine was increased to allow for the observed poor solution of Fmoc-Lys(BOC)-OH in NMP.

The analytical RP-HPLC of fraction 1 is shown in Figure 41.

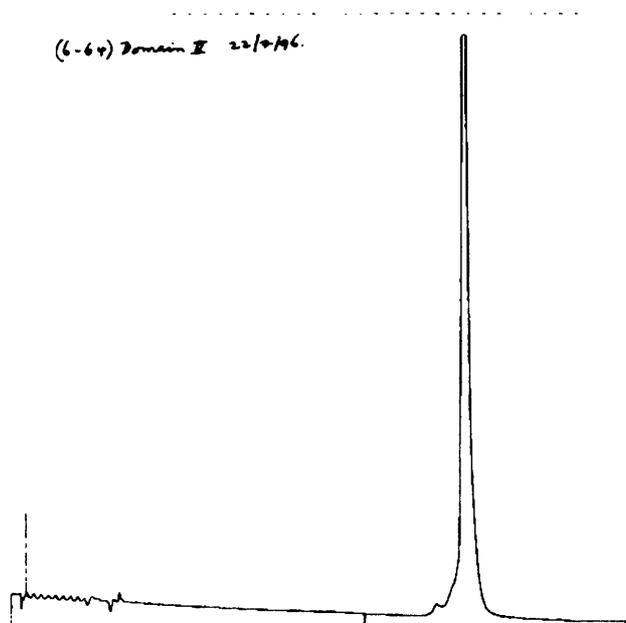


Figure 41. Reverse Phase HPLC: Hypersil C<sub>8</sub> column, 225 nm detection. 10 to 50% acetonitrile in 0.05% TFA.

The ES-MS spectrum corresponding to this peak is shown in Figure 42.

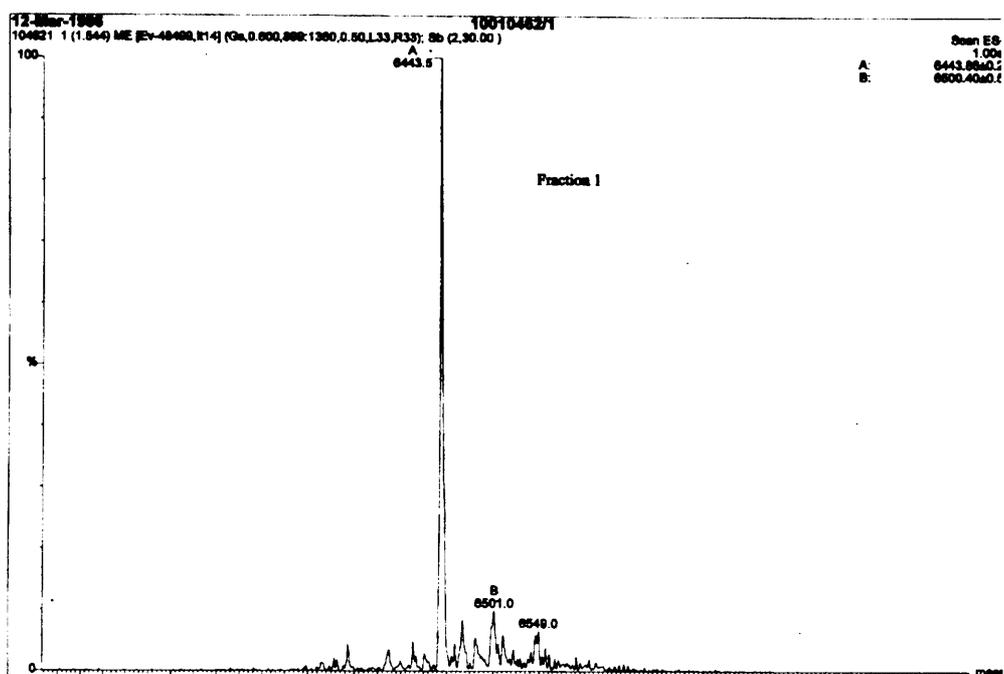


Figure 42. Deconvoluted electrospray mass spectrum of synthetic B2-domain after purification.

The LD-TOF spectrum corresponding to this peak is shown in Figure 43. The major peak marked “int.std” to the left of the product peak is an internal standard.

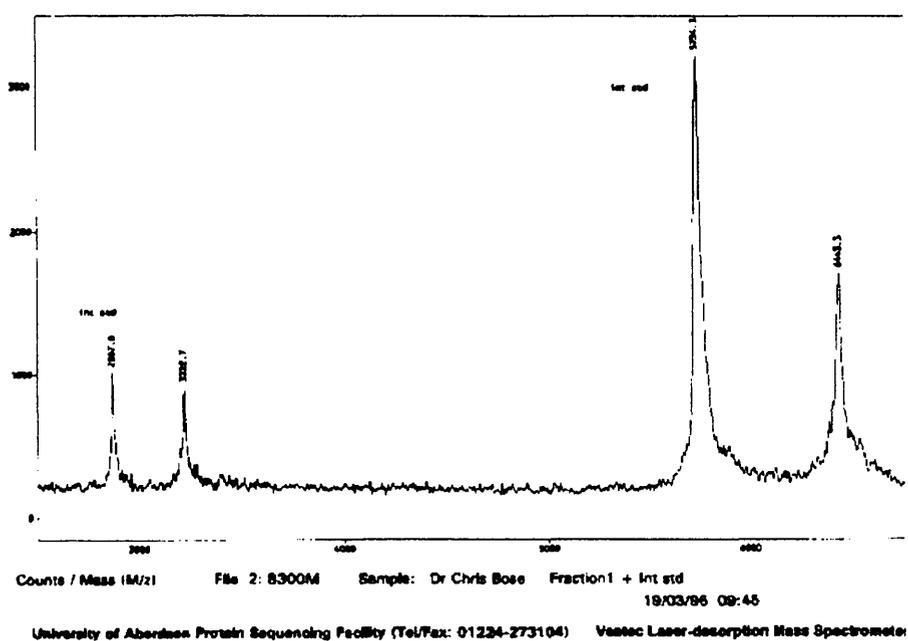


Figure 43. Laser Desorption-Time of Flight mass spectrum of synthetic B2-domain

The yield after purification was 29mg. The calculated molecular weight is 6444. The experimentally determined molecular weights were 6443.3 from laser time of flight, and 6444 from electrospray mass spectrometry.

Figure 134 shows the amide region of the NMR spectrum of wild type B2 domain from initial studies that compares well with the same region of the spectrum for recombinant material.

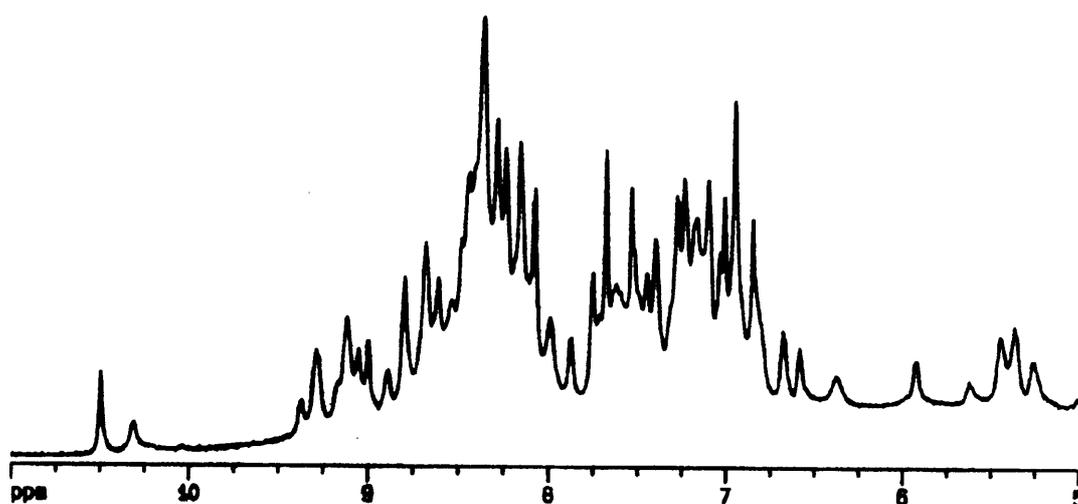


Figure 134. 500Mhz 1-D NMR spectrum of synthetic wild-type B2-domain.

When the spectra of recombinant material were compared in more detail, it appeared that the dispersion of signals in the amide region for the synthetic material was not as discrete as the signals from the recombinant material. Concerns were raised about the suitability of the refolding and purification procedure for synthetic protein G. It was suggested that the lack of dispersion could be explained by incomplete folding of the synthetic domain. Also, reverse phase HPLC had been used as the final step and it

was suggested that some denaturation could occur during this stage because of the presence of acetonitrile.

Because of these concerns, the wildtype domain was remade using O- (7-Azahydroxybenzotriaz-1-yl)-1,1,3,3, -tetramethyl uronium hexafluorophosphate (Figure 44), in place of HBTU to allow a direct comparison of the two reagents.

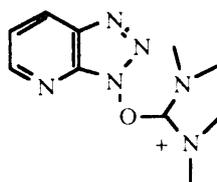


Figure 44. Coupling agent O- (7-Azahydroxybenzotriaz-1-yl)-1,1,3,3, -tetramethyl uronium hexafluorophosphate.

Refolding was accomplished by heating a solution of the domain, which had been desalted by reverse phase HPLC, in TRIS buffer at pH 7.6 to 75°C. The mixture was then cooled to 4°C rapidly and kept at this temperature throughout the purification process. The refolded material was purified on a Poros HQ/M anion exchange column. The eluted peaks were gel-filtered into water and lyophilised.

Figure 45 shows an analytical HPLC of the material.

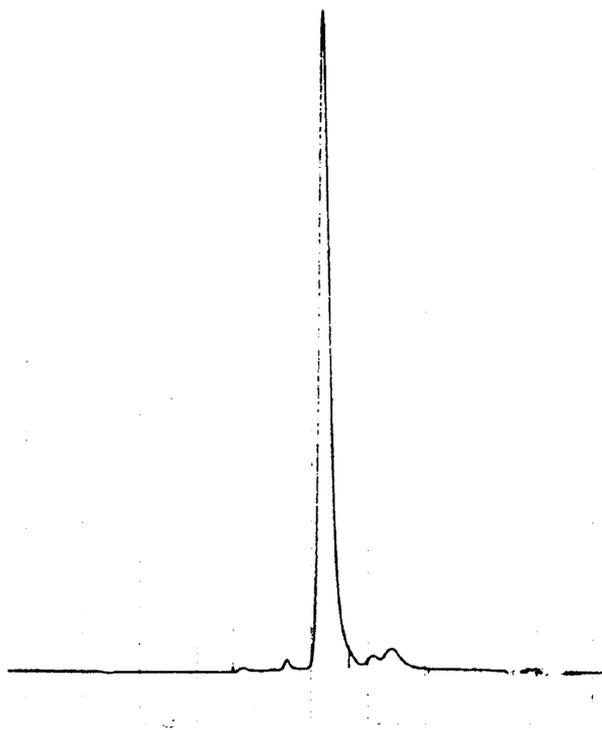


Figure 45. Analytical HPLC of HATU synthesized wildtype B2 domain

Figure 46 shows the electrospray mass spectrum of the peak in figure 45.

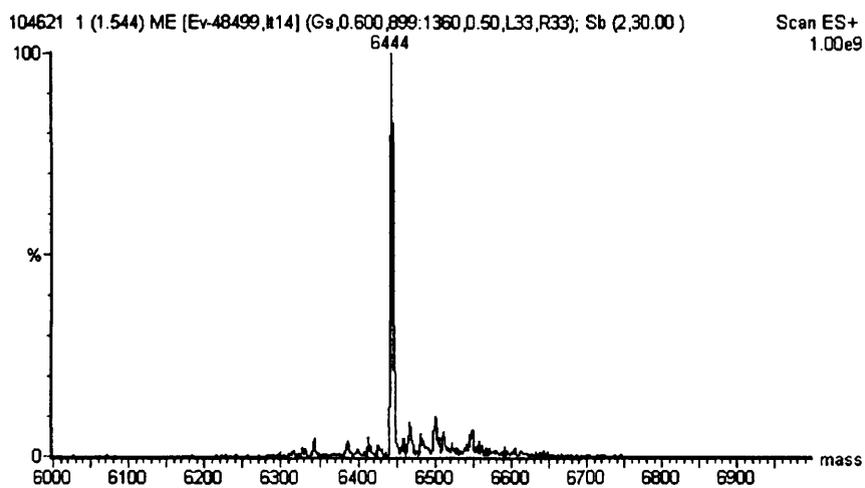


Figure 46. Deconvoluted electrospray mass spectrum of HATU synthesized B2-domain.

The yield was 33mg compared to the original synthesis result of 29mg. The new material was of satisfactory homogeneity, had been refolded by the same method as used for the recombinant material, and had avoided the use of reverse phase HPLC as the last step in the purification.

The 1-D NMR experiments were tried again and the result is shown in Figure 47.

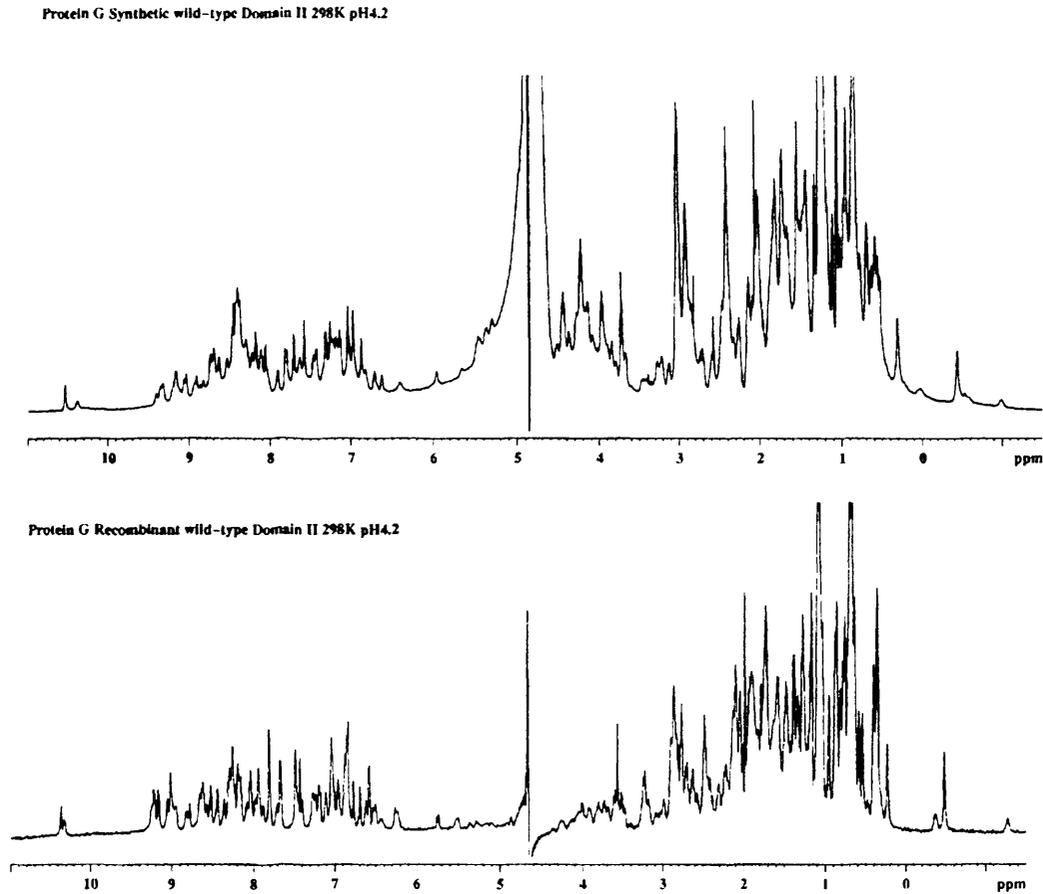


Figure 47. 1-D NMR spectra of synthesized B2 domain (6-64) and recombinant B2 domain (1-64)

This comparison suggests that there are overlapping signals in the amide region of the synthetic B2 domain. The synthetic samples appear to be homogeneous. One possible explanation is that although the N-terminal 5 residues have been shown to be freely moving in solution, they may be making a contribution to the stability of the domain.

Q37Aib

Q37Aib (6-64) was made using the same procedure for assembly and removal of protecting groups as the wildtype sequence with the exception that HATU was used as the coupling agent.

The crude material was purified by reverse phase HPLC as shown in Figure 48.

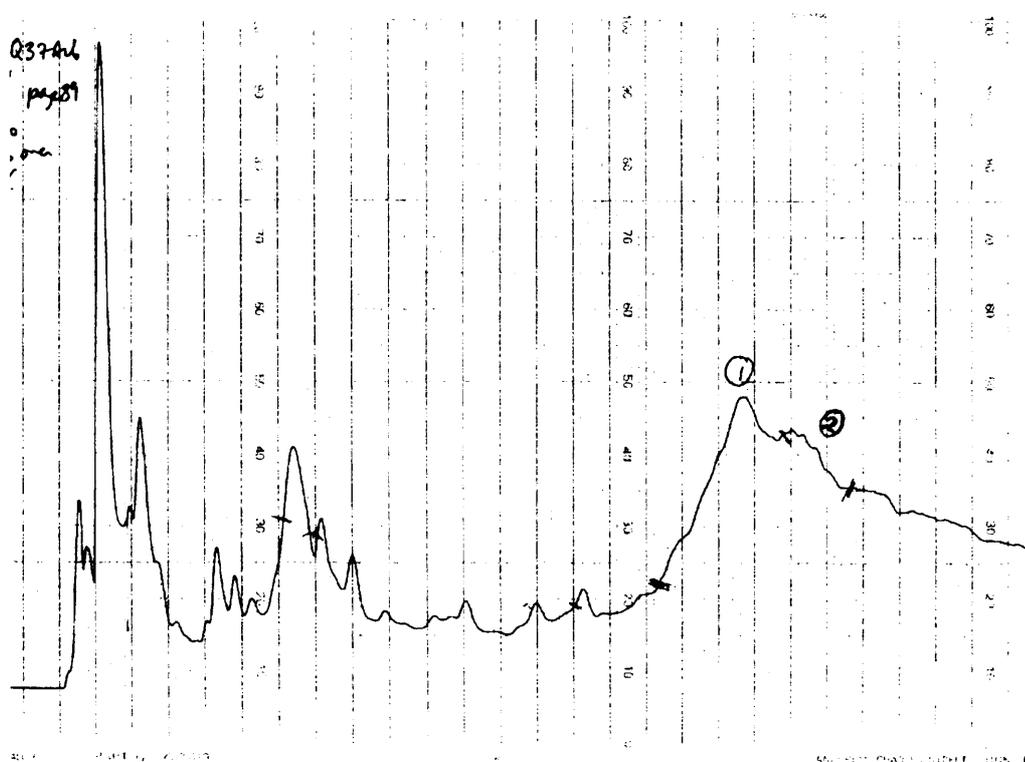


Figure 48. Hypersil C<sub>8</sub> column, 225 nm detection. 20 to 40% acetonitrile in 0.05% TFA over 80 minutes.

The material from fraction 1 was refolded using the method for the recombinant material as described in the previous section, and purified on a Poros anion exchange column.

The yield was 10mg of material with an ES-MS of 6400 (Figure 50). An analytical anion-exchange HPLC is shown in Figure 49.

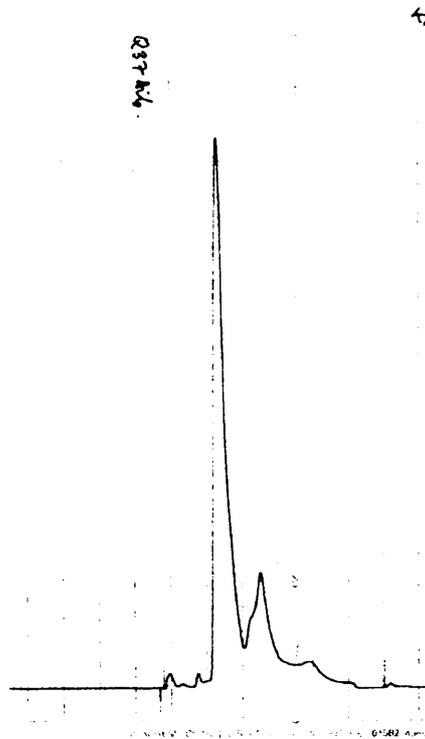


Figure 49. Analytical anion-exchange HPLC of Q37Aib (6-64). Poros HQ/M column, 225 nm detection. 0 to 30% 0.5M NaCl in 20mM TRIS, pH7.64 over 10 minutes.

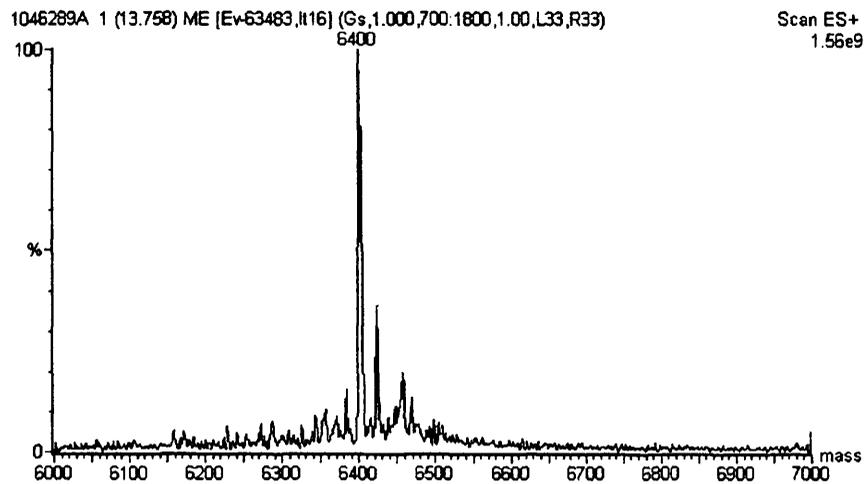


Figure 50. Deconvoluted electrospray mass spectrum of Q37A1b

The ability to incorporate Aib residues chemically into the middle of a protein domain is significant. It is one of the most hindered amino acids. To date, there have been no published chemical syntheses of domains containing Aib residues.

## K15T16BTD

Fmoc-BTD-OH was not commercially available at the time of this synthesis and was made from Benzyloxycarbonyl-glutamic acid [140] (Figure 51) as shown in the reaction sequence below:

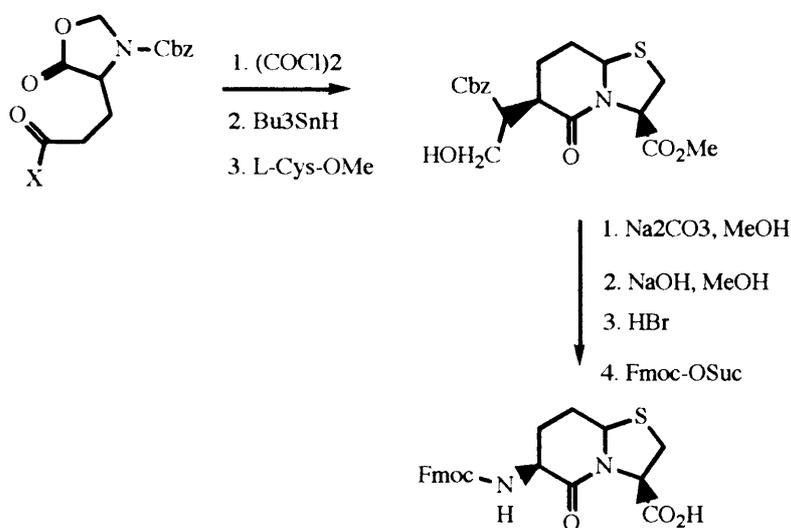


Figure 51. Synthesis of Fmoc-BTD-OH from Cbz-glutamic acid.

The synthesis gave 9.0g of material that was homogeneous on TLC. Nagai and Sato [141] first introduced the bicyclic dipeptide as a  $\beta$ -turn mimic. More recently Chalmers and Marshall [39] have suggested that its effect more closely mimics a reverse turn. In the following syntheses BTD is used as a rigid backbone substitute in a loop of the B2 domain molecule.

K15T16 BTD (6-64) was made using the same procedures for assembly and removal of protecting groups as the wildtype sequence. Refolding was done in an analogous way to the first wildtype synthesis using dilution from a 6M-guanidine solution

Figure 52 shows the reverse phase HPLC of K15T16 BTD B2 Domain. The yield was 27mg of homogeneous material.

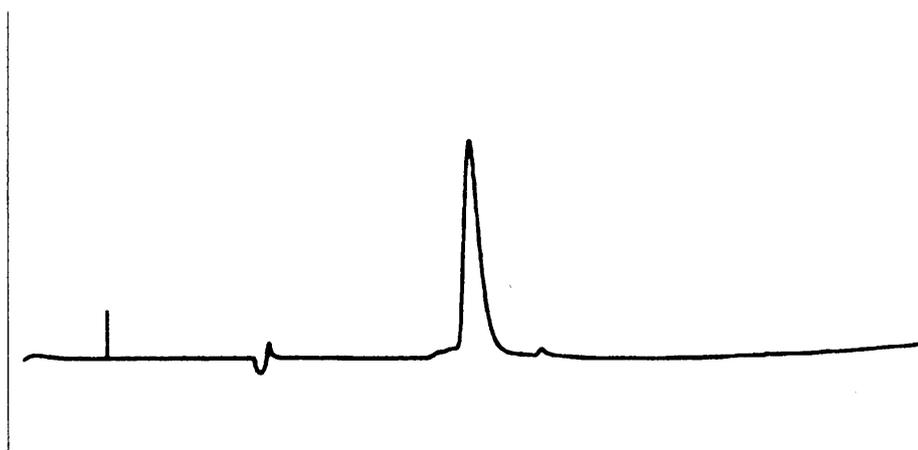


Figure 52. Analytical Reverse Phase HPLC of K15, T16 BTD Domain II. Hypersil C<sub>8</sub> column, 225 nm detection. 10 to 50% acetonitrile in 0.05% TFA.

The calculated molecular weight is 6412 and its corresponding ES-MS is 6412.

Figure 53 shows the mass spectrum.

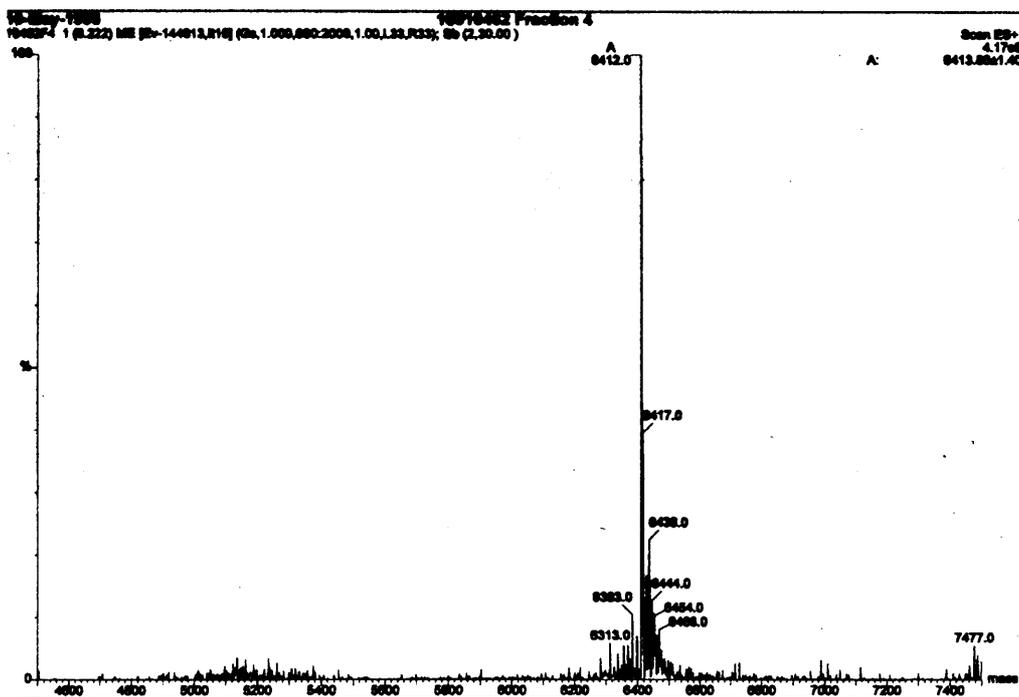


Figure 53. Deconvoluted electrospray mass spectrum of K15, T16 BTD Domain II

The NMR spectra of K15T16BTD showed the same type of inhomogeneity as the wildtype.

The material from the NMR tube was reanalysed on anion exchange HPLC and the result is shown in Figure 54.

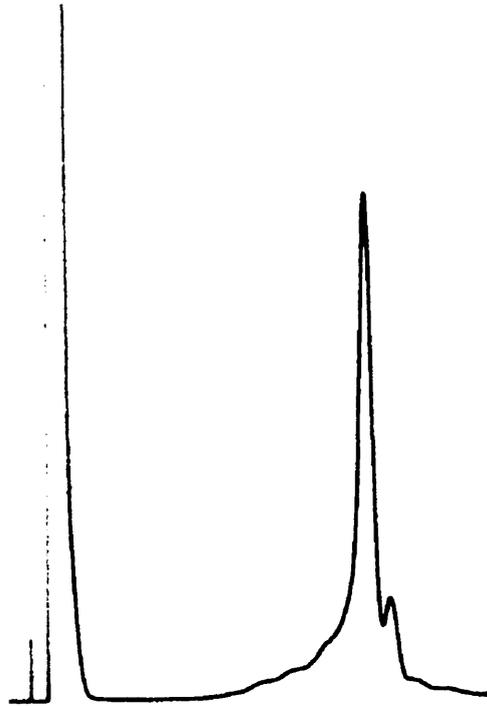


Figure 54. K15T16BTD diluted from 6M guanidine 20 fold into 20mM TRIS pH 7.6. Anion exchange on Resource Q. 0 to 15% 0.5M NaCl in 20mM TRIS pH 7.6 over 10 minutes.

K15T16BTD was resynthesized using the same method as the second synthesis of the wildtype sequence.

The analytical ion exchange HPLC is shown in Figure 55.

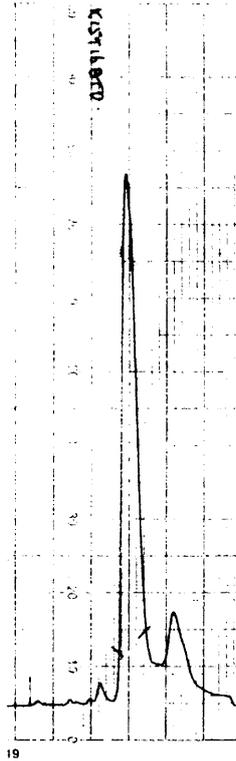


Figure 55. Analytical anion exchange HPLC of K15T16BTD. Poros HQ/M column. 10ml/min. 0 to 15% 0.5M NaCl in 20mM TRIS pH 7.4 over 10 minutes

Figure 56 shows the ES-MS. This preparation gave 20mg of material.

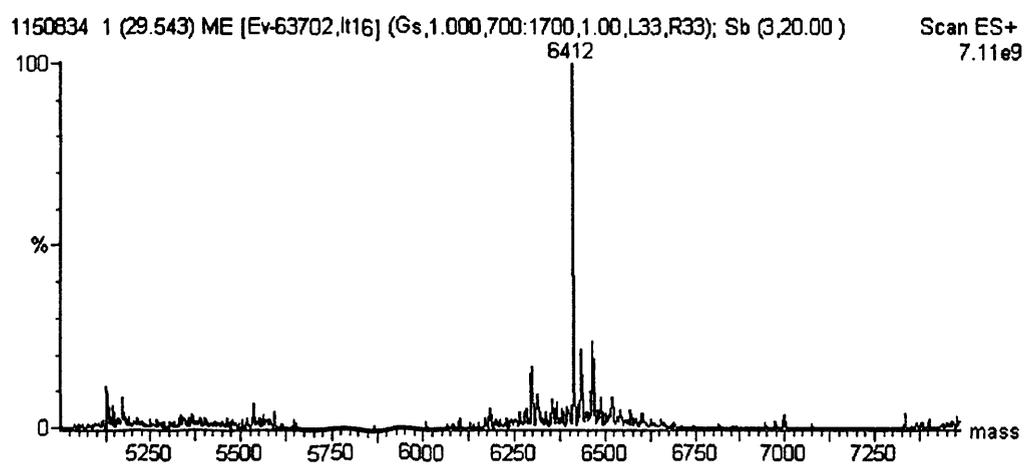


Figure 56. Deconvoluted electrospray mass spectrum of K15, T16 BTB B2 Domain.

1-D NMR spectra of newly refolded and recombinant B2 domain were run and Figure 57 shows the result.

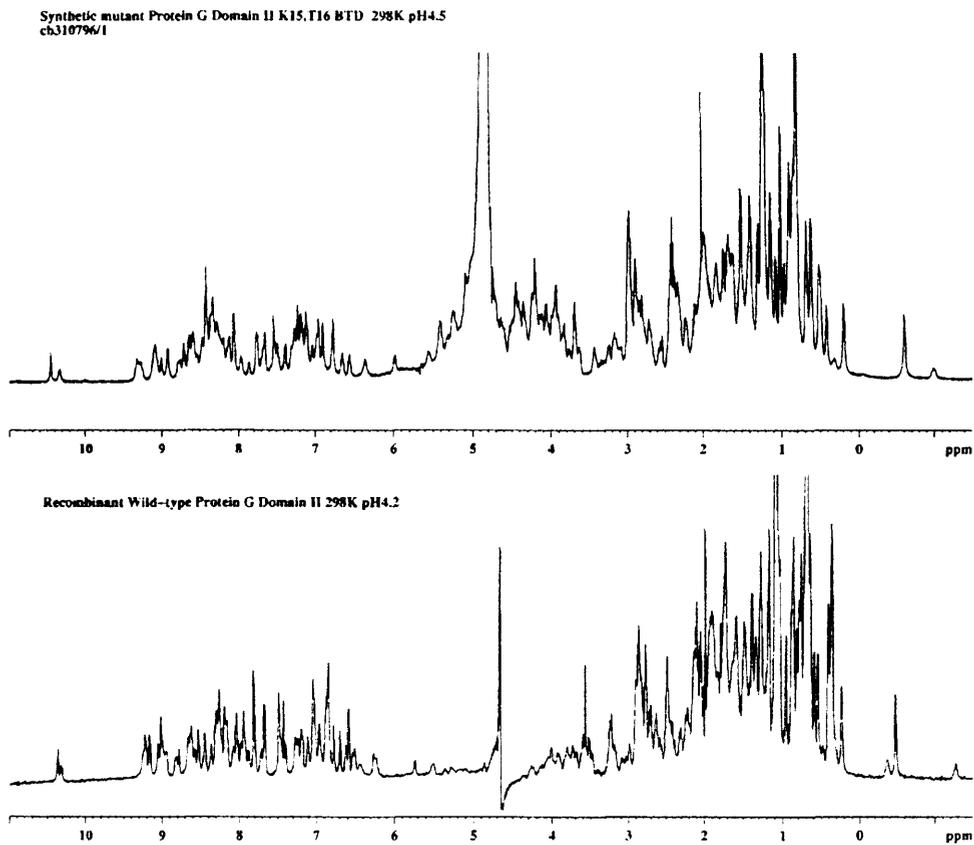


Figure 57. 1-D NMR spectra of synthesized K15T16BTD domain (6-64) and recombinant B2 domain (1-64)

As for the wildtype sequence, the amide signals are overlapping to a greater extent than those for the recombinant domain. This is shown more clearly in Figure 58, which is the fingerprint region of a NOESY spectrum of wildtype B2 domain superimposed on the equivalent spectrum of K15T16BTD. The signals for the BTD analogue are shown in red.

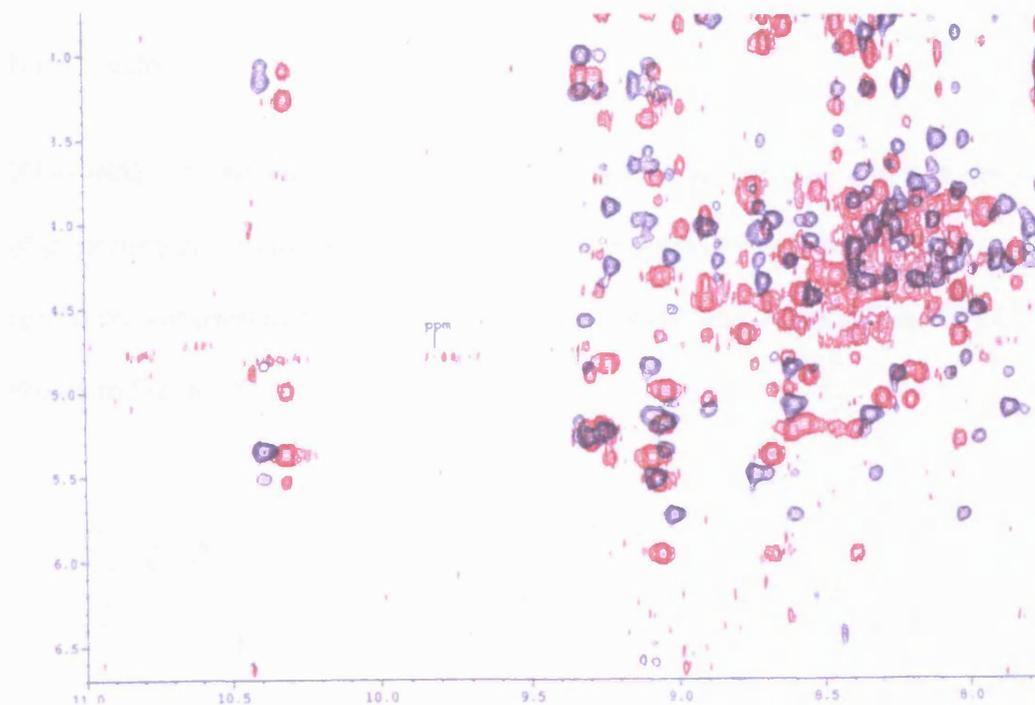


Figure 58. NOESY spectra. Wildtype signals are black; K15T16BTD signals are red.

There is certainly more overlap in the signals for the BTD analogue. The possibilities are that (a) the peptide is impure; (b) the preparation contains other forms of the domain; (c) the analogue is degrading under the conditions; or (d) overlapping signals are normal for proteins containing BTD and a workaround has to be found.

## K15T16Ahx

K15T16Ahx (6-64) was made using the same procedures for assembly and removal of protecting groups as the wildtype sequence. Refolding was done using the method for the recombinant material. The isolated yield was 35mg. An analytical HPLC is shown in Figure 59.

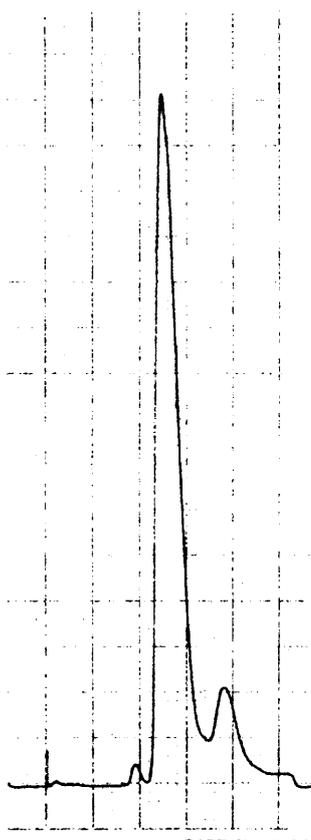


Figure 59. Analytical anion-exchange HPLC of K15T16Ahx (6-64). Poros HQ/M column, 225 nm detection. 0 to 30% 0.5M NaCl in 20mM TRIS, pH7.64 over 10 minutes.

The molecular weight of the analogue is calculated to be 6328 and ES-MS shows an experimental molecular weight of 6326 (Figure 60).

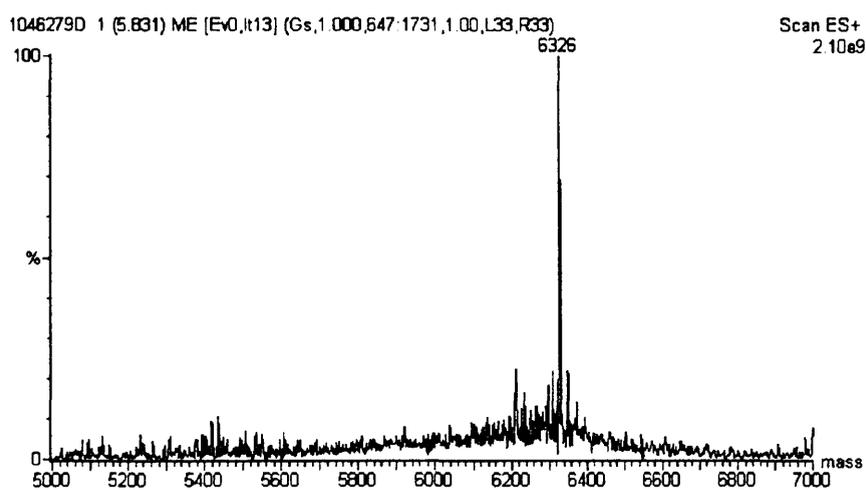


Figure 60. Deconvoluted electrospray mass spectrum of K15T16Ahx (6-64)

## G14K15BTD

G14K15 BTD (1-64) was made using the same procedures for assembly and removal of protecting groups as the wildtype sequence. After precipitation with ether, the peptide was dissolved in 6M guanidine, 0.1M TRIS, 2 mM EDTA pH 7.2 to break up any aggregated and misfolded material. The solutions were then gel-filtered into 20 mM TRIS pH 7.5 for refolding, but gave a cloudy mixture, which was immediately diluted to 1L of 20 mM TRIS pH 7.5. The mixture was left to stand in a cold room at 4°C for seven days.

Anion exchange and reverse phase HPLC purified the crude peptide. Figure 61 shows an analytical HPLC of G14K15BTD (1-64).

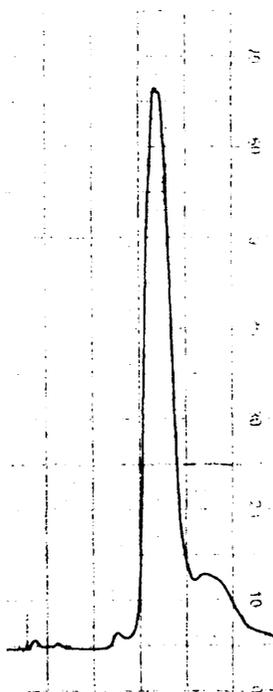


Figure 61. Analytical anion-exchange HPLC of G14K15BTD (1-64). Poros HQ/M column, 225 nm detection. 0 to 30% 0.5M NaCl in 20mM TRIS, pH7.64 over 10 minutes.

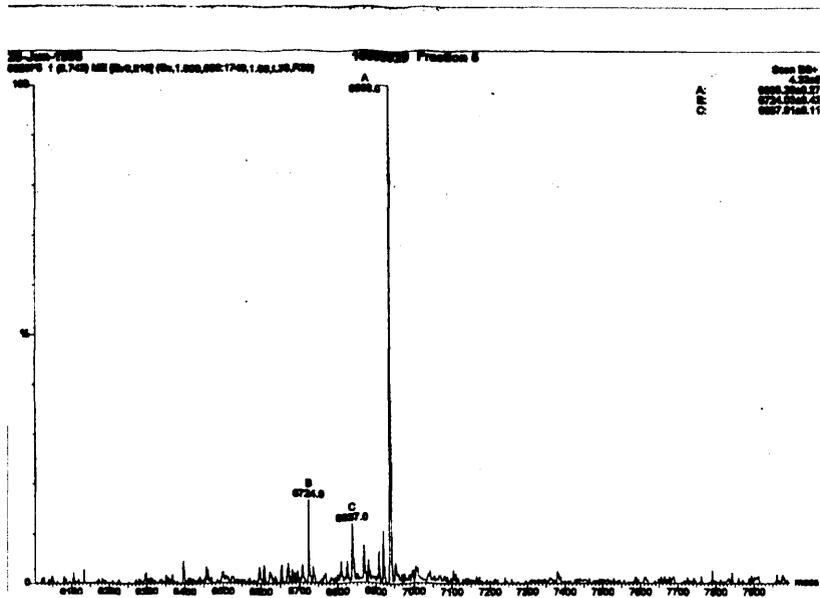


Figure 62. Electrospray mass spectrum of G14K15 BTD (1-64)

Figure 62 is an electrospray mass spectrum of the same material.

## Biotin 0, G14K15BTD

This domain was synthesized in exactly the same way as the previous molecule because at the end of the synthesis of G14K15 BTD (1-64) the resin was split into two. The first half was deprotected and became G14K15 BTD (1-64). The second half was biotinylated using biotin and HBTU activation. Figure 63 below shows the reverse phase HPLC of Biotin 0, G14, and K15 BTD B2 Domain.



Figure 63. Analytical reverse phase HPLC of Biotin 0, G14, K15 BTD Domain II. Hypersil C<sub>8</sub> column, 225 nm detection. 10 to 50% acetonitrile in 0.05% TFA.

Figure 64 shows the electrospray mass spectrum of Biotin 0, G14, K15 BTD.

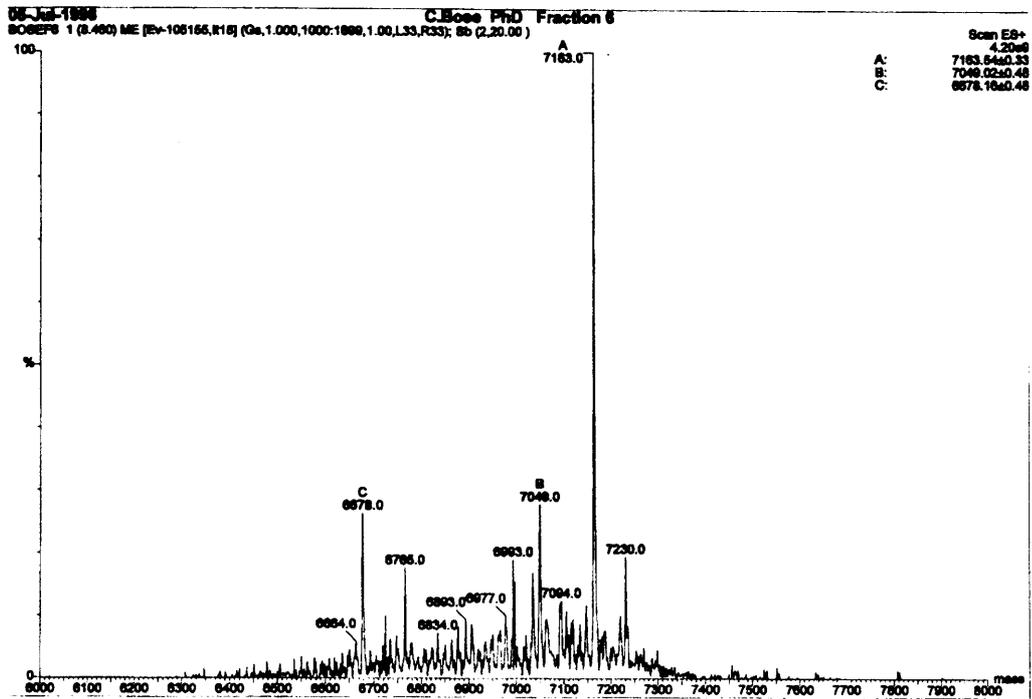


Figure 64. Deconvoluted electrospray mass spectrum of Biotin 0, G14, K15 BTD

## K15dP T16N-MeA

K15dPT16N-MeA (6-64) was made using exactly the same procedure as used for Q37Aib. The final product was lyophilised to yield 20mg of solid. Figure 65 shows the analytical HPLC of the purified material.

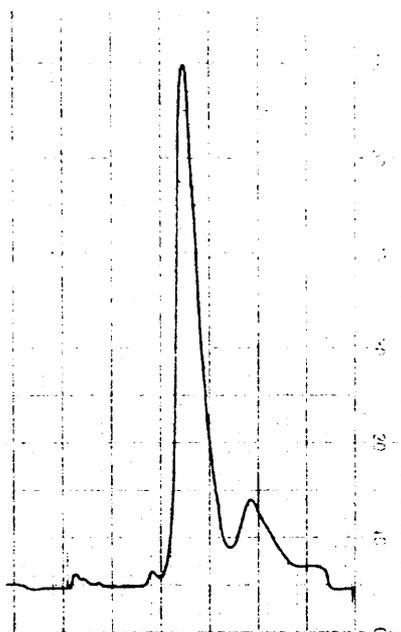


Figure 65. Analytical anion-exchange HPLC of K15dPT16N-MeA (6-64). Poros HQ/M column, 225 nm detection. 0 to 30% 0.5M NaCl in 20mM TRIS, pH7.64 over 10 minutes.

The calculated molecular weight is 6397 and the ES-MS shows a mass of 6396 as shown in Figure 66.

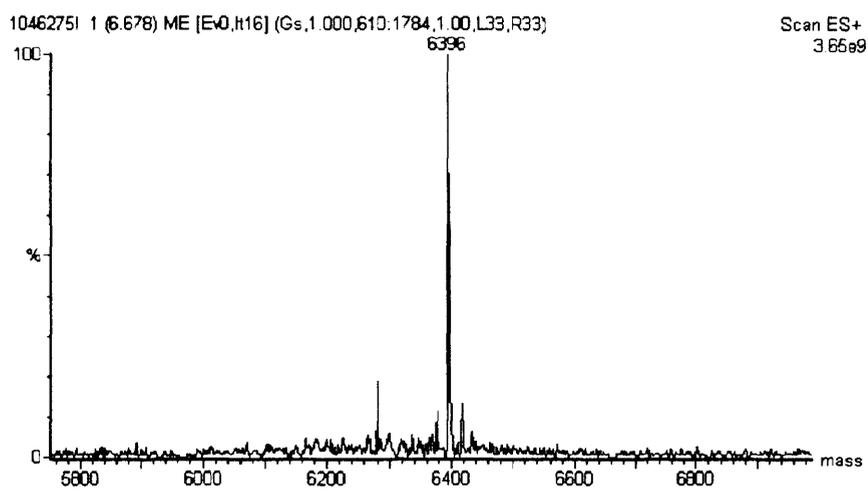


Figure 66. Deconvoluted electrospray mass spectrum of K15dP, T16N-MeA (6-64)

## G14K15Ahx

G14K15Ahx (6-64) was made using the same procedures for assembly and removal of protecting groups as the wildtype sequence. Refolding was done using the method for the recombinant material. The yield was 33mg. An analytical HPLC is shown in Figure 67.

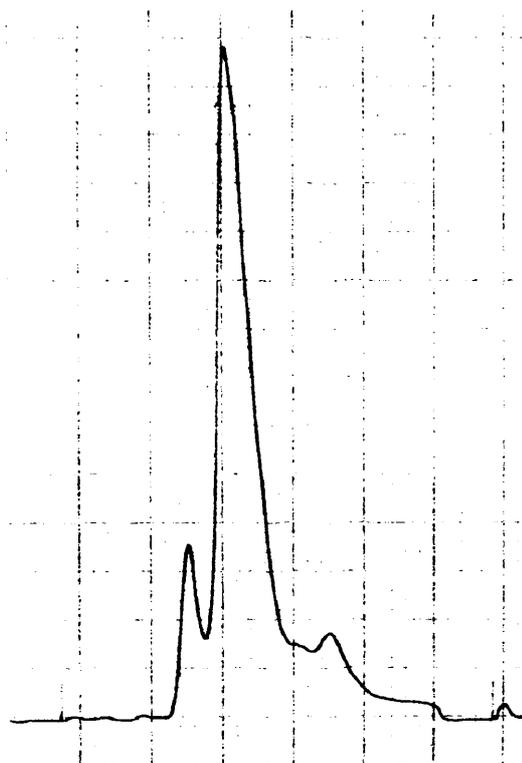


Figure 67. Analytical anion-exchange HPLC of G14K15Ahx (6-64). Poros HQ/M column, 225 nm detection. 0 to 30% 0.5M NaCl in 20mM TRIS, pH7.64 over 10 minutes.

The calculated molecular weight is 6372. The ES-MS shows a molecular weight of 6371 as shown in Figure 68.

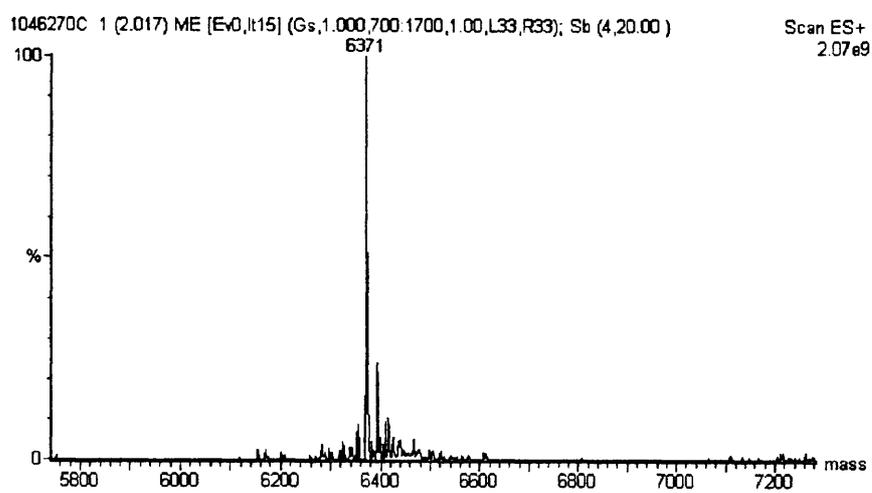


Figure 68. Deconvoluted electrospray mass spectrum of G14K15Ahx (6-64)

## G14K15BTD

G14K15BTD (6-64) was made using the same procedure for assembly, removal of protecting groups and refolding as Q37Aib. The final product was lyophilised to yield 21mg. An analytical HPLC is shown in Figure 69.

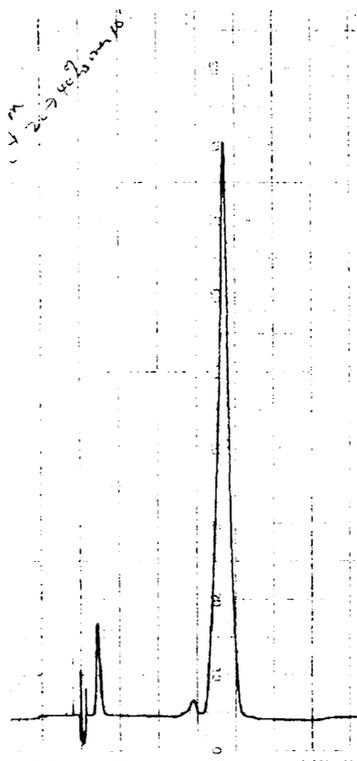


Figure 69. Analytical anion exchange HPLC of G14K15BTD B2 Domain.

The calculated molecular weight is 6457. Figure 70 shows the experimentally determined molecular weight of 6456.

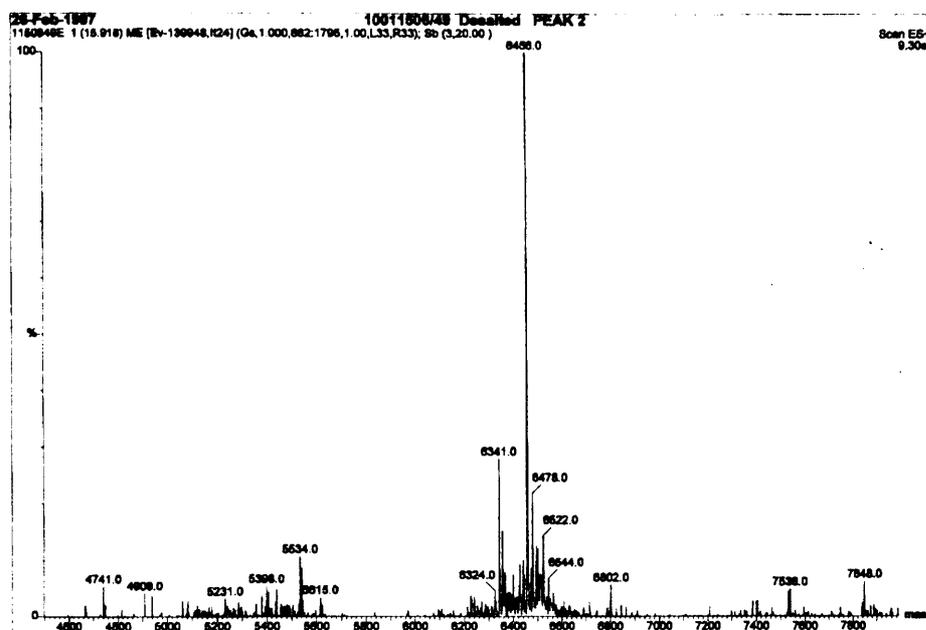


Figure 70. Deconvoluted mass spectrum of G14K15BTD (6-64)

15BTD16

15BTD16 (6-64) was made using the same procedure for assembly, removal of protecting groups and refolding as Q37Aib. The final product was lyophilised to yield 14mg of homogeneous material. An analytical HPLC is shown in Figure 71.

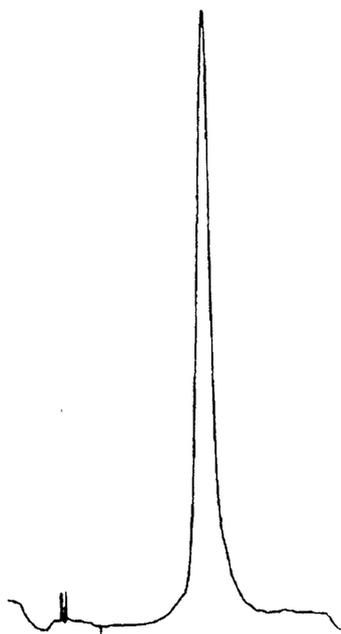


Figure 71. Analytical ion exchange HPLC. Poros HQ/M column. 10ml/min. 0 to 15% 0.5M NaCl in 20mM TRIS pH 7.4 over 10 minutes.

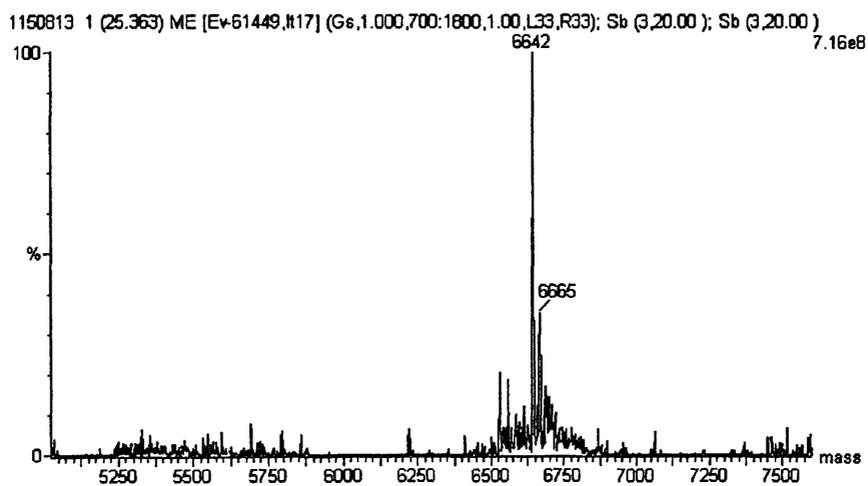


Figure 72. Deconvoluted electrospray mass spectrum of 15BTD16

The calculated molecular weight is 6641. The ES-MS shows a molecular weight of 6642 as shown in Figure 72.

## 15Ahx16

15Ahx16 (6-64) was made using the same procedure for assembly, removal of protecting groups and refolding as Q37Aib. The final product was lyophilised to yield 7mg. An analytical HPLC is shown in Figure 73 and ES-MS in Figure 74.

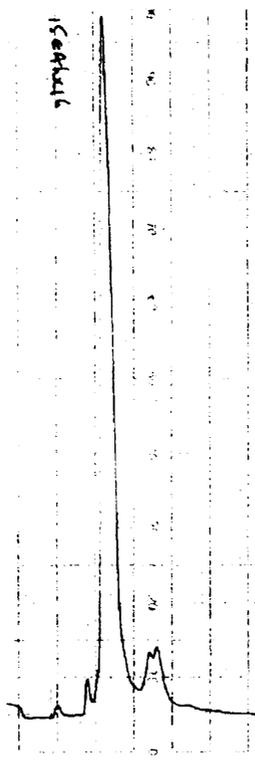


Figure 73. Analytical anion exchange HPLC of 15Ahx16 Poros HQ/M column. 10ml/min. 0 to 15% 0.5M NaCl in 20mM TRIS pH 7.4 over 10 minutes.

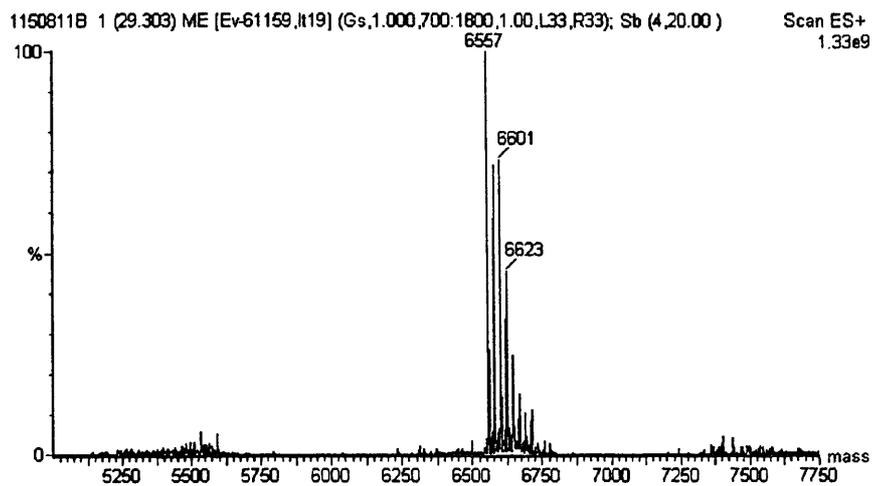


Figure 74. Deconvoluted electrospray mass spectrum of 15Ahx16

The multiple peaks in the spectrum can all be assigned to sodium adducts occurring at  $M+22n$  where  $n$  is an integer.

## Synthesis Discussion

Ten domains and the wildtype sequence were successfully made by solid phase peptide synthesis. Analogues containing hindered residues were made using an analogue of the standard coupling reagent. HATU (Figure 75) is the 7-azahydroxybenzotriazole analogue of HBTU, whose structure by analogy with HBTU is shown.

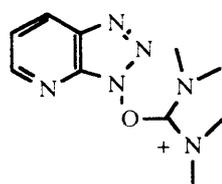


Figure 75. Peptide coupling agent HATU.

X-ray studies have shown that the crystalline structure is actually the N-oxide guanidinium structure (Figure 76).

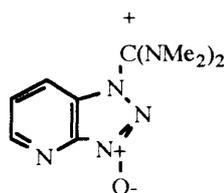


Figure 76. X-ray structure of HATU.

Many studies have shown its efficacy in coupling hindered residues in peptides, but to date there had been no applications to large peptides containing hindered residues. Its advantages were that it was six to thirty-two times more reactive than HBTU, that it lessens loss of chiral integrity, and that it provides a visual endpoint (yellow to colourless) of the reaction endpoint. Table 1 summarises the synthetic work.

Table 1

Peptide	Coupling Agent	Calc. Mass	Exp. Mass	Yield/mg
WildType	HBTU	6444	6443	29
K15T16BTD	HBTU	6412	6412	27
K15T16Ahx	HATU	6328	6326	35
G14K15BTD	HBTU	6457	6456	21
G14K15Ahx	HATU	6372	6371	33
15BTD16	HATU	6641	6642	14
15Ahx16	HATU	6555	6557	7
Q37Aib	HATU	6400	6400	30
K15dP,T16NMeA	HATU	6397	6395	20
Biotin 0G14K15BTD(1-64)	HBTU	7162	7163	N/D
G14K15BTD(1-64)	HBTU	6938	6938	N/D

HATU was used successfully to make Q37Aib and K15dPT16NMeA analogues.

These types of protein G analogues have not been made before. Their synthesis demonstrates that it is now possible to introduce any hindered amino acid type entity into any position in a protein or peptide up to 64 residues in length.

### 3.3 Guanidine unfolding

Guanidine denaturation curves for wildtype protein G B2 domain and eight analogues were determined by monitoring the change in fluorescence of tryptophan and tyrosine residues in the peptides at pH 7.6 and 4.7. Guanidine was used as denaturant because urea solutions are chemically unstable and they would have to be freshly made up for each new experiment. The analogues can be divided into four groups, characterized by the position and nature of the introduced changes. The first group contains the insertion analogues, 15BTD16 and 15Ahx16. The second contains those with changes in the loop region at residues 14 to 16. The third contains a change in the helical region. The guanidine unfolding was carried out at two pH's because the NMR work required a pH of 4.7, while the anion-exchange HPLC purification required a pH of 7.6.

Analysis of the denaturation curve was performed using the two-state unfolding model and the linear extrapolation method. The free energy of unfolding  $\Delta G_{obs}$  can be calculated as a function of guanidine concentration and temperature by assuming a two-state mechanism and using the equation

$$\Delta G_{obs} = -RT \ln K_{obs} = -RT \ln \left( \frac{F_n - F}{F - F_u} \right)$$

$K_{\text{obs}}$  is the equilibrium constant,  $F$  is the observed fluorescence, and  $F_n$  and  $F_u$  are the fluorescence values characteristic of the folded and unfolded conformations of the protein, respectively. The simplest method of estimating the conformational stability in the absence of denaturant,  $\Delta G_{\text{water}}$ , is to note that  $\Delta G$  varies linearly with denaturant concentration in the region where  $\Delta G$  can be measured. This is represented by the equation shown below

$$\Delta G_{\text{obs}} = \Delta G_{\text{water}} - m[D]$$

where  $m$  is a measure of the dependence of  $\Delta G$  on guanidine concentration and  $[D]$  is the concentration of the denaturant.

At the midpoint of the curve  $\Delta G_{\text{obs}} = 0$  and  $\Delta G_{\text{water}} = m[D_{1/2}]$  where  $[D_{1/2}]$  is the denaturant concentration at the midpoint of the curve.

$$\text{So, } \Delta G_{\text{obs}} = m[D_{1/2}] - m[D]$$

Assuming a linear relationship between the observed fluorescence and denaturant concentration the following relationships are derived

$$F_n = F_n(0) + m_n [D] \text{ and } F_u = F_u(0) + m_u [D].$$

Simple algebraic manipulation combines the equations for the two-state mechanism and the linear extrapolation method into a single equation to describe the shape of the denaturation curve:

$$F = \frac{F_n(0) + m_n[D] + (F_u(0) + m_u[D] \exp \frac{m([D] - [D_{1/2}]}{RT})}{1 + \exp \frac{m([D] - [D_{1/2}]}{RT}}$$

where F is the observed fluorescence (after allowing for the fluorescence of guanidine and buffer),  $F_n$  and  $m_n$  are the intercept and slope, respectively, of the pre-transition baseline,  $F_u$  and  $m_u$  are the intercept and slope of the post-transition baseline. Using this equation fits all the experimental parameters simultaneously to the model and so gives a better estimate of the errors in the experiment. The observed fluorescence and denaturant concentration were fitted to the above equation using the general non-linear least squares analysis in the program Kaleidagraph™. The free energy of unfolding (defined as the conformational stability of the protein)  $\Delta G_{\text{water}}$  is equal to  $m[D_{1/2}]$ . For this work,  $\Delta\Delta G$ , the difference in conformational stability is defined in the equation shown below.

$$\Delta(\Delta G) = \Delta G (\text{mut}, \text{H}_2\text{O}, 298\text{K}) - \Delta G (\text{wt}, \text{H}_2\text{O}, 298\text{K})$$

A positive  $\Delta(\Delta G)$  indicates a more stable analogue.  $\Delta\Delta G$  is then found from the relation

$$\Delta\Delta G = [D_{1/2}(\text{mut}) - D_{1/2}(\text{wt})][m(\text{wt}) + m(\text{mut})]/2$$

By calculating  $\Delta\Delta G$  from the difference in midpoints, a long extrapolation back to 0M guanidine is avoided.

The overlay plots are shown to compare the fitted curves to wildtype and the derived values for the slope of the denaturant curve and  $D_{1/2}$ , the midpoint of the denaturation curve and the  $\Delta\Delta G$  values are tabulated.

### pH 7.6

The data for each domain are presented individually to emphasize the raw data. All the raw data for the pH 7.6 results were collected with an excitation wavelength of 280 nm and an emission wavelength of 350 nm. Figures 77 to 85 show the raw data for wildtype, 15BTD16, 15Ahx16, G14K15BTD, G14K15Ahx, K15T16BTD, K15T16Ahx, K15dP,T16N-MeA and Q37Aib.

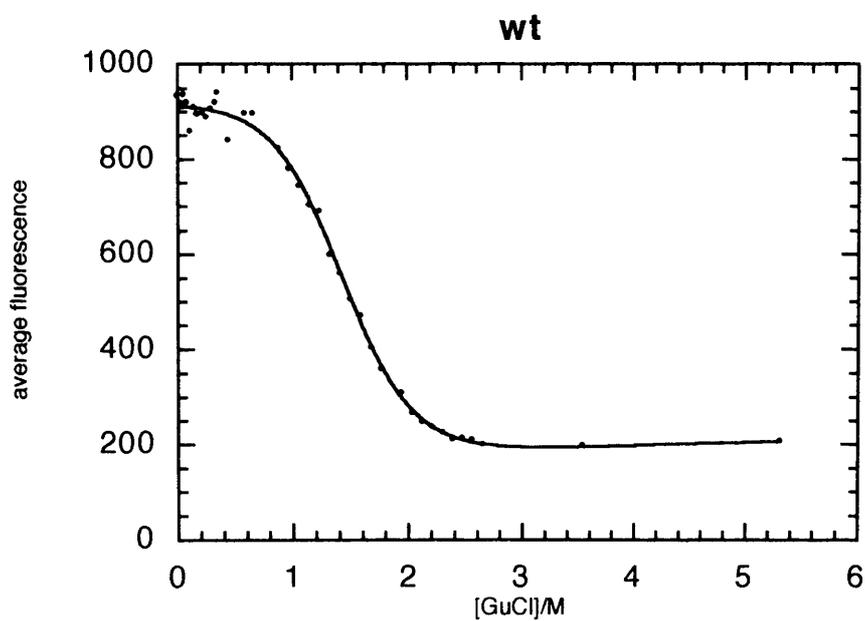


Figure 77.

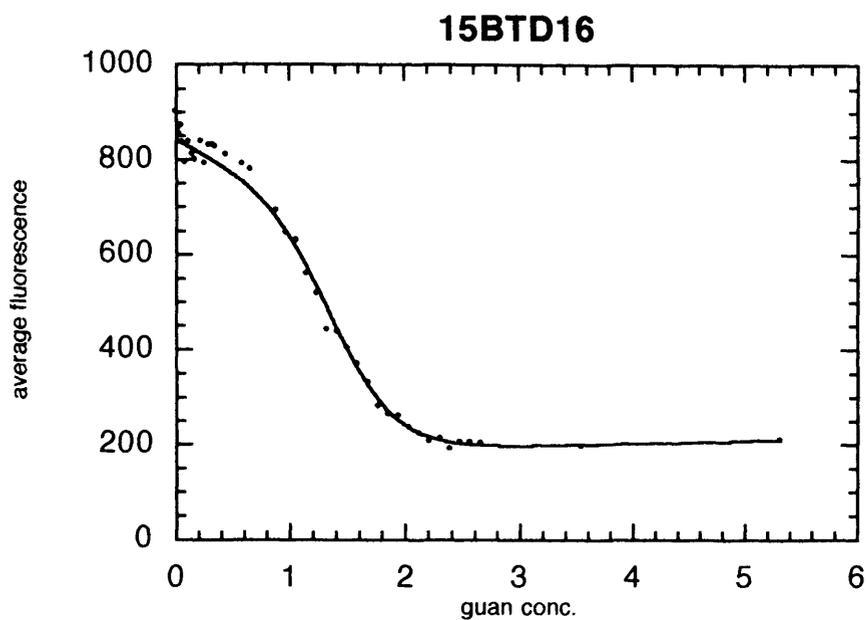


Figure 78.

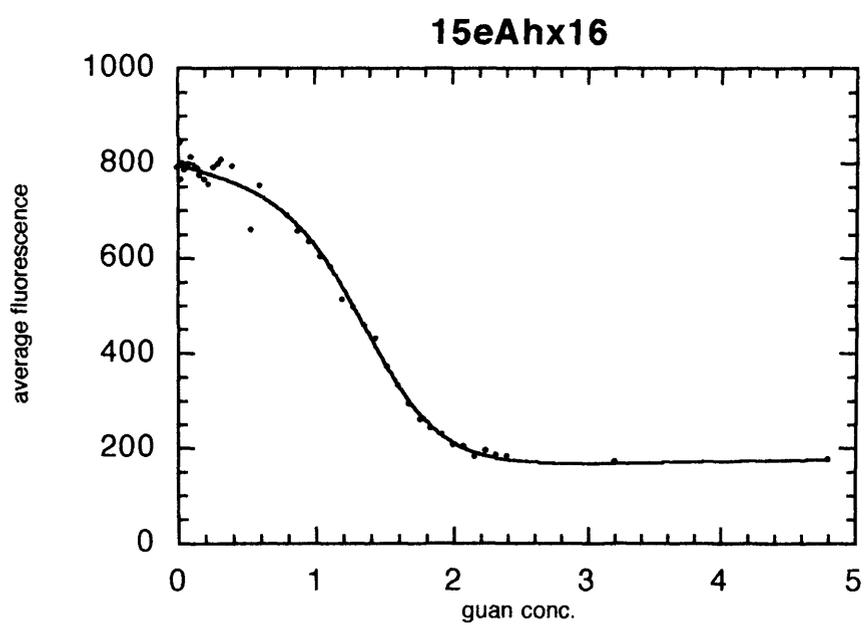


Figure 79.

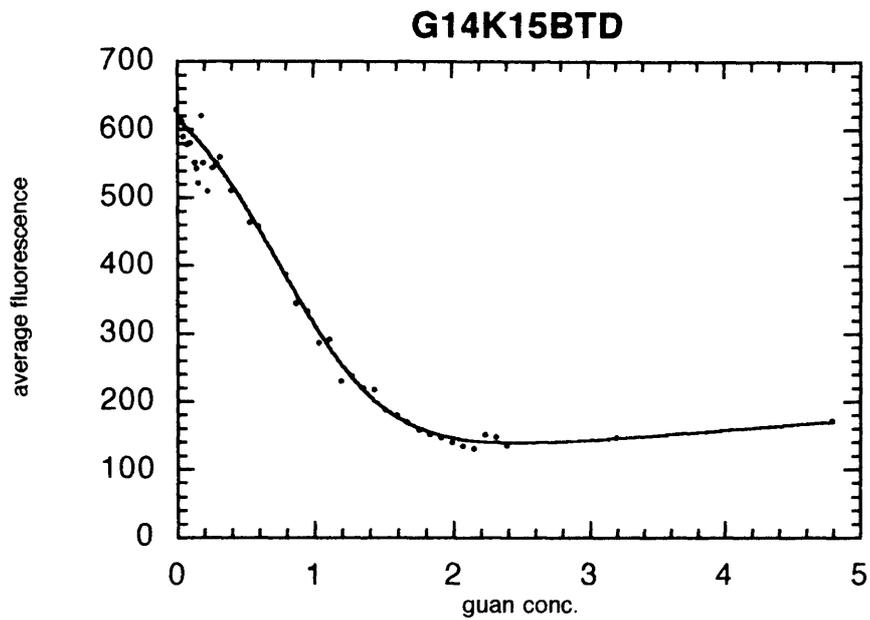


Figure 80.

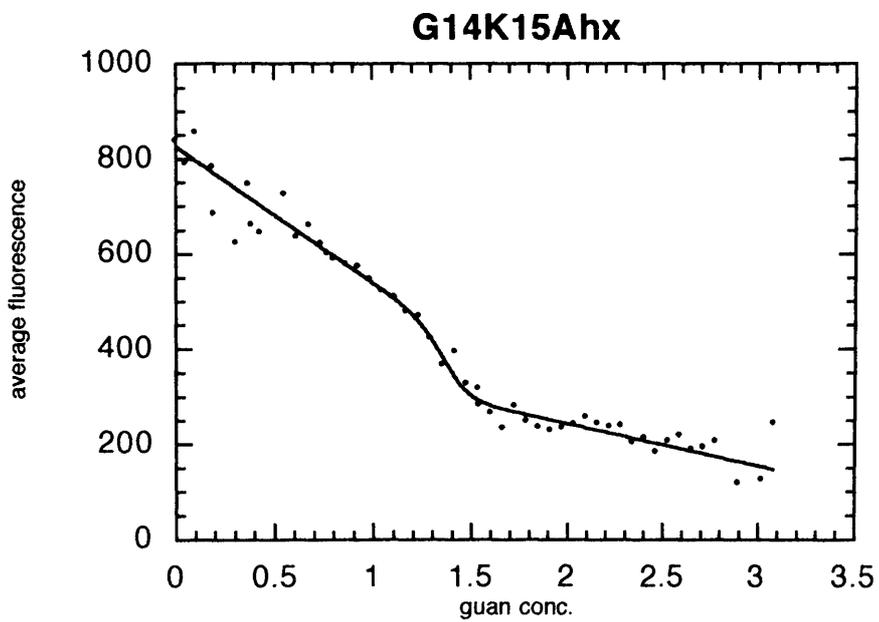


Figure 81.

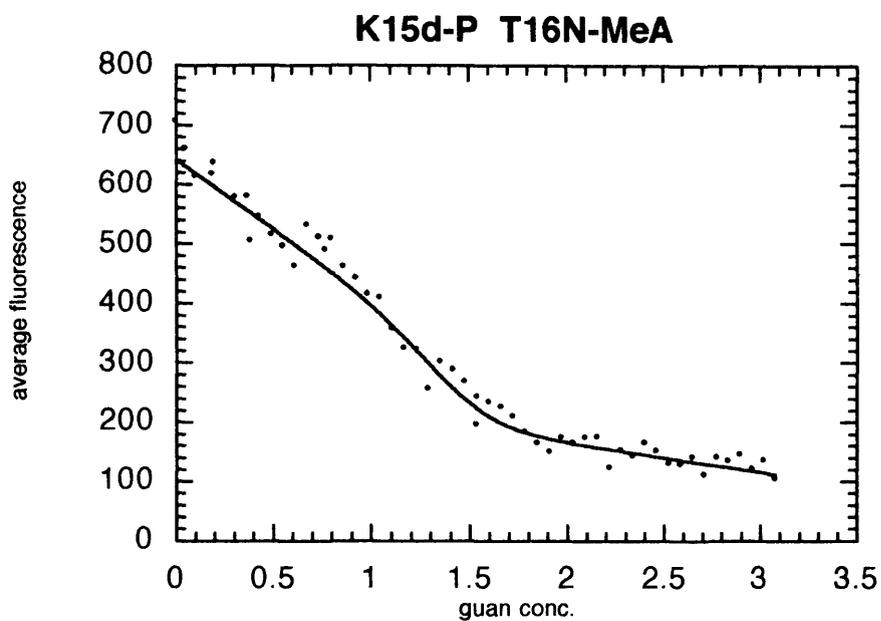


Figure 82.

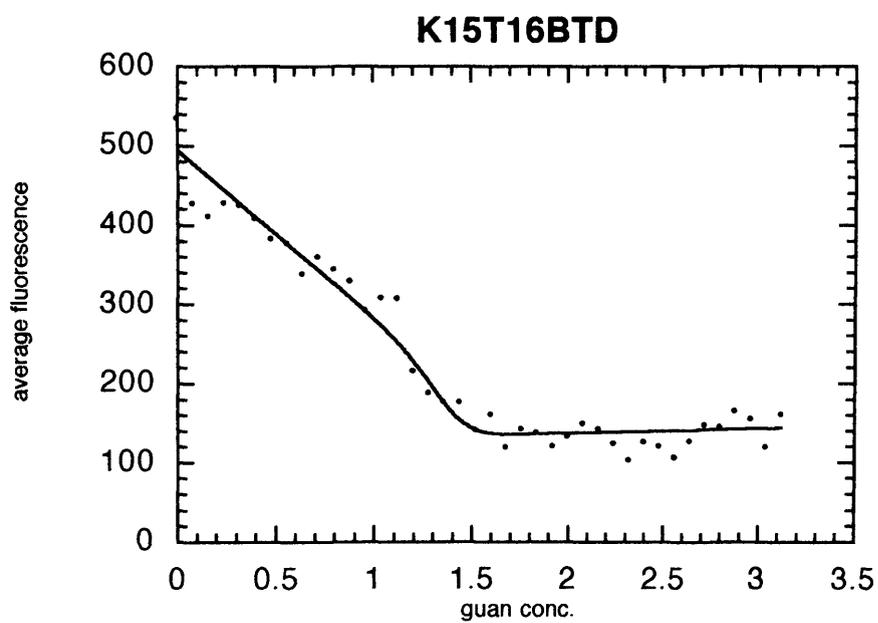


Figure 83.

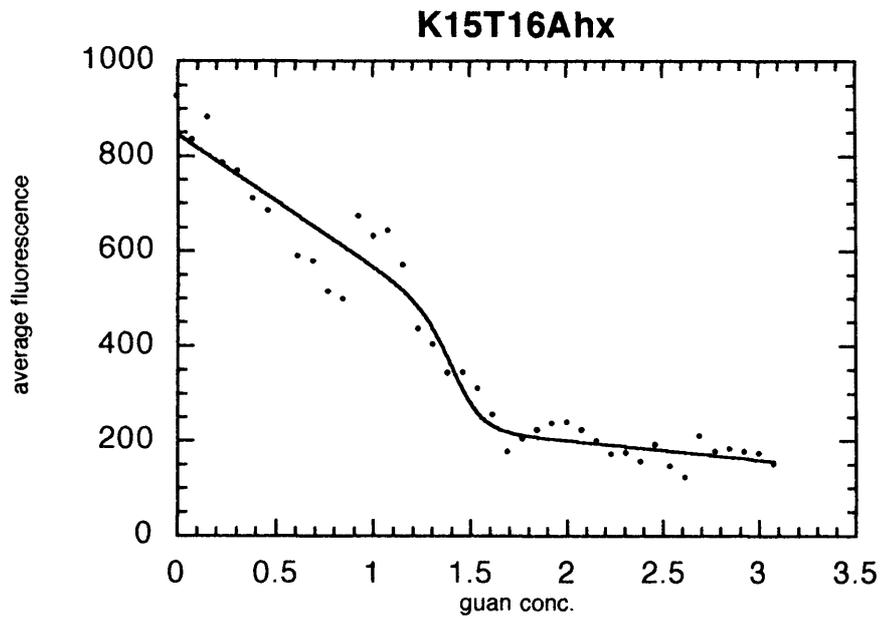


Figure 84.

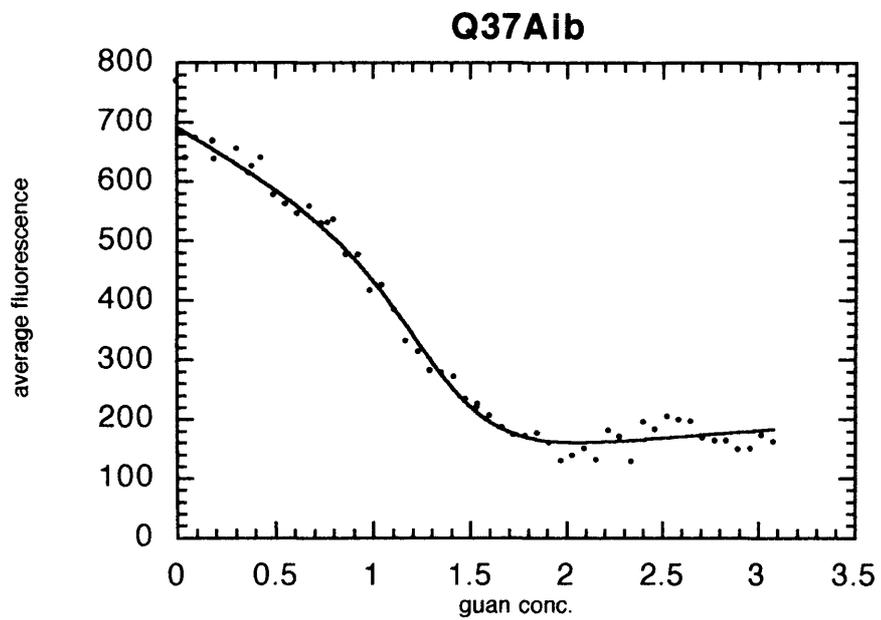


Figure 85.

Figure 86 shows the overlay plots of the fitted curves of 15BTD16, 15Ahx16 and wt.

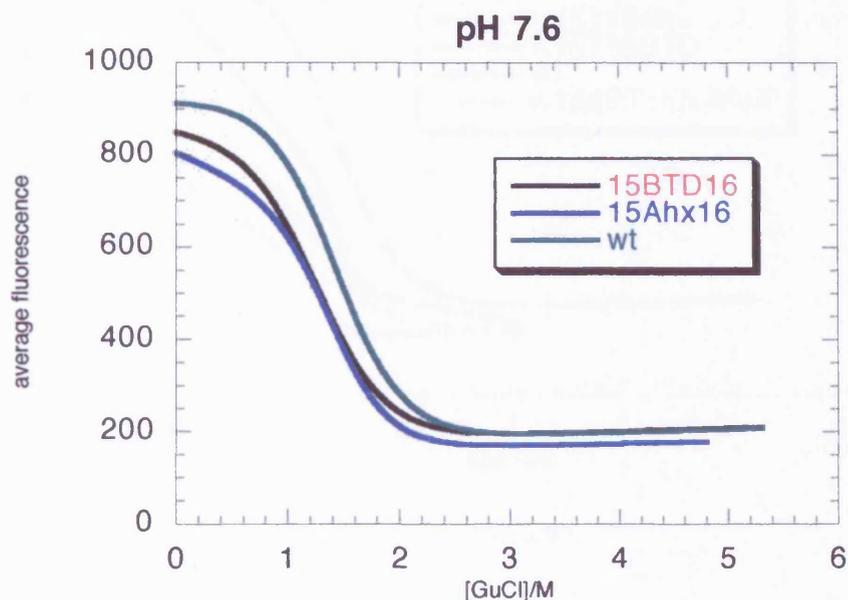


Figure 86.

The two 15-16 insertion analogues appear to destabilize the domain slightly at pH 7.6. The  $m$  values for the three do not differ significantly (see Table 2) suggesting that there is little difference in the folded and unfolded states of the three proteins. The loop appears to be flexible enough to adjust to the rigid dipeptide mimetic as well as the inherently flexible carbon chain. In this case, it is reasonable to calculate  $\Delta\Delta G$  values because the difference in stability in the absence of denaturant is the same as the stability difference in the transition region.

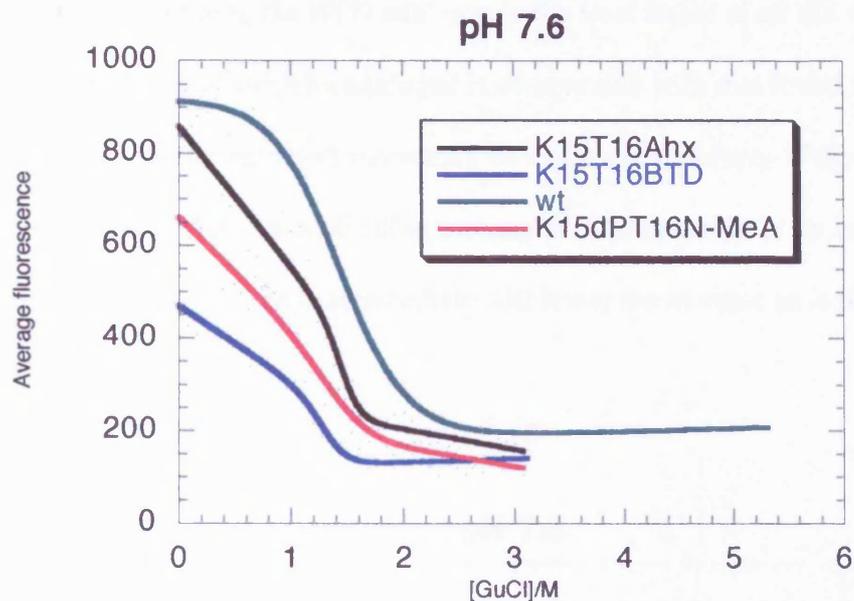


Figure 87.

Figure 87 shows the overlay plot for K15T16Ahx, K15T16BTD, K15dPT16N-MeA, and wildtype. The K15T16 analogues all have substantially higher  $m$  values from the wildtype protein. This indicates that something interesting is happening with these analogues at pH 7.6. One possibility is that there is a significant change in the conformation of the unfolded state that changes the accessibility of the state to the denaturant. Shortle has suggested that larger values of  $m$  over wildtype in staphylococcal nuclease variants reflect more extended and solvent accessible conformations. The analogues also begin to denature at very small denaturant concentrations. In these cases, it may be unreasonable to assign any significance to the  $\Delta\Delta G$  values, because these values depend on there being a perturbation of the wildtype state by the substitution. If the analogue behaves significantly differently then differences in  $\Delta G$  may have little meaning as comparisons.

The G14K15 analogues shown in Figure 88 also have significant differences in their  $m$  values. In addition, the BTD analogue is the least stable of all the variants at pH 7.6. The  $m$  value of the Ahx analogue is comparable with that found for K15T16Ahx suggesting that the unfolded states may be similar in structure. If there is a deviation from the assumed two-state folding mechanism, the presence of an increased concentration of a folding intermediate will lower the  $m$  value as is the case for the BTD.

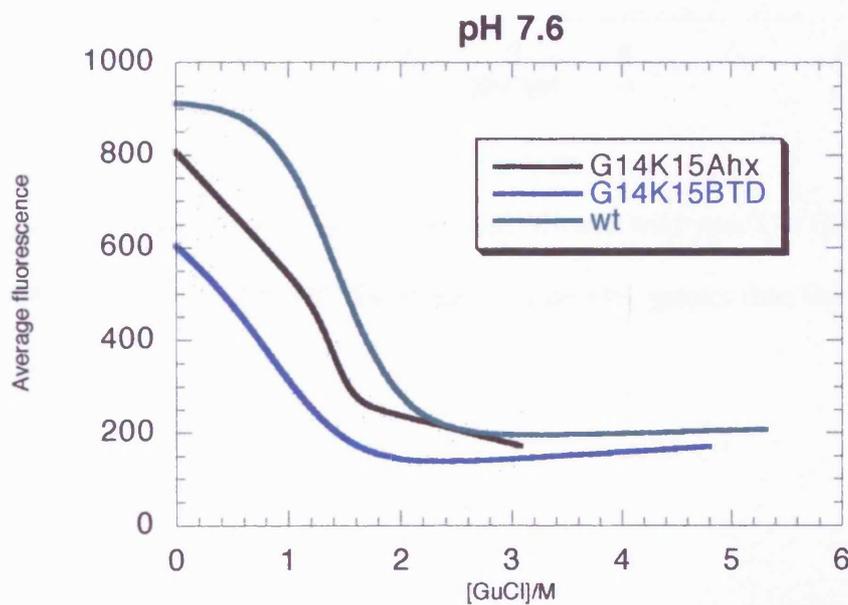


Figure 88.

Figure 88 shows the overlay plots for G14K15BTD, G14K15Ahx and wildtype. The analogues also begin to denature at very small denaturant concentrations. In these cases, it again may be unreasonable to assign any significance to the  $\Delta\Delta G$  values for the same reasons given for the K15T16 s in the previous paragraph.

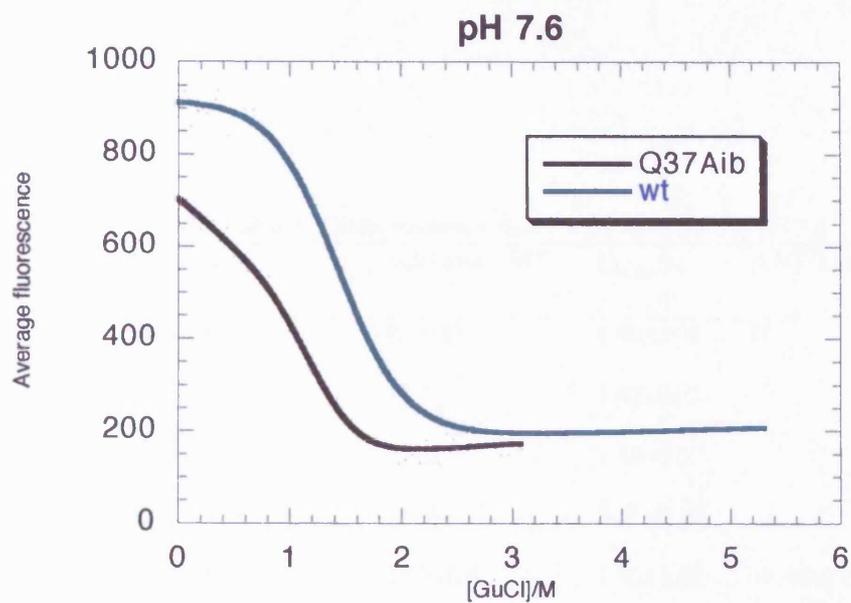


Figure 89.

Figure 89 shows the overlay plots for Q37Aib and wildtype. The Q37Aib is less stable than wildtype at pH 7.6 and the m value 33% greater than the value for the wildtype.

Table 2 gives the results of the thermodynamic analysis on the analogues at pH 7.6.

Table 2. Thermodynamic data for B2 analogues at pH 7.6

Domain	$m/\text{kJ mol}^{-1} \text{ M}^{-1}$	$D_{50\%}/\text{M}$	$\Delta\Delta G/\text{kJmol}^{-1}$
Wildtype	$8.9\pm 0.3$	$1.46\pm 0.01$	0
K15T16Ahx	35.5	$1.42\pm 0.01$	
K15T16BTD	$27\pm 8$	$1.38\pm 0.02$	
K15dPT16N-MeA	$15\pm 8$	$1.42\pm 0.15$	
Q37Aib	$11.7\pm 0.5$	$1.32\pm 0.02$	$-1.42\pm 0.03$
15BTD16	$8.9\pm 0.3$	$1.42\pm 0.02$	$-0.36\pm 0.01$
15Ahx16	$9.1\pm 0.1$	$1.39\pm 0.01$	$-0.63\pm 0.01$
G14K15Ahx	$37\pm 3$	$1.38\pm 0.01$	
G14K15BTD	$6.4\pm 0.1$	$0.87\pm 0.05$	$-4.5\pm 0.07$

$m$  is the dependence of free energy of unfolding on guanidine concentration.

$D_{50\%}$ , is the midpoint of the denaturation curve.

$$\Delta\Delta G = [D_{50\%,(mut)} - D_{50\%,(wt)}][m(wt) + m(mut)]/2$$

pH 4.7

First, the data for each domain are again presented individually to emphasize the raw data. The data for 15BTD16, 15Ahx16 and G14K15Ahx analogues were collected with an excitation wavelength of 295 nm and an emission wavelength of 350 nm, to ensure that only the tryptophan fluorescence was measured. This was an attempt to improve the signal to noise ratio of the experiments. The other analogues and wildtype experiments were conducted with an excitation wavelength of 280 nm and an emission wavelength of 350 nm.

There was no significant reduction in the error in the calculated result. Overlay plots are then shown to compare the fitted curves to wildtype. The fitted curves were found from the same analysis as for the pH 7.6 data. The derived values for the slope of the denaturation curve and  $D_{50\%}$ , the midpoint of the denaturation curve and the  $\Delta\Delta G$  values are tabulated. Figures 90 to 98 show the raw data for 15BTD16, 15Ahx16, G14K15BTD, G14K15Ahx, K15T16BTD, K15T16Ahx, K15dP,T16N-MeA, Q37Aib and wildtype.

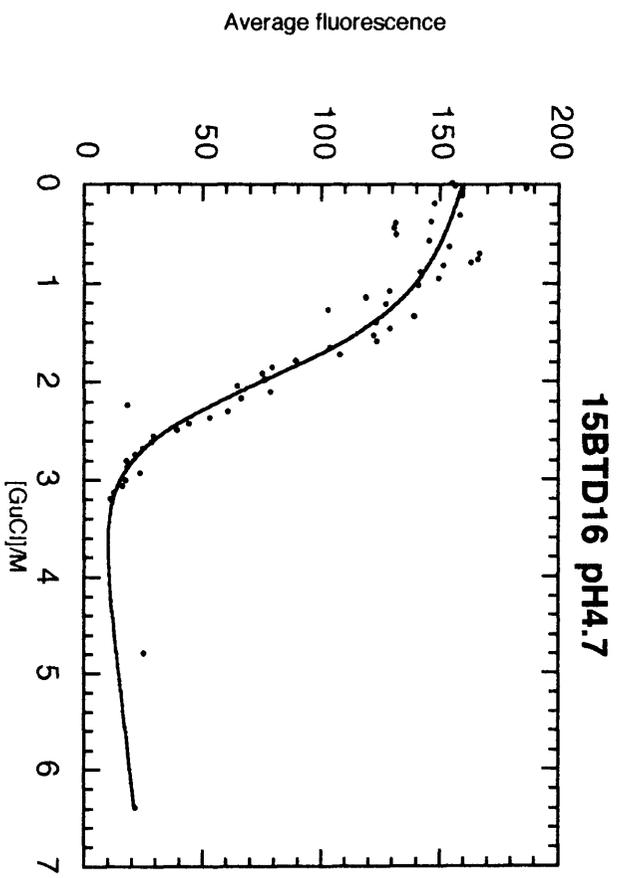


Figure 90

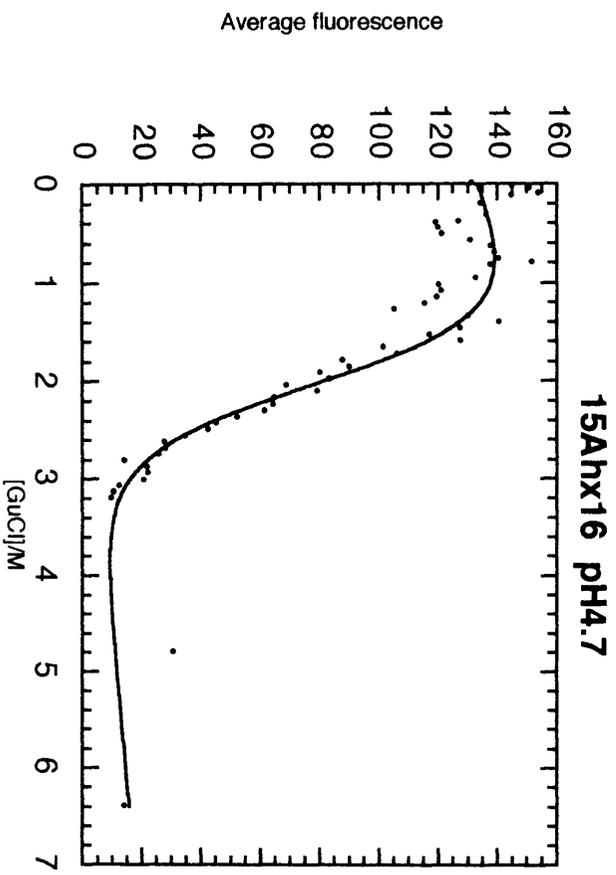


Figure 91.

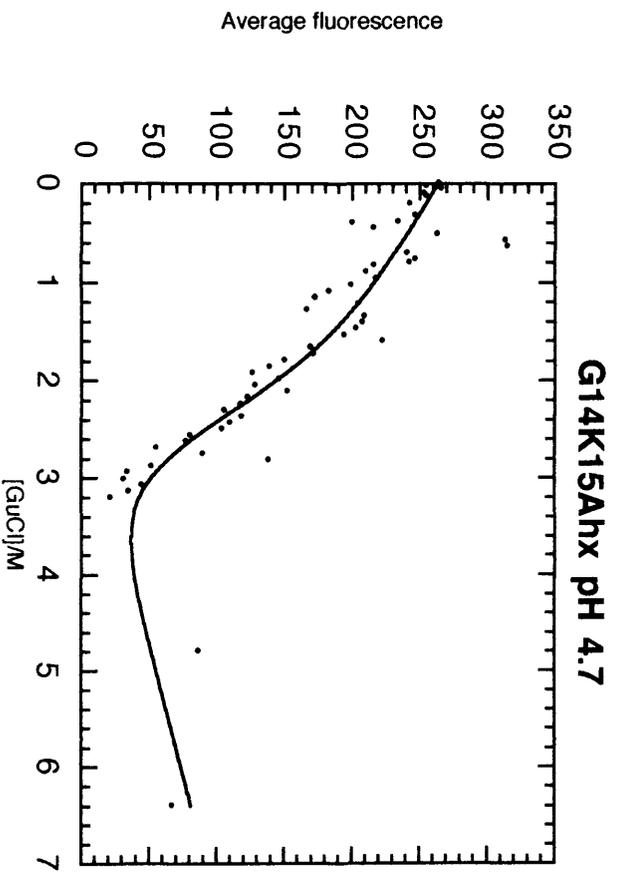


Figure 92.

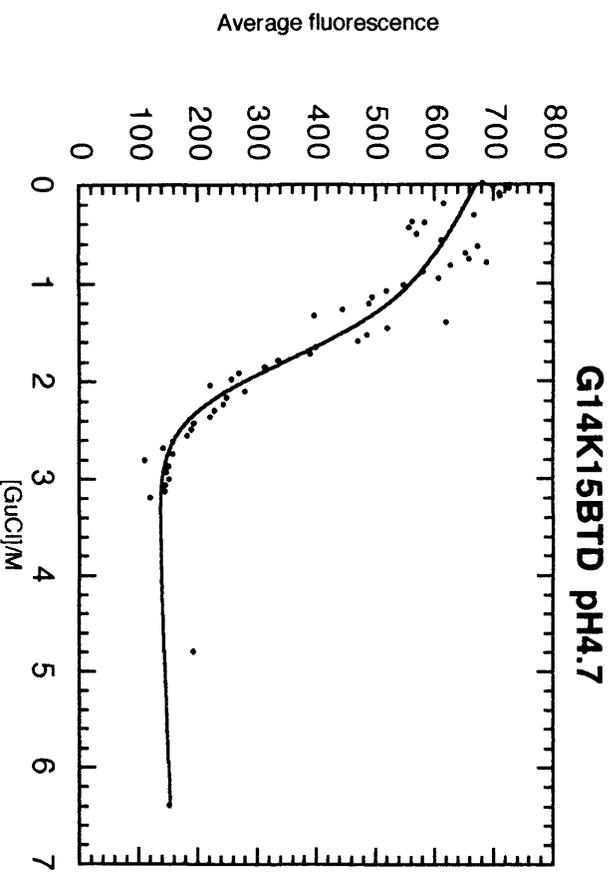


Figure 93.

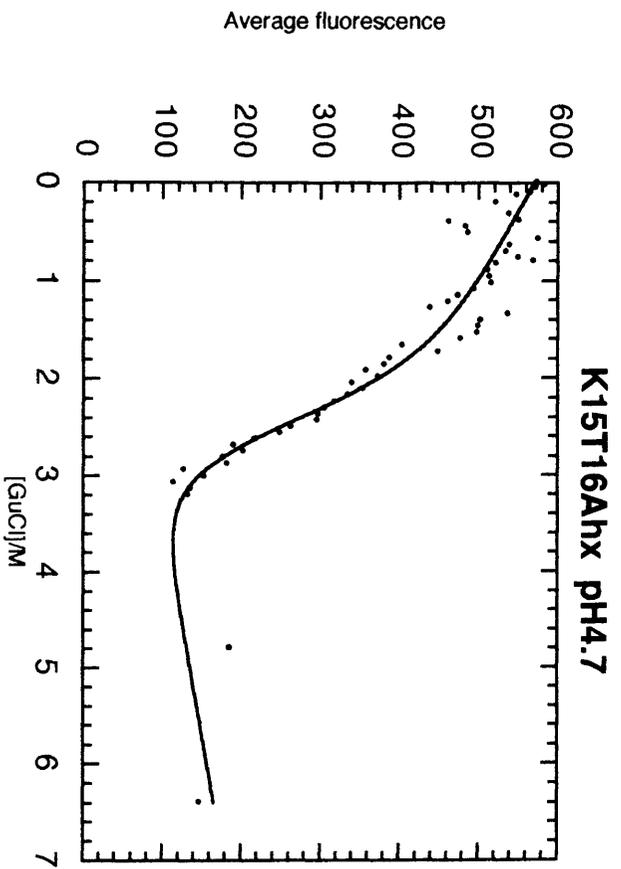


Figure 94.

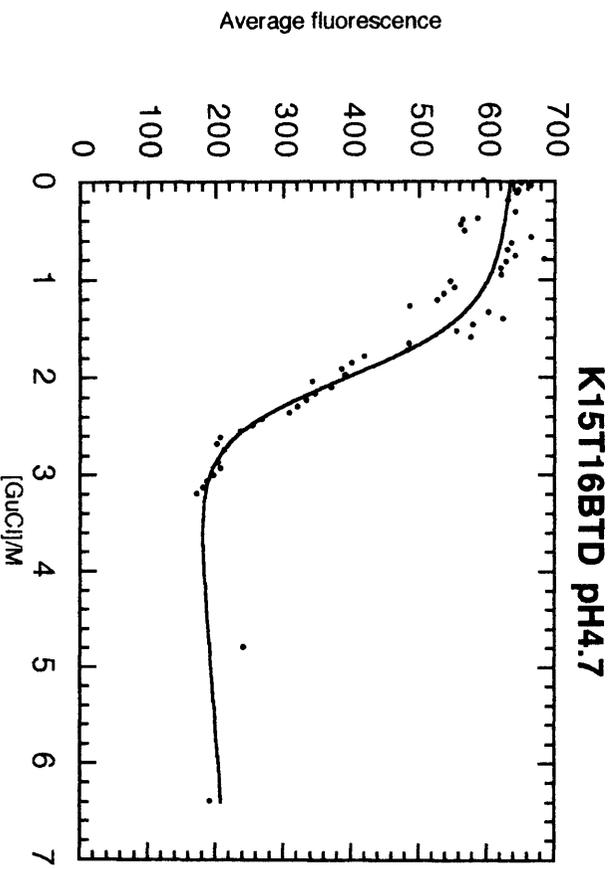


Figure 95.

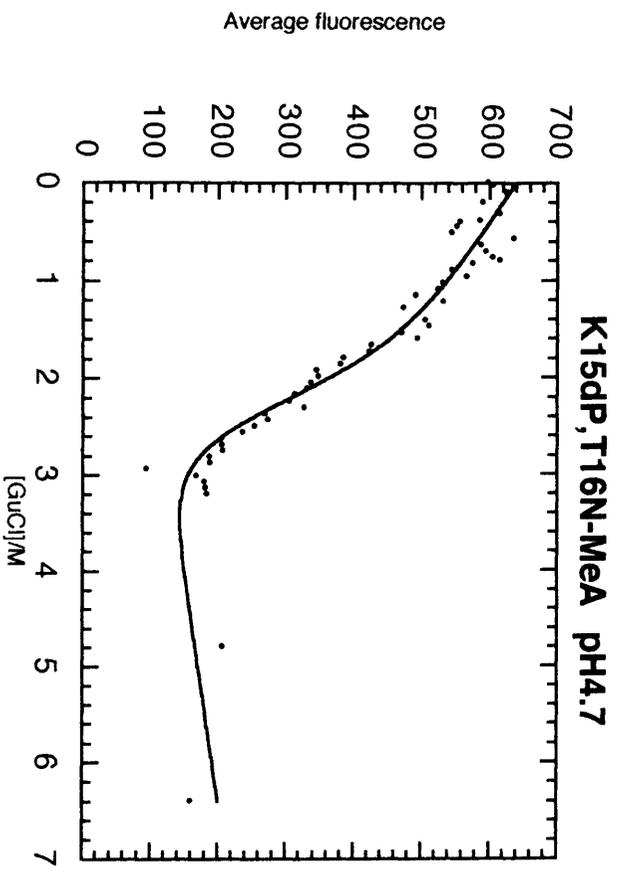


Figure 96.

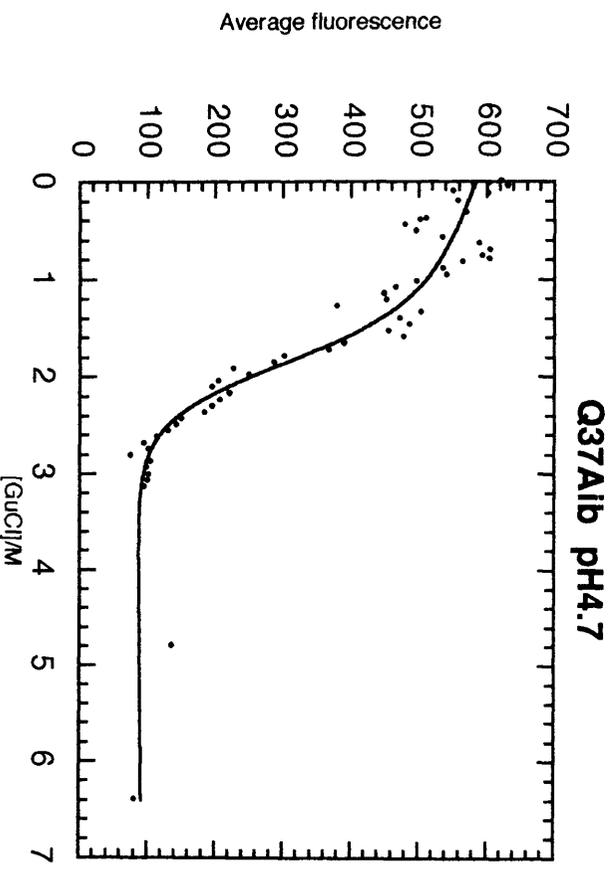


Figure 97.

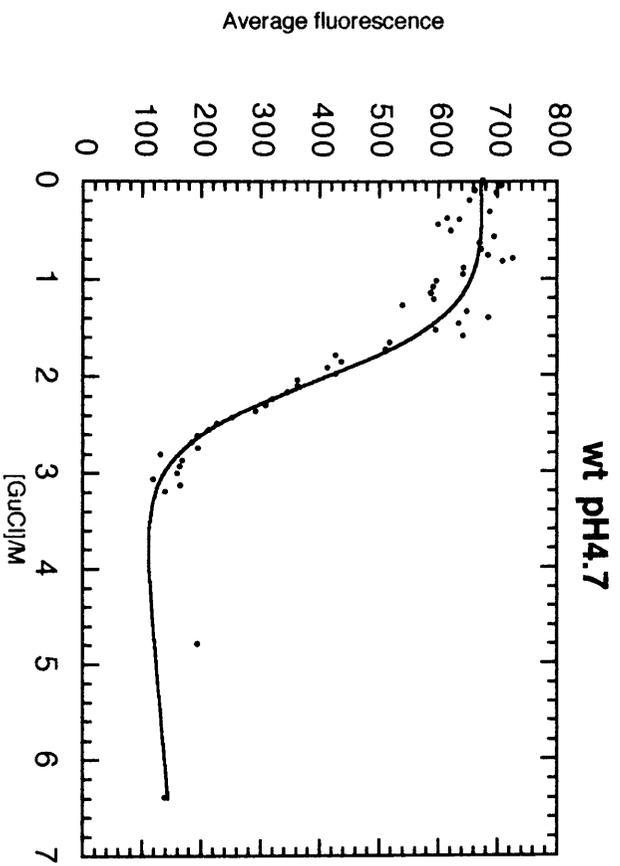


Figure 98.

Figure 99 shows the overlay plots of the fitted curves of 15BTD16, 15Ahx16 and wt.

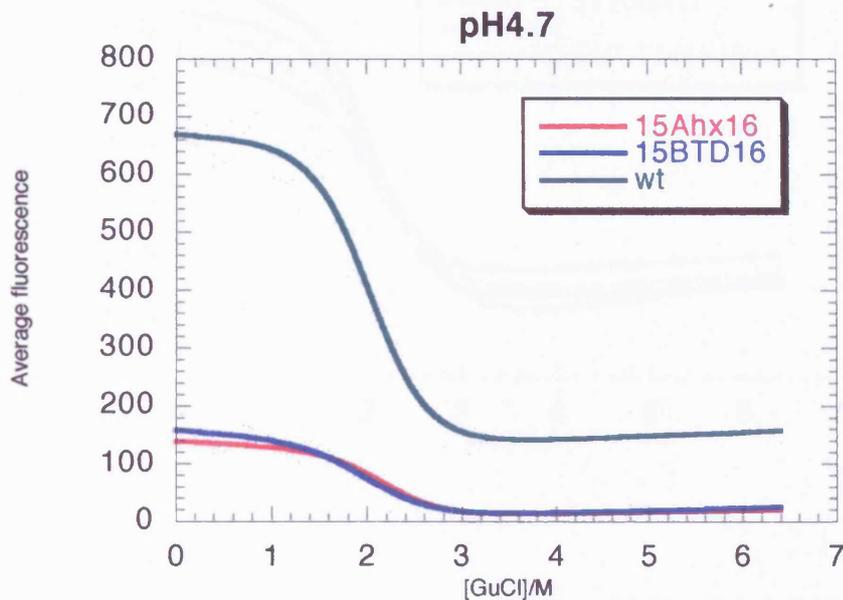


Figure 99.

The insertion analogues and wildtype at pH 4.7 behave in a similar way to their behaviour at pH 7.6. The difference in amplitude between the curves in Figure 99 is because the experiment was conducted with an excitation wavelength of 295nm for the 15-16 insertion analogues, and with an excitation wavelength of 280nm for the wildtype. In the first instance only the tryptophan is excited, while in the second, tryptophan and tyrosines are excited. The experiments were conducted in this way to see if the error estimates in the analyses could be reduced. The conclusion was that it made no difference to the result if the excitation wavelength was changed.

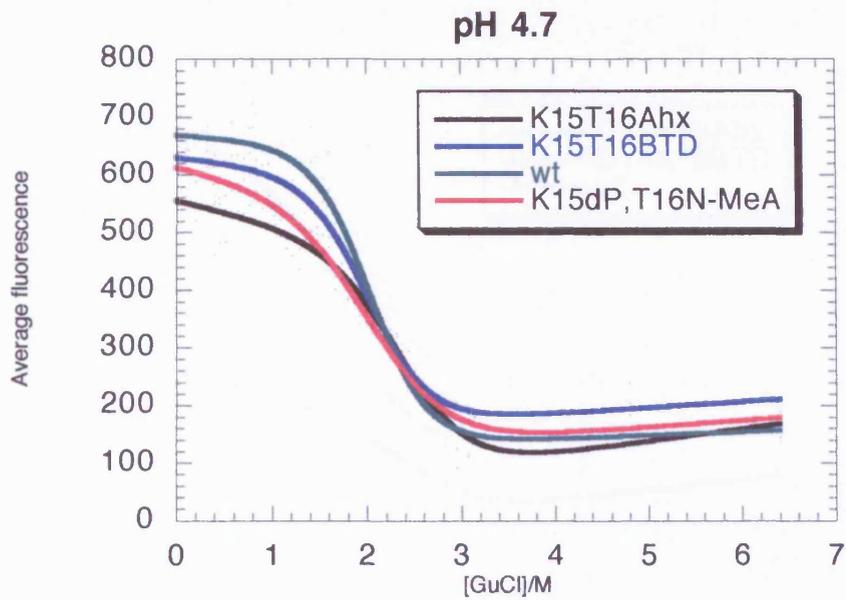


Figure 100.

Figure 100 shows the overlay plots for K15T16Ahx, K15T16BTD, K15dPT16N-MeA, and wildtype. The K15T16 analogues all have similar  $m$  values to wildtype at pH 4.7. Two of the variants, K15T16Ahx and K15dPT16N-MeA are slightly more stable than wildtype at this pH.

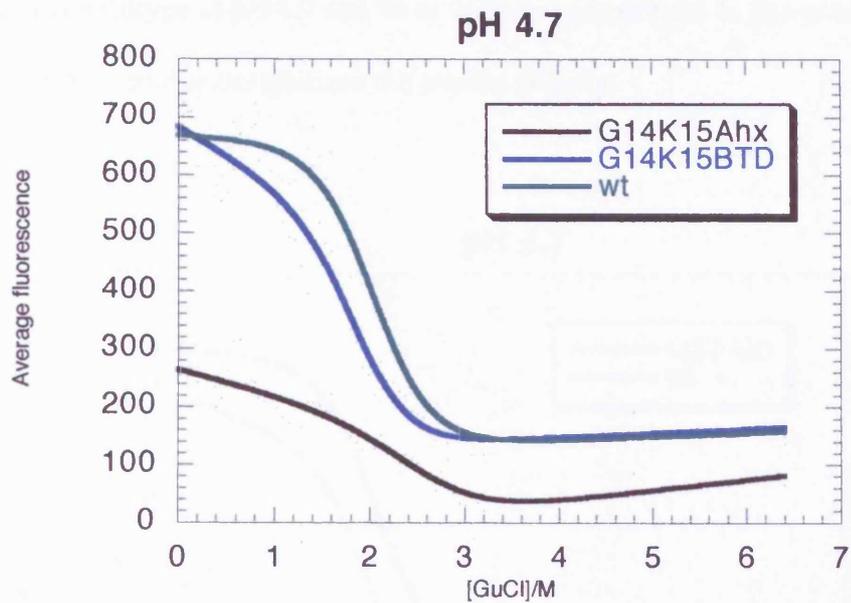


Figure 101

Figure 101 shows the overlay plots for G14K15BTD, G14K15Ahx and wildtype. G14K15BTD is much more stable at pH 4.7 than at pH 7.6, but is still the least stable of all the variants (Figure 101). The  $m$  value for the Ahx variant is 30% lower than the value for wildtype and it would not be reasonable to assign any meaning to the  $\Delta\Delta G$  value.

Figure 102 shows the overlay plots for Q37Aib and wildtype. The Q37Aib is less stable than wildtype at pH 4.7 and its  $m$  value is comparable. In this position in the helix, the Aib residue destabilizes the protein slightly.

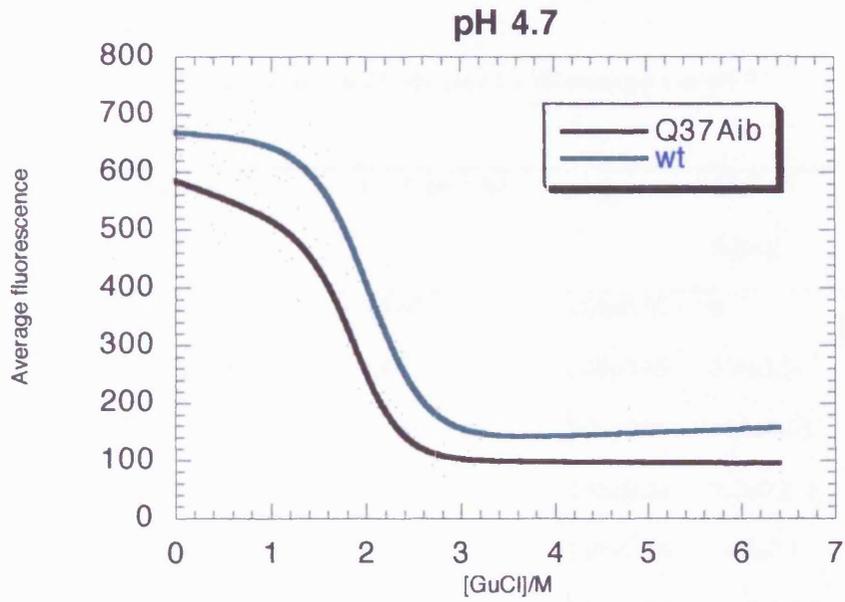


Figure 102.

Table 3 summarizes the results of the thermodynamic analysis for the analogues at pH 4.7.

Table 3. Thermodynamic data for B2 analogues at pH 4.7

Domain	$m/\text{kJ mol}^{-1} \text{ M}^{-1}$	$D_{50\%}/\text{M}$	$\Delta\Delta G$ /kJmol <sup>-1</sup>
Wildtype	6.9±0.5	2.05±0.02	0
K15T16Ahx	7.4±0.7	2.53±0.03	3.4±0.3
K15T16BTD	7±1	2.03±0.06	-0.1±0.01
K15dPT16N-MeA	7.7±0.7	2.35±0.04	2.2±0.2
Q37Aib	8±1	1.91±0.08	-1.0±0.1
15BTD16	6.2±0.5	2.06±0.06	0.07±0.01
15Ahx16	6.7±0.5	2.00±0.04	-0.34±0.02
G14K15Ahx	5±5	2.6±0.3	3±3
G14K15BTD	8±2	1.87±0.14	-1±0.25

$m$  is the dependence of free energy of unfolding on guanidine concentration.

$D_{50\%}$ , is the midpoint of the denaturation curve.

$$\Delta\Delta G = [D_{50\%}(\text{mut}) - D_{50\%}(\text{wt})][m(\text{wt}) + m(\text{mut})]/2$$

At pH 7.6 all the variants are less stable than wildtype, with the insertion s having the least effect of all. At pH 4.7, the insertion analogues are as stable as wildtype. Both Ahx variants and the d-proline, N-methyl alanine substitutions are more stable than wildtype. All others are only marginally less stable.

All of the proteins are less stable at pH 7.6 than at pH 4.7. At 298K, RNase T<sub>1</sub> is most stable at pH 4.5 and this is near the pH where the net charge is zero. Protein G B2 domain has a pI of 4.7. At this pH the seven positive charges will be neutralized by seven of the twelve negative charges. It is reasonable to assume that the charges on globular proteins are generally arranged so that the electrostatic interactions contribute favourably to the conformational stability. At pH 7.6, the ratio of the charged groups will be changed and if the ratio at pH 4.7 is an optimum, then a change in protonation will result in destabilization. This may tend to favour the maximum conformational stability occurring near the pI. This reasoning may apply to the observed differences in  $m$ , the dependence of  $\Delta G$  on guanidine concentration, at pH 7.6. The accepted mechanism of guanidine function is that it binds to sites on the protein and disrupts the binding forces on the chain. Guanidine may be binding to different sites in the folded and unfolded states at pH 7.6 because of the difference in protonation states at this pH, which is less pronounced at pH 4.7, a pH very close to the pI for this protein.

### 3.4 Circular Dichroism

Circular dichroism measurements on wildtype protein G B2 domain and the analogues K15T16Ahx, K15T16BTD, G14K15Ahx and G14K15BTD show that the structures of the analogues are similar to the wild type with minima at 205 to 208 nm. Figure 103 shows the spectra, corrected for concentration. Table 4 contains a summary of the experimental conditions and the secondary structure analysis performed using the CONTIN program.

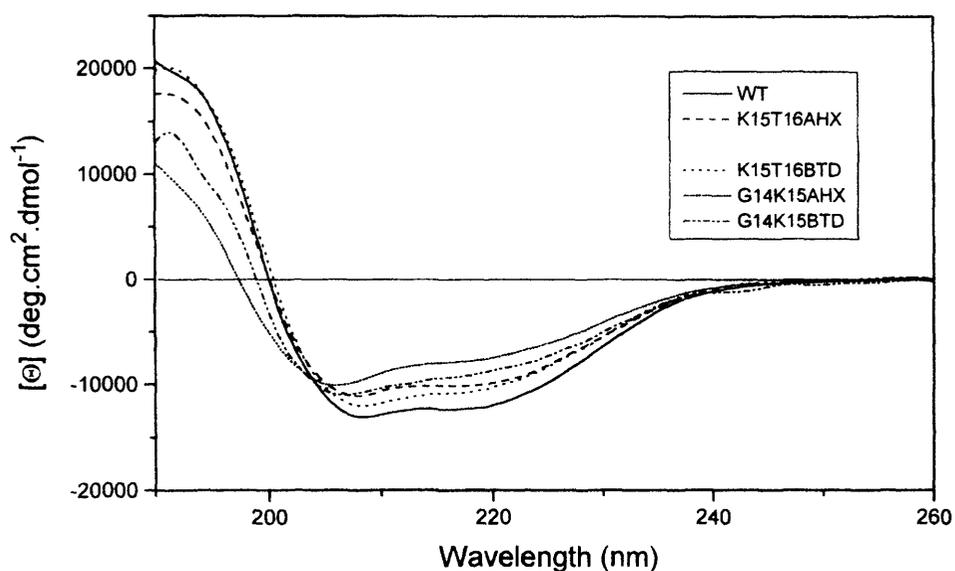


Figure 103. CD analysis of wildtype and four analogues. Sample buffer: 20mM acetate, pH4.7. Protein concentrations are given in Table 4. Spectra were run in cuvettes of pathlength 0.02 cm at T=298K .

Table 4.

Domain	[ ] in mg/ml	%helix	%sheet	Remainder
Wildtype	0.33	24±0.61	34±0.64	42±1.1
K15T16Ahx	0.32	24±.69	45±0.72	32±1.3
K15T16BTD	0.30	31±0.53	48±0.55	22±0.97
G14K15Ahx	0.32	15±0.48	41±0.5	44±0.88
G14K15BTD	0.16	27±0.56	47±0.86	25±1.3

While the percentage of helix is broadly the same in three of the analogues compared to wildtype, the fourth, G14K15Ahx seems to have 30% less. All the analogues have 20% to 40% more sheet structure than wildtype. With the exception of G14K15Ahx all the analogues have much less random structure than wildtype.

CD was also used to monitor the effect of increasing the concentration of guanidinium hydrochloride on the protein solutions. The ellipticity at 225nm is a typical index of loss of secondary structure. The information in Table 4 suggests that a direct comparison of the denaturing curves is possible because the percentage of secondary structure elements to be monitored is roughly the same in the absence of denaturant.

Figure 104 shows the CD spectra of wildtype protein G B2 domain at different guanidine concentrations and Table 5 shows the data for the ellipticity at 225 nm as measured from the CD spectra.

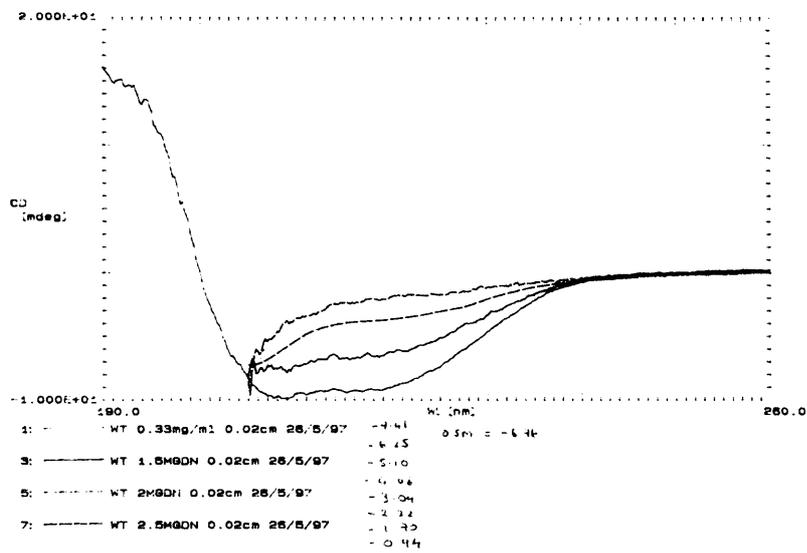


Figure 104.

Table 5.

[GuClI]/M	Ellipticity at 225nm/mdeg	% change
0	-7.41	0
0.5	-6.76	10
1.0	-6.25	17.9
1.5	-5.10	35.7
1.75	-4.06	51.8
2.0	-3.04	67.5
2.25	-2.22	80.2
2.5	-1.7	88.3
4	-0.94	100

Figure 105 shows the CD spectra of K15T16Ahx B2 domain at different guanidine concentrations and Table 6 shows the data for the ellipticity at 225 nm as measured from the CD spectra.

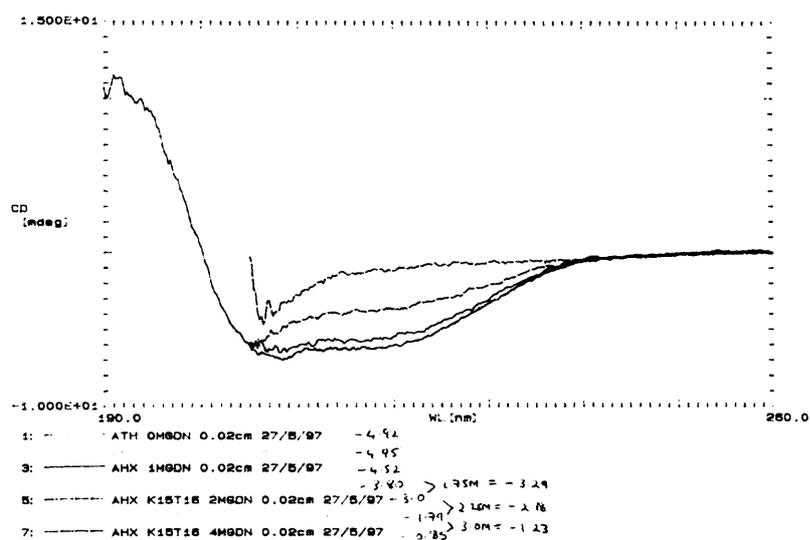


Figure 105.

where  $\Theta$  is the observed ellipticity,  $\Theta_n$  and  $m_n$  are the intercept and slope, respectively, of the pre-transition baseline,  $\Theta_u$  and  $m_u$  are the intercept and slope of the post-transition baseline,  $m$  is the dependence of the free energy of unfolding on guanidine concentration, and  $D_{1/2}$  is the midpoint of the denaturation curve. The experimental points were fitted using Kaleidagraph™ software. The free energy of unfolding (defined as the conformational stability of the protein)  $\Delta G_{\text{water}}$  is equal to  $m[D_{1/2}]$ . By calculating  $\Delta\Delta G$  from the difference in midpoints, a long extrapolation back to 0M guanidine is avoided.

Figures 106 to 110 show the raw data and the calculated curves.

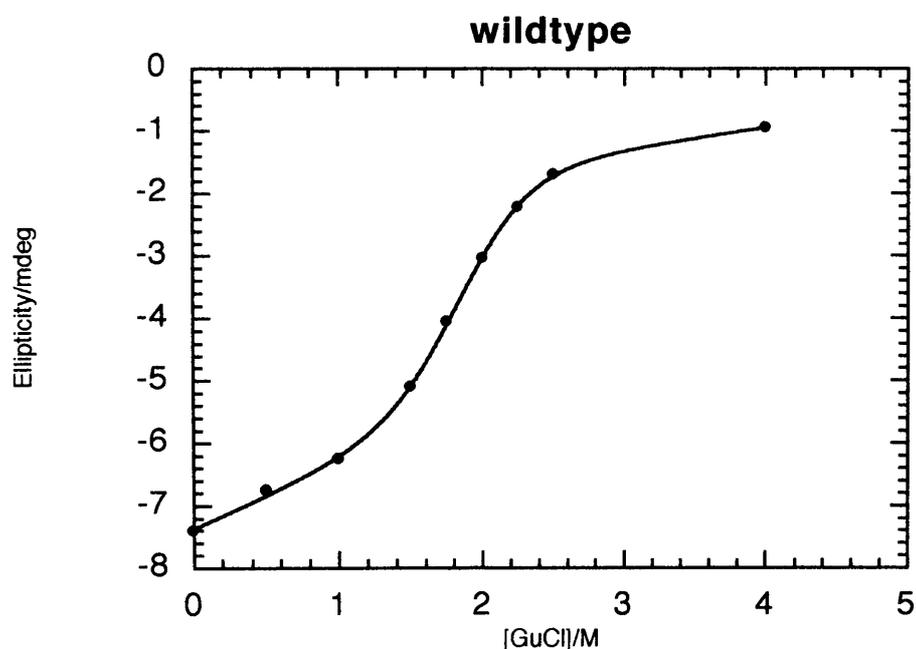


Figure 106.

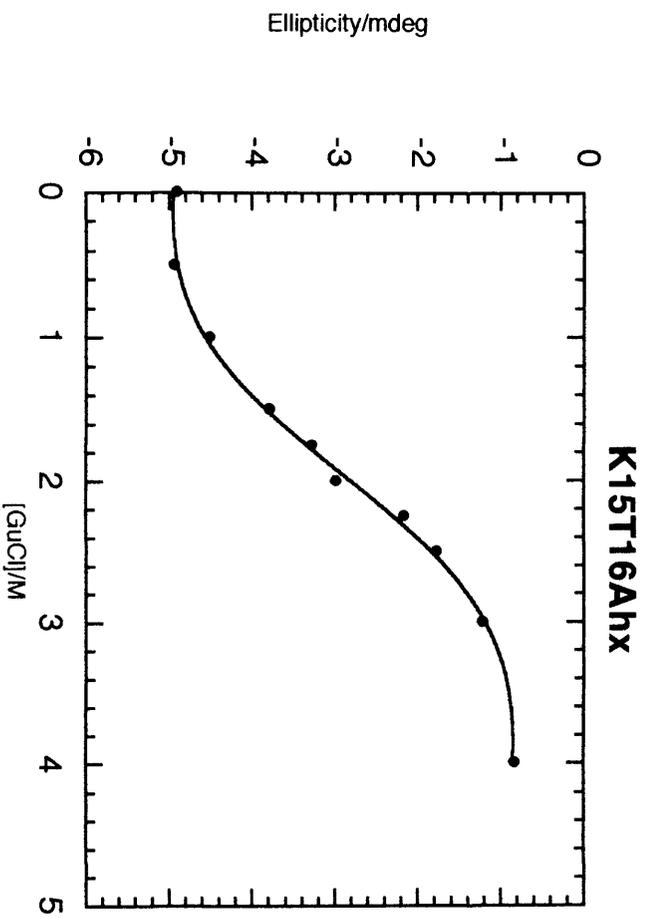


Figure 107.

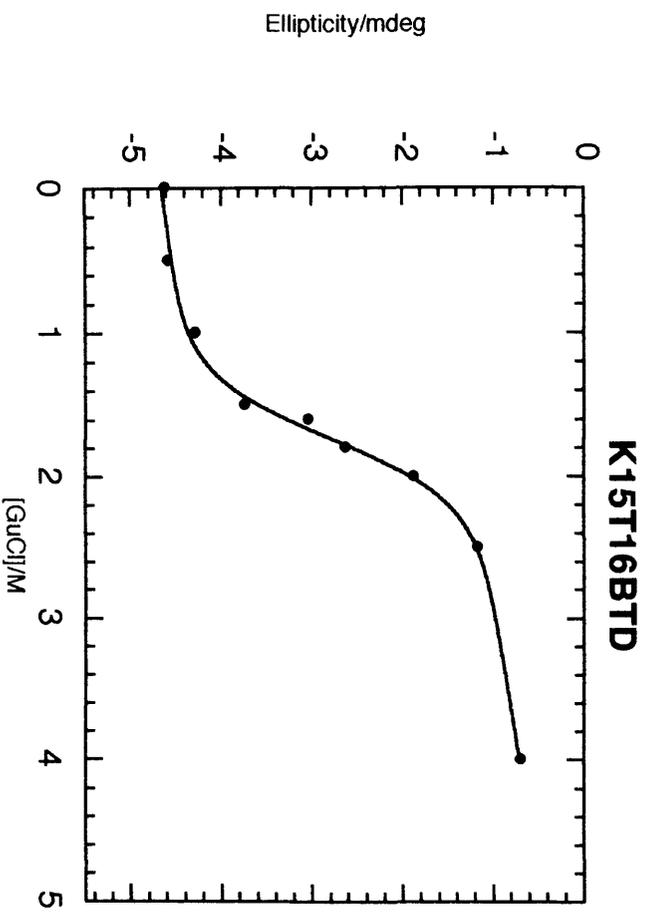


Figure 108.

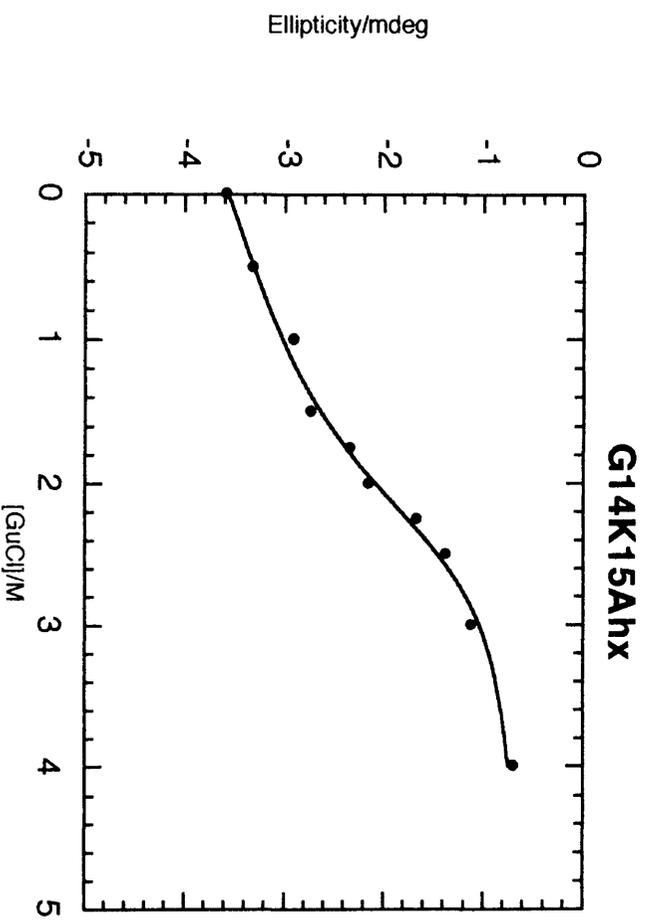


Figure 109.

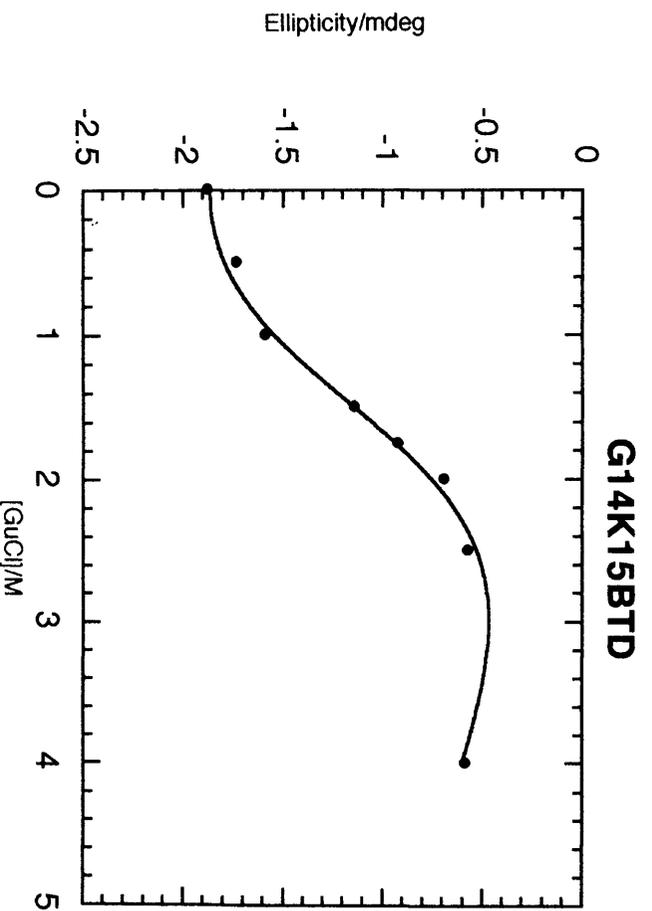


Figure 110.

Figure 111 shows an overlay plot of all the calculated curves, with the denaturant concentration plotted against percentage of the total change.

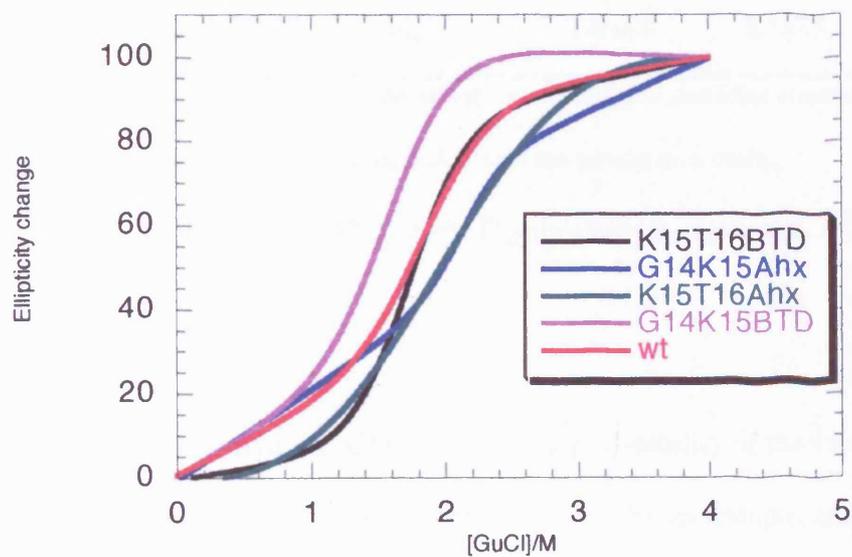


Figure 111.

Table 7 shows the results of the curve fitting analysis.

Table 7. Thermodynamic data for B2 analogues at pH 4.7

Domain	$m/\text{kJ mol}^{-1} \text{ M}^{-1}$	$D_{50\%}/\text{M}$	$\Delta\Delta G/\text{kJmol}^{-1}$
Wildtype	$10\pm 1$	$1.87\pm 0.04$	
K15T16Ahx	$4\pm 2$	$1.92\pm 0.25$	$0.4\pm 0.2$
K15T16BTD	$11\pm 4$	$1.73\pm 0.09$	$-1.5\pm 0.5$
G14K15Ahx	$15\pm 7$	$2.2\pm 0.1$	$4\pm 2$
G14K15BTD	$10\pm 2$	$1.6\pm 0.7$	$-2.7\pm 0.5$

$m$  is the dependence of free energy of unfolding on guanidine concentration.

$D_{50\%}$ , is the midpoint of the denaturation curve.

$$\Delta\Delta G = [D_{50\%,(mut)} - D_{50\%,(wt)}][m(wt) + m(mut)]/2$$

According to the analysis of CD data, the order of stability of the variants with respect to wildtype is as follows: G14K15BTD is the least stable, then K15T16Ahx, then K15T16BTD, and G14K15Ahx is the most stable. However the  $m$  value for the K15T16Ahx seems anomalous when compared to the other variants.

### 3.5 Calorimetry

The thermal unfolding transitions of wildtype Protein G B2 domain and the analogues K15T16Ahx, K15T16BTD, G14K15Ahx and G14K15BTD were studied using differential scanning calorimetry. The raw data for a typical DSC scan of the unfolding transition for B2 is shown in Figure 112.

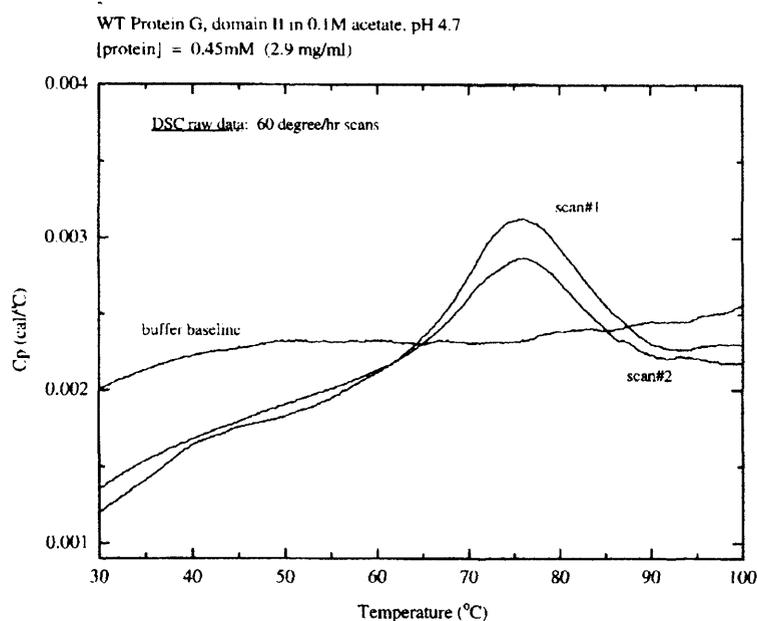


Figure112.

The unfolding reaction is quantitatively reversible throughout the transition. Scan #2 above represents the unfolding transition of the material in scan#1 that had been allowed to return to room temperature. When the buffer scan is subtracted from the raw data, a single symmetrical transition is observed with a midpoint of 75°C at pH 4.7 (Figure 113).

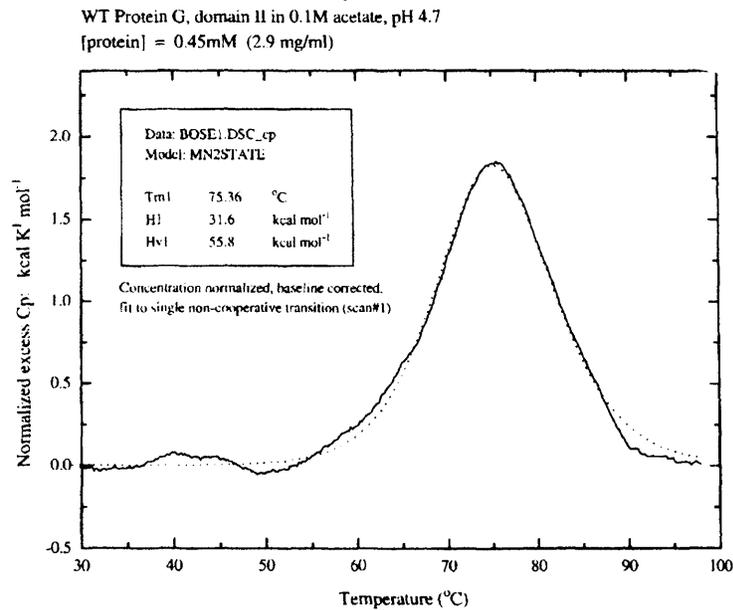


Figure 113.

The dotted line in Figure 113 also shows a concentration normalized, baseline corrected, fit to a single non-cooperative transition. A non-linear least squares method fit the experimental data to a theoretical curve by varying the parameters  $T_m$  and  $\Delta H_{vH}$  to minimize the standard deviation of the observed values of the excess heat capacity from the calculated curve.

The equation used to fit the data is given below.

$$C_p(T) = \frac{K_A(T)\Delta H_A(T)^2}{(1 + K_A(T))^2 RT^2} + \dots$$

where

$$K_A(T) = \exp\left\{\frac{-\Delta H_{m_A}}{RT}\left(1 - \frac{T}{T_{m_A}}\right)\right\}$$

and

$$\Delta H_A(T) = \Delta H_{m_A} + \Delta C_{p_A}(T - T_{m_A})$$

$C_p(T)$  is the system's heat capacity at any temperature  $T$ .  $\Delta H_A$  is the van't Hoff heat change for the cooperative unit, which participates in the unfolding.  $\Delta H_{n_A}$  is the heat change at the midpoint.  $\Delta C_{p_A}$  is the heat capacity change for the transition.  $T_{m_A}$  is the temperature of the midpoint.  $K_A$  is the equilibrium constant for the unfolding reaction.

The closeness of fit shows that the data fits a simple two-state model for the wildtype.

Figure 114 shows the fit to a theoretical curve for the concentration normalized, baseline corrected, fit to a single non-cooperative transition for the G14K15Ahx analogue.

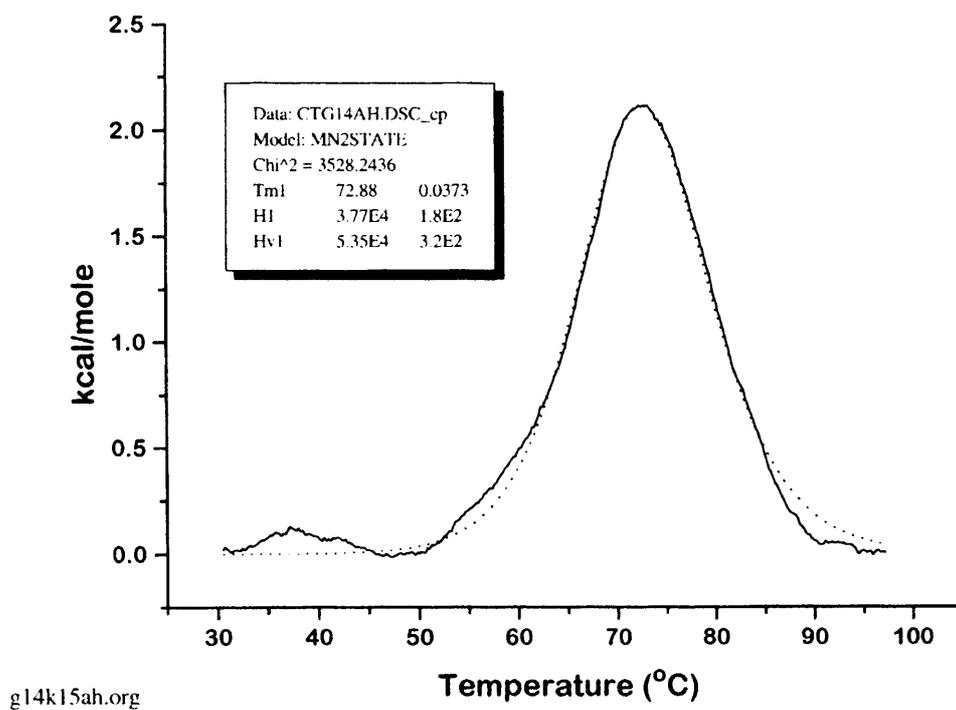


Figure 114.

The  $T_m$  and  $\Delta H_{vH}$  for this domain are  $72.88 \pm 0.04^\circ\text{C}$  and  $209 \pm 1$  kJ/mol respectively.

Figure 115 shows the fit to a theoretical curve for the concentration normalized, baseline corrected, and the fit to a single non-cooperative transition for the G14K15BTD.

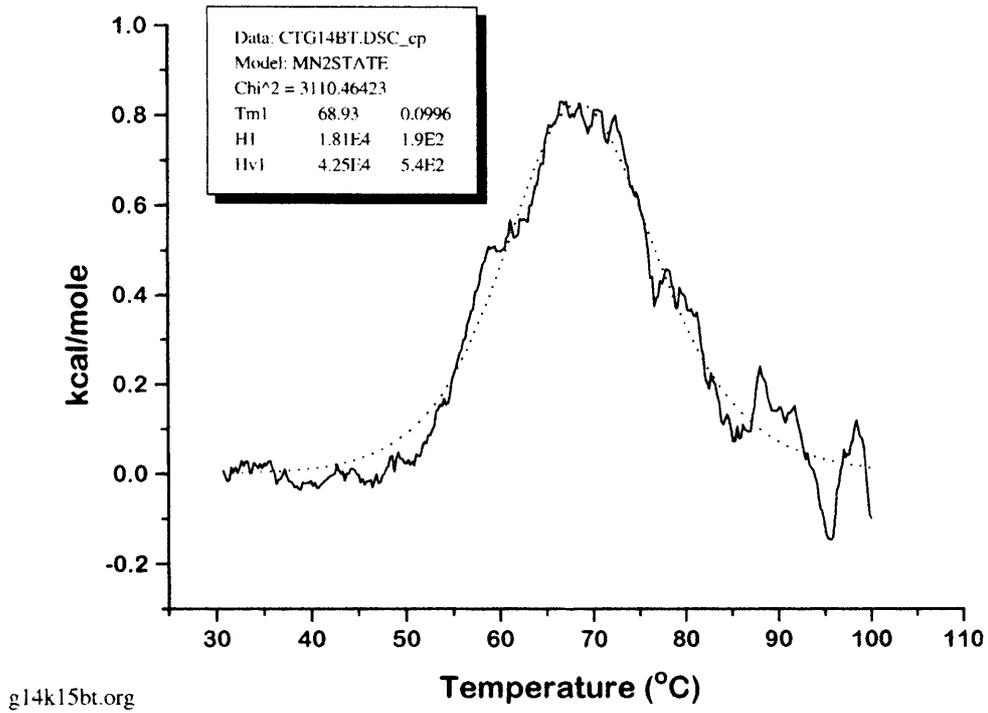


Figure 115.

The  $T_m$  and  $\Delta H_{vH}$  for this domain are  $68.93 \pm 0.1^\circ\text{C}$  and  $178 \pm 2$  kJ/mol respectively.

The signal to noise ratio is worse than the other determinations because the concentration of G14K15BTD was about 50% of the other domains.

Figure 116 shows the fit to a theoretical curve for the concentration normalized, baseline corrected, and the fit to a single non-cooperative transition for the K15T16Ahx.

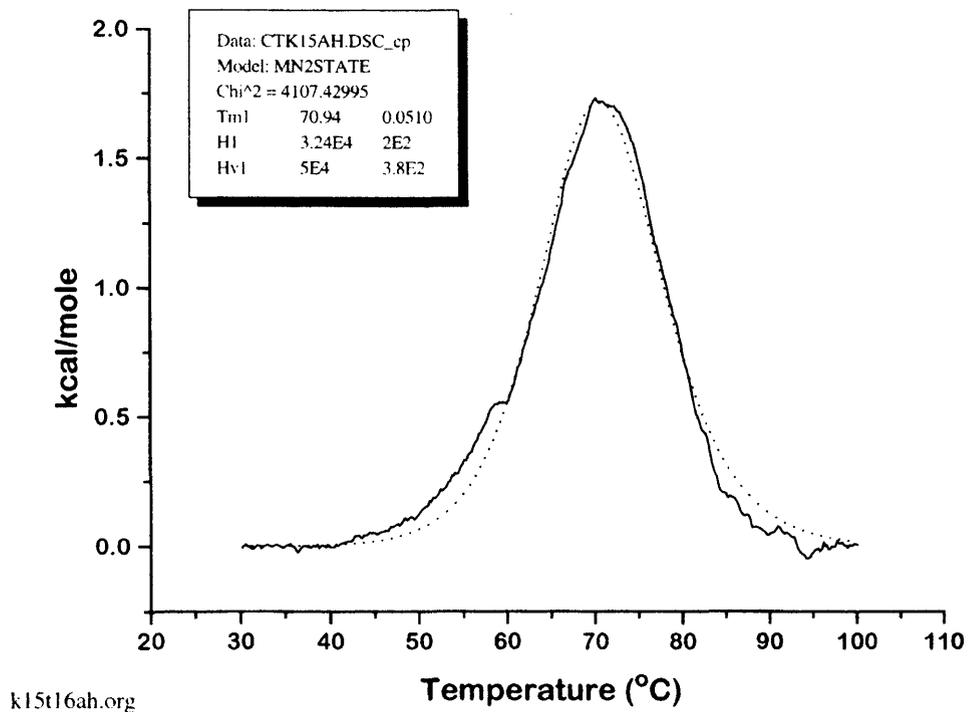


Figure 116.

The  $T_m$  and  $\Delta H_{vH}$  for this domain are  $70.94 \pm 0.05^\circ\text{C}$  and  $209 \pm 2$  kJ/mol respectively.

Figure 117 shows the fit to a theoretical curve for the concentration normalized, baseline corrected, and the fit to a single non-cooperative transition for the K15T16BTD.

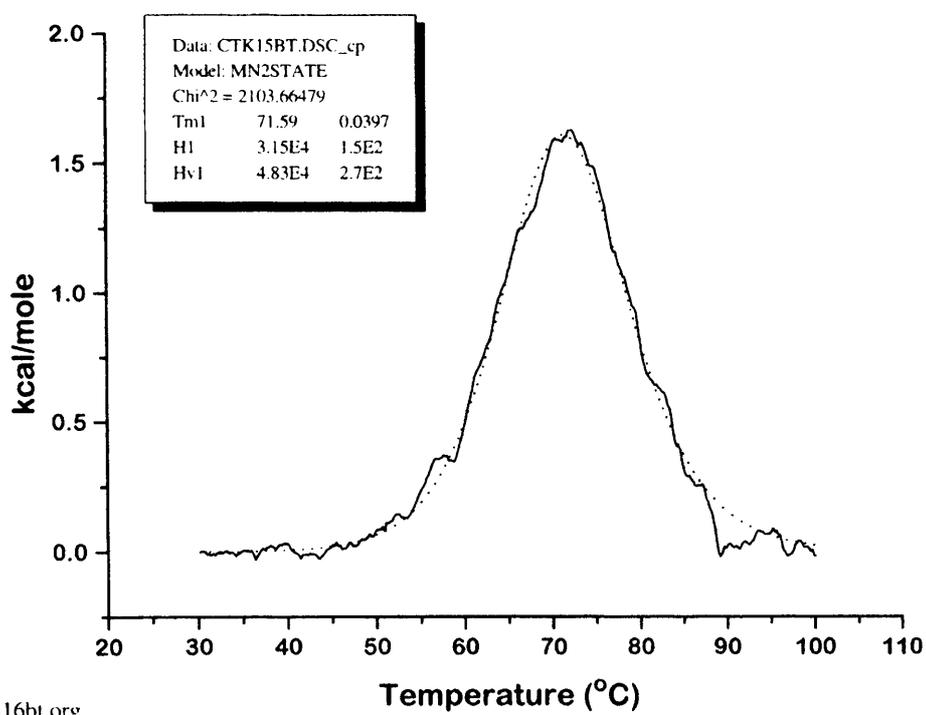
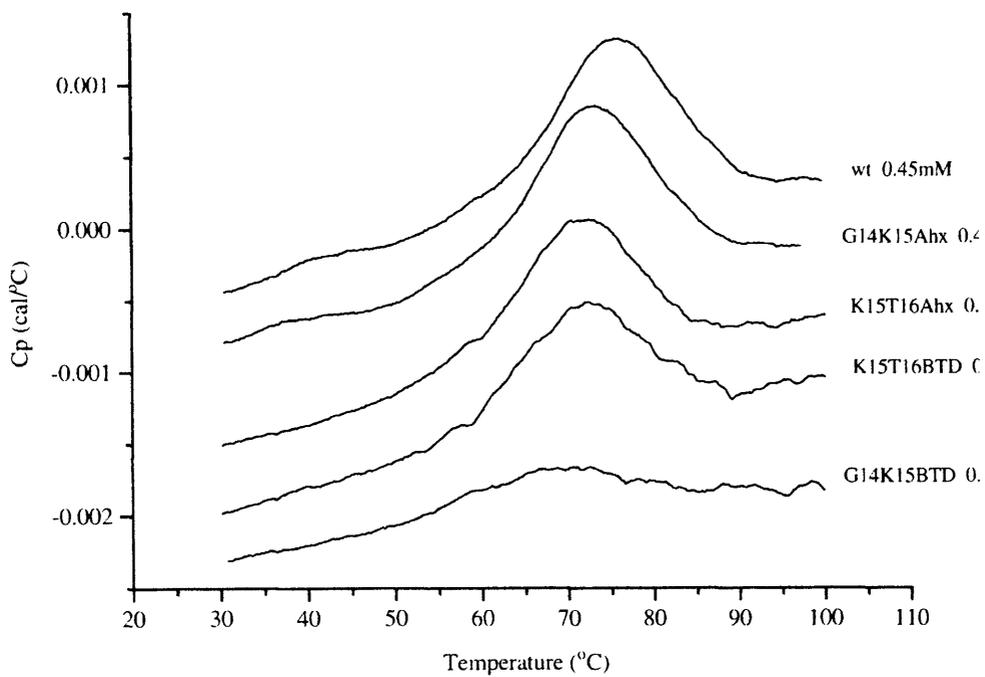


Figure 117.

The  $T_m$  and  $\Delta H_{vH}$  for this domain are  $71.59 \pm 0.04^\circ\text{C}$  and  $202 \pm 1$  kJ/mol respectively.

Figure 118 shows an overlay of the raw data for all the domains, with the buffer baseline subtracted, and with the curves offset for clarity.



bose-mut.org

Figure 118.

For this work,  $\Delta\Delta G$ , the difference in conformational stability is defined by the equation

$$\Delta(\Delta G) = \Delta G (\text{mut}, \text{H}_2\text{O}, 298\text{K}) - \Delta G (\text{wt}, \text{H}_2\text{O}, 298\text{K})$$

The simplest equation that can be used to estimate the  $\Delta(\Delta G)$  value is

$$\Delta(\Delta G) = [(T_m(\text{mut}) - T_m(\text{wt}))][\Delta H_m(\text{wt})/T_m(\text{wt})]$$

A positive  $\Delta(\Delta G)$  indicates a more stable analogue. Table 8 shows the results of applying this relation to the calorimetric data.

Table 8. Thermodynamic data for B2 analogues at pH 4.7

Domain	[ ] /mM	$T_m$ /K	$\Delta H_{vH}$ /kJmol <sup>-1</sup>	$\Delta\Delta G$ / kJmol <sup>-1</sup>
Wildtype	0.45	348.51±0.04	234±1	0
K15T16Ahx	0.46	344.09±0.05	209±2	-2.96±0.01
K15T16BTD	0.43	344.74±0.05	202±1	-2.53±0.01
G14K15Ahx	0.41	346.03±0.04	224±1	-1.67±0.01
G14K15BTD	0.25	342.09±0.1	178±2	-4.31±0.02

According to the analysis of the calorimetric data, the order of the variants with respect to wildtype is as follows: G14K15BTD is the least stable, then K15T16Ahx, then K15T16BTD, and G14K15Ahx is the most stable. All the variants are slightly less stable than wildtype.

The same order of stability is observed in the guanidine denaturation studies at pH 4.7. In this case, however, the analysis shows that the Ahx analogues are slightly more stable than wildtype, and the BTD analogue slightly less stable.

### **3.6 Binding Studies**

The binding of the protein G analogues to IgG molecules was measured using a BIAcore 2000 instrument (Pharmacia). Within a microfluidic flow cell, the BIAcore monitors soluble protein binding to an immobilized ligand using changes to the surface plasmon resonance that are caused by a change in solution refractive index near a plane surface. Surface plasmon resonance is observed when polarized light is reflected from a metal surface. It is a decrease in light intensity for a specific angle of incidence. This angle changes with changes in the refractive index in the vicinity of the surface due to adsorption of large molecules on the immobilized ligand on the sensor chip. The association rates were determined from the concentration dependence on the rate of association of the protein G molecules.

The initial measurements shown below were made to determine

1. The order of magnitude of binding that could be expected from the analogues.
2. Whether to use whole IgG, Fab2 or Fab.
3. Which antibody to measure against.
4. The most appropriate measuring procedure to adopt.

Initial measurements with whole protein G from Pharmacia and synthetic protein G B2 domains on goat IgG showed large bulk effects from the synthetic material. The bulk effect is the sudden large increase in refractive index due to high protein concentration near the chip surface. The goat IgG was immobilized on the sensor chip and varying dilutions of either the synthetic or the Pharmacia protein G were injected. Figure 119 shows the change in refractive index units for the on and off rates of synthetic B2 domain.

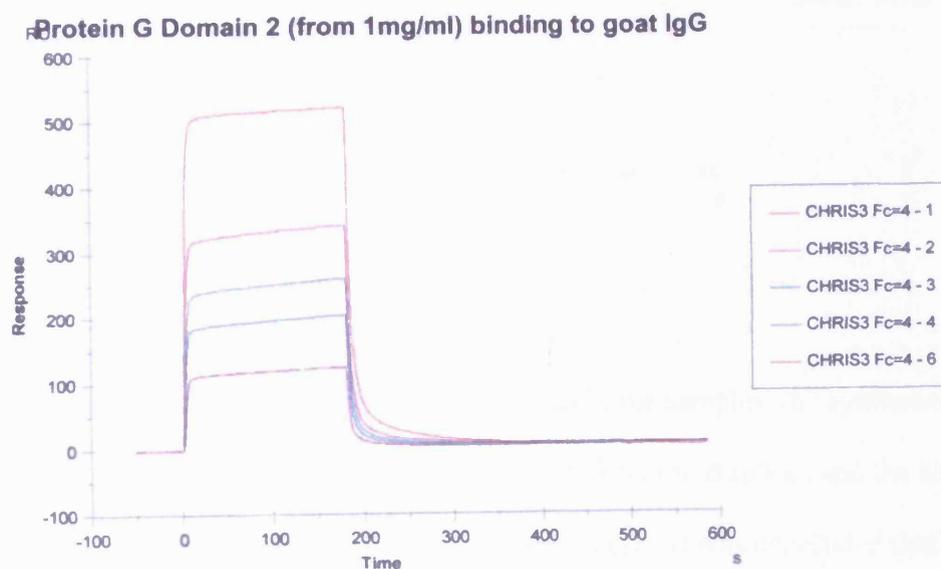


Figure 119.

Figure 120 shows the change in refractive index units for the on and off rates of protein G (Pharmacia).

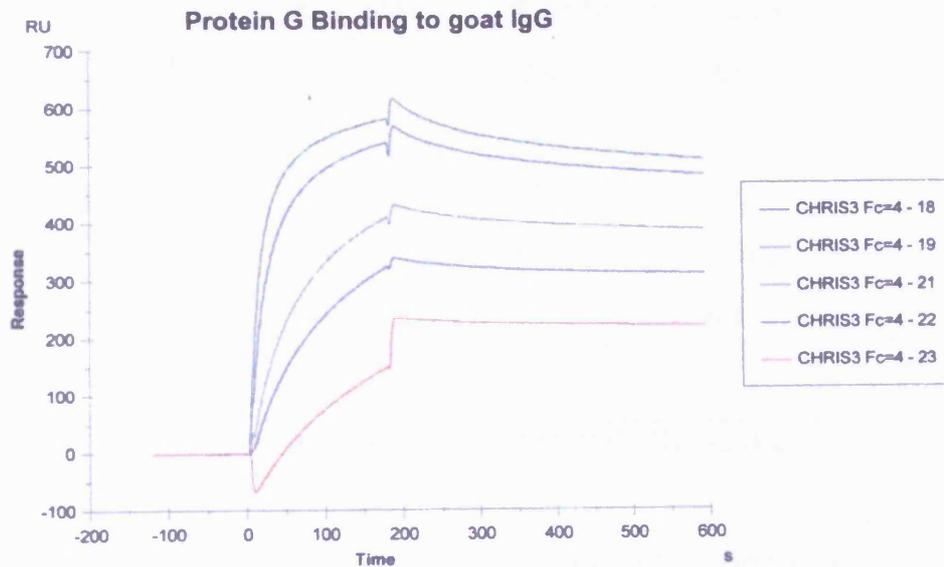


Figure 120.

Assuming that the bulk effects were caused by salt in the samples, the synthetic B2 domain was desalted into the running buffer, HEPES buffered saline, and the binding repeated. The trace in Figure 119 was reproduced exactly. It was concluded that the experimentally observed effect was not due to salt effects but to the high concentration of synthetic protein used. The concentration range in the experiment with protein G (Pharmacia) was from 15.6  $\mu\text{g/ml}$  down to 0.9  $\mu\text{g/ml}$ . For the synthetic domain the concentration range was from 1mg/ml down to 31  $\mu\text{g/ml}$  in order to get a measurable effect with the 6443 MW domain.

The results are shown in Table 9 for synthetic B2 domain.

Table 9: Kinetic and Equilibrium Constants for Synthetic Protein G B2 Domain binding to goat IgG.

[ ]/ $\mu\text{M}$	$k_{\text{ass}} \times 10^3/\text{M}^{-1} \text{s}^{-1}$	$K_{\text{diss}} \times 10^3/\text{s}^{-1}$	$K_{\text{d}}/\mu\text{M}$
200	2.85	12.4	4.4
100	5.92	9.93	1.68
50	1.23	9.27	0.75
25	0.26	8.88	1.34
12.5	0.57	6.65	0.12

---

Equilibrium Constants were determined by  $K_{\text{d}} = k_{\text{diss}}/k_{\text{ass}}$

For comparison, the data in Table 10 are for the protein G (Pharmacia)

Table 10: Kinetic and Equilibrium Constants for Protein G (Pharmacia) binding to goat IgG.

[ ]/ $\mu\text{M}$	$k_{\text{ass}} \times 10^3/\text{M}^{-1} \text{s}^{-1}$	$K_{\text{diss}} \times 10^3/\text{s}^{-1}$	$K_{\text{d}}/\mu\text{M}$
1	0.82	1.12	0.014
0.5	0.92	1.19	0.013
0.125	0.016	0.68	0.0042
0.06	0.022	0.42	0.0019
0.03	0.021	0.26	0.0012

---

Equilibrium Constants were determined by  $K_{\text{d}} = k_{\text{diss}}/k_{\text{ass}}$

At the high concentrations of the synthetic B2 domain, there appears to be a mixture of Fab and  $F_c$  binding being measured. At the lower concentrations of the protein G (Pharmacia), only the  $F_c$  binding remains.

IgG and F(ab')<sub>2</sub> fragments were immobilized on a sensor chip and dilutions of synthetic B2 domain and protein G (Pharmacia) injected over them. Table 11 shows the data for synthetic B2 domain

Table 11: Kinetic and Equilibrium Constants for Synthetic Protein G B2 Domain binding to goat F(ab')<sub>2</sub>, rabbit IgG antikappa chain, and rabbit IgG anti-humanF<sub>c</sub>.

Antibody	$k_{\text{ass}} \times 10^3 / \text{M}^{-1} \text{s}^{-1}$	$K_{\text{diss}} \times 10^{-3} / \text{s}^{-1}$	$K_d / \mu\text{M}$
goat F(ab') <sub>2</sub>	1.45	15.3	10.55
rabbit IgG antikappa chain	1.84	1.25	0.679
rabbit IgG anti-humanF <sub>c</sub>	1.48	1.29	0.872

---

Equilibrium Constants were determined by  $K_d = k_{\text{diss}}/k_{\text{ass}}$

Because the changes in the synthesized protein G analogues were mostly located in the Fab-binding loop between residues 10 to 18, we wanted to measure the binding to Fab alone and remove the F<sub>c</sub> binding component from the analysis.

The slope in Figure 121 represents the increase in binding and shows that the binding of the K15T16BTD B2 domain was much less than wt B2 domain on goat F(ab')<sub>2</sub>. The change in refractive index for the wildtype domain is 200 units, compared to 20 units for the analogue.

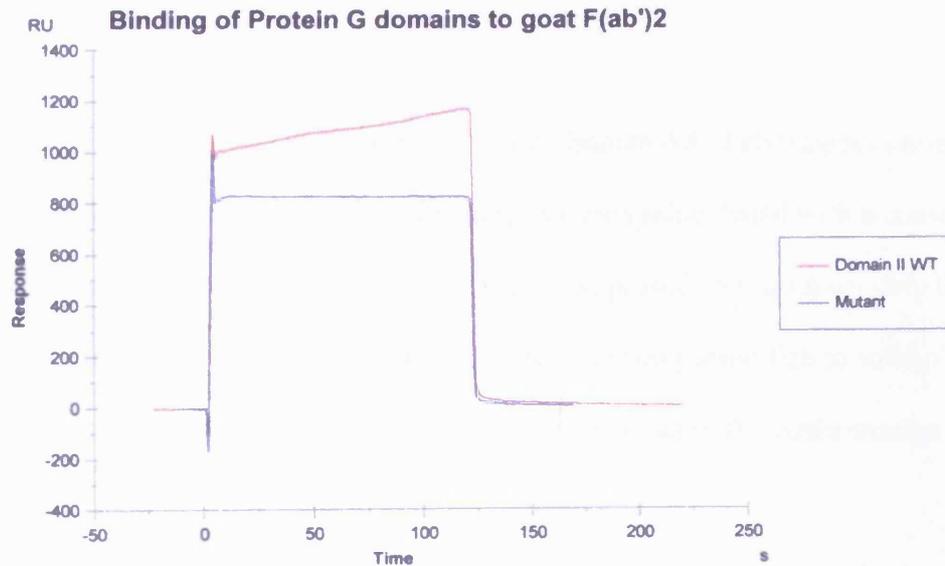


Figure 121.

After the large bulk effect, the response at the highest concentration is still very low for the analogue. The preliminary results indicated that the  $K_d$  of the analogues would be in the high micromolar to millimolar range. Therefore it was decided to use the solution affinity method rather than any direct method of determining binding constants because the affinities were lower than wildtype. It was also decided to measure the affinity of the analogues to a humanised antibody Fab fragment because there was no previous data on the affinity of protein G domains for human antibodies.

The antibody chosen was humanised A33 Fab fragment. The mAb A33 detects a membrane antigen that is expressed in normal human colonic and small bowel epithelium and > 95% of human colon cancers. It is absent from most other human tissues and tumour types. The murine A33 mAb has been shown to target colon cancer in clinical trials, and the therapeutic potential of a humanized antibody is currently being evaluated.

The affinity of the analogue was measured for a human A33 Fab fragment using the solution affinity method. Dilutions of analogue were preincubated with a constant concentration of human A33 Fab. The mixture was passed over a sensor chip that had protein G immobilized on it. The initial on rate of uncomplexed Fab to solid phase protein G was measured. This binding rate is proportional to the concentration of free Fab.

A standard curve for the binding of human A33 Fab to protein G is shown in Figure 123. The points for this curve were derived from the data in Figure 122. The highlighted sections of the curves are the data selected for initial binding rates, typically between 10 and 25s.

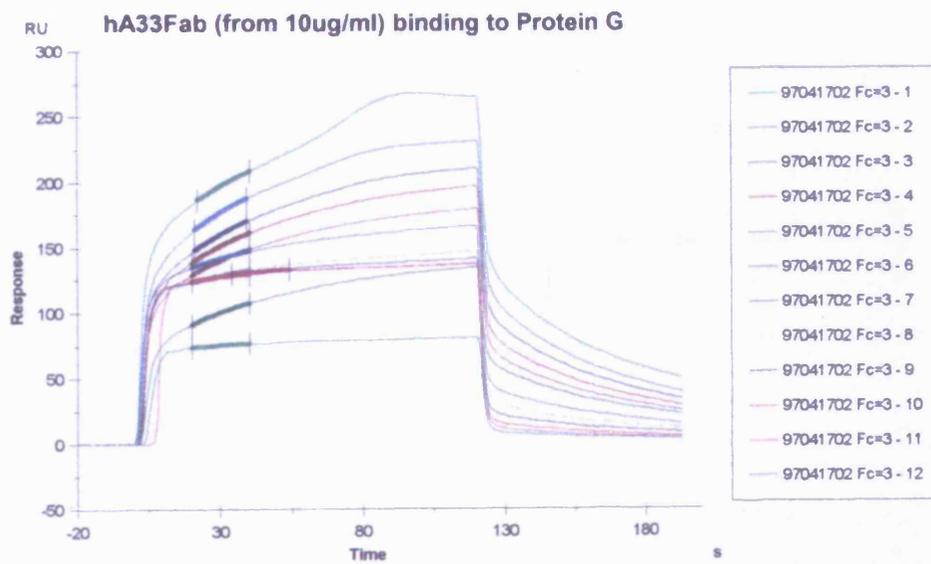


Figure 122.

Model: Linear  
Curve: r0 vs. C  
Data range: 9.99999939e-9 to 1.660000066e-7, 10 points  
R^2: 0.984  
LRSd: 0.0684

Constant	Value	Standard Error
Slope	9.49e6	4.31e5
icpt	0.154	0.039

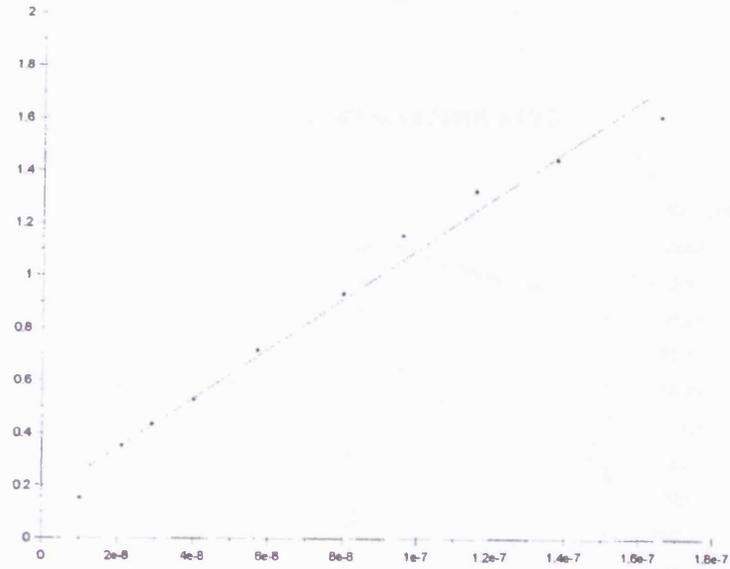


Figure 123.

Figure 124 shows the data for the effect of the 15BTD16 on human A33 Fab binding to immobilized protein G. Similar curves for the other analogues are given in Appendix 1.

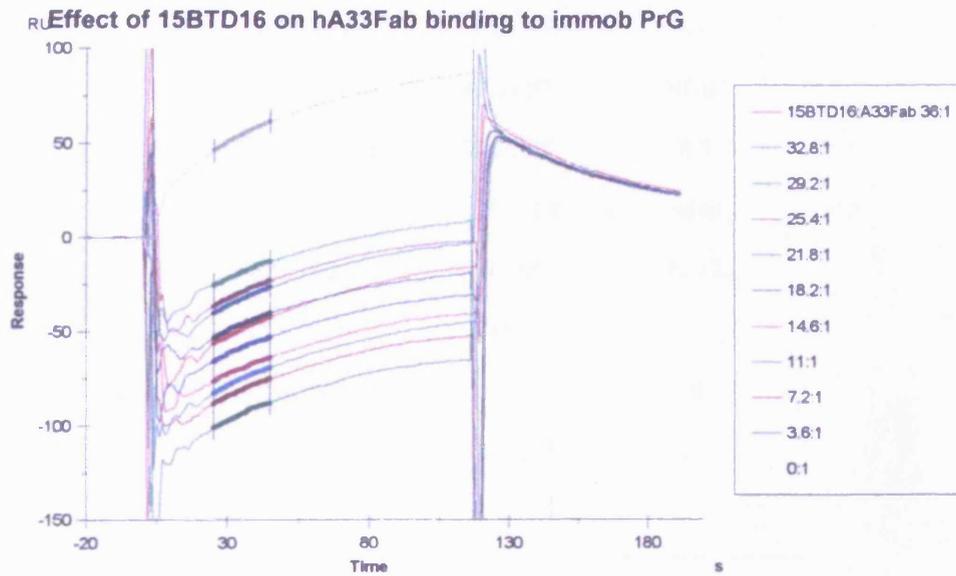


Figure 124.

The affinity was calculated by plotting the concentration of free Fab, derived from the initial binding rate, against the total concentration of the domains.

The calculated affinities are given in Table 12.

Table 12: Binding Constants for Synthetic Protein G B2 analogues against Human A33 Fab.

Domain	$K_d/\mu\text{M}$	$K_d(\text{mut})/K_d(\text{wt})$
Wt	35±7	
15BTD16	132±57	3.8
15Ahx16	45.9x10 <sup>6</sup>	NDB <sup>†</sup>
G14K15BTD	290±125	8.3
G14K15Ahx	25.58x10 <sup>6</sup>	NDB
K15T16BTD	4x10 <sup>6</sup>	NDB
K15T16Ahx	102±11	3
K15dPT16N-MeA	280±130	8
Q37Aib	251±227	7.1

The introduction of the rigid reverse turn mimetic BTD between residues 15 and 16 still binds to Fab, but the more “flexible” aminohexanoic acid does not. A possible explanation is the lack of contact surfaces for Van der Waal’s interactions with the alkyl chain.

Moving the BTD residue one residue along the chain from 14 to 15 abolishes binding. Moving the amino hexanoic acid along the chain from 14 to 15 recovers binding. The introduction of the unnatural amino acids D-proline, N-methyl alanine and aminoisobutyric acid does not abolish binding.

<sup>†</sup> NDB is no detectable binding

## Chapter 4: Discussion

### 4.1 Discussion

The B2 domain of protein G has 64 residues. The 3-D structure derived from NMR data is shown in Figure 125, with the NMR derived residues affected by Fab binding shown in blue.

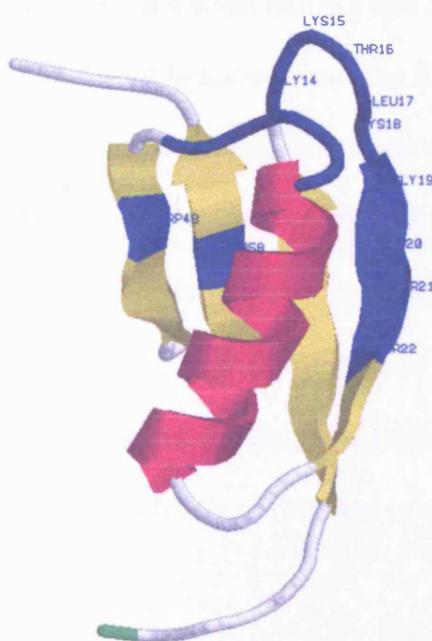


Figure 125. RasMol drawn representation of protein G B2 domain. The N-terminal leucine is shown in green.

The aim of this work was to use chemical synthesis to make analogues of protein G B2 domain that were not generally accessible by biochemical and to analyse the

effect of the introduction of substituents on stability and binding to the Fab fragment of a monoclonal antibody.

It was accepted at the beginning of this work that NMR data had shown that the N-terminal five residues were unstructured in solution, and made no difference to the overall stability of the domain. The analogues were therefore made from the 59-residue domain to make the syntheses easier, but for the sake of clarity the numbering system of the full sequence was maintained.

Protein G B2 domain was chosen because it is a small domain that had been made before by chemical synthesis. It was also used because it contains  $\beta$ -sheets and a helix, but no disulphide bonds. It is important biologically because of its high specificity for IgG. It is also important to know whether modifications to the side chains and backbone interfere with Fab binding and if they do, how the changes make this happen. A better understanding of the factors that contribute to the stability of proteins is very important because it will help in the design of proteins as drugs, vaccines or biosensors. It will contribute to advances in the design of proteins with novel sequences.

It was decided to make analogues by replacing the residues Gly14 to Thr16 with mimetics that constrained the turn and mimetics that made the region more flexible. Gly14 to Thr16 were chosen because they appeared to be in a  $\beta$ -turn.

The mimetic bithiazolidine dipeptide (BTD) (Figure 126) was used to constrain the turn

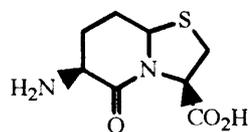


Figure 126. BTD mimetic.

and 6-amino hexanoic acid (Figure 127) was used to make the turn more flexible.



Figure 127.

These molecules were also used to make insertion analogues between residues 15 and 16. These analogues would be used to estimate how much of a perturbation the unnatural molecules were making on the structure. These insertion analogues are vital to the study because to make valid comparisons of stability, the wildtype structure must only be changed slightly. If drastic changes are made to either the folded or unfolded state then the studies are being made on two or more different structures and no valid comparisons can be made. If the insertions are well behaved, then the likelihood is that substitutions in the same region will only perturb the system to be measured.

These substitutions also alter the peptide backbone itself. This is important because site directed mutagenesis replaces side chains only, which represent only 40% of the

average protein mass. The other 60%, the backbone, is not altered by amino acid substitutions. BTD had been used to constrain turns in peptides, and its use to mimic a  $\beta$ -turn in HIV protease had been reported. In this latter work, no stability data had been reported. The reason for using this mimetic was as a constraint on the loop with no real concern for whether it was replacing a  $\beta$ -turn.

The amino hexanoic acid molecule was used as a spacer molecule in peptide synthesis methodology, and might make the loop more flexible, although it is possible that the alkyl chains might have a tendency to curl in on themselves in the secondary structure by expelling water cages. The turn residues were also replaced with the dipeptide D-ProN-MeAla, (Figure 128)

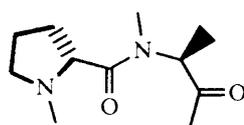


Figure 128.

For the helix, aminoisobutyric acid replaced the residue Q37. There were three reasons for this. The first is to investigate the synthetic issues. Aib, (Figure 129) is a sterically hindered amino acid which would be inserted in an ordered part of the structure.



Figure 129.

This would be a good test of the efficacy of the synthetic strategy. Secondly, Aib is a known stabilizer of helical structures so the result could be measured. Thirdly, the

analogue Q37A had already been made by mutagenesis techniques and stabilities could be compared. Figure 130 shows the structure of the B2 domain with the areas that were replaced by analogue structures in blue.



Figure 130. RasMol drawn representation of protein G B2 domain with G15 to T16 and Q37 shown in blue. The N-terminal residue is at the bottom.

Obtaining protein analogues by chemical synthesis is increasing in importance. The precision of stepwise addition of amino acids on to a suitable support matrix is now enough to allow the synthesis of proteins of about 100 amino acids. The incorporation of sterically hindered residues such as Aib and N-methyl alanine into protein sequences is not standard practice however. At the time of writing (December 2000) there were no examples in the literature of proteins being synthesized chemically with either secondary amino acids or Aib residues in the middle of the sequence. Also, the methods of purification of proteins synthesized chemically are long and complicated

with an attendant loss of product from dilution effects and absorption to purification supports.

To obtain sufficient quantities of homogeneous analogues containing Aib and N-methyl alanine to perform calorimetry, circular dichroism, denaturant unfolding, binding studies and NMR structure determination, two improvements were made to the standard methods of peptide synthesis. The first was to use HATU in place of HBTU as coupling agent. HATU, (Figure 131), is the 7-aza analogue of HBTU and the structure is shown. HATU had just been introduced as a coupling agent in the synthesis of small peptides but had not yet been evaluated in the synthesis of proteins.

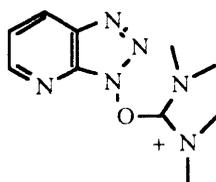


Figure 131. HATU

This reagent was less stable than HBTU and its side reactions had not been investigated to the same level as HBTU, but it was the only way to incorporate Aib residues into protein G B2 domain. HBTU could not add Aib in an efficient enough manner to give the required 30 mg yields of analogue.

The second improvement was to use perfusion chromatography in the POROS range of columns for purification of the domains. This flow through particle chromatography performed separations that were complete in two to five minutes without loss of capacity or resolution compared to conventional HPLC. Bulk purifications of large scale folding reactions that previously had taken days, were

accomplished in minutes. Electrospray mass spectroscopy and analytical HPLC were used to analyse the product.

A synthetic strategy that could make analogues of protein G B2 domain that contain any unnatural amino acid, however difficult to couple, was shown to be as follows: (a) peptide synthesis using HATU, (b) reverse phase HPLC, (c) folding by heating and rapid cooling, (d) perfusion chromatography, (e) analysis. Ten analogues and wildtype were made or purified using this method in isolated yields from 1% to 4% and purity that varied between 85% to 95%.

The most significant result from the synthetic point of view only, is the successful synthesis of an analogue containing an alpha disubstituted amino acid and one containing an N-methylated amino acid. These results extend the scope of solid phase peptide synthesis. It is now possible to make protein analogues containing any hindered amino acid or to have analogues containing a number of hindered residues in multi-milligram quantities in a few days.

The method's limitations lie in the final analysis. Electrospray mass spectrometry gives the mass of the product to 0.02%, but it is difficult to estimate purity.

Analytical HPLC was used for this, but it is important to have an orthogonal method of analysis. A synthetic protein should be analysed by a method that is different in mechanism to the method used to purify it. CZE would have been an orthogonal analytical method. Protein gels are the standard biochemical analytical tool, but were not used in this work because the material was not of biological origin and the impurities, being chemical in nature, would be best analysed by chemical methods only. The initial NMR studies on K15T16BTD show poor chemical shift dispersion

in the amide region of the spectrum. In the 2-D spectrum there appears to be a mixture of resonance characteristics. In the amide region, there are sharp well defined cross-peaks superimposed on broad, poorly defined cross peaks. This preliminary NMR data shows evidence of improperly folded material.

#### Wildtype protein G B2 domain

Table 14 shows the values for  $\Delta G_{\text{water}}(298\text{K})$  calculated for the wildtype molecule.

The values for CD and denaturant are from the equation  $\Delta G_{\text{water}}(298\text{K}) = mD_{1/2}$ . The calorimetric value is derived from the equation

$$\Delta G(298) = \Delta H_{\text{vh}}(1 - 298/T_m) - (\Delta C_p)(298 - T_m) + T \ln 298/T_m$$

where  $\Delta C_p$  is the literature value of 2.9 kJ/K.mol.

Table 14.

Exp.	$\Delta G_{\text{water}}(298\text{K})/$ kJmol <sup>-1</sup>
Calorimetry	22.70±0.02
CD	19±2
guanidine	14±1

This compares with the figure derived by Bryan of 19.5 kJ/mol. These results show that the wildtype is marginally stable at room temperature. This is consistent with the literature results on small globular proteins. The difference in the values derived from calorimetry and guanidine denaturation may be explained as a result of the high ionic strength of the guanidine solutions near the  $D_{1/2}$  values of 2M. The lower  $\Delta G$  values may reflect the destabilization of the protein at increased salt concentrations. This is

why urea might be preferred over guanidine in the denaturation studies. Guanidine was used in this study because urea decomposes over time in solution and it was necessary to be able to replicate experiments on different days without having to remake the dilutions. The guanidine denaturation data were fitted to an equation that assumes two state behaviour in the unfolding of protein G B2 domain and a linear dependence of  $\Delta G$  on denaturant concentration. These methods have been applied to other protein G analogues and are accepted literature methods for investigating the thermodynamics of small globular proteins. All the evidence in our denaturation curves suggests that the unfolding is a two state process.

Fifty-five data points were used in the analysis. However, there were few data points in the range 3.5M to 6.5M guanidine. This was a mistake. The denaturant concentrations should have been distributed equally between the maximum and minimum values to reduce fitting errors. This argument applies to all the domains.

15Ahx16 and 15BTD16

Figure 132 shows a schematic of the hairpin turn from protein G B2 domain.

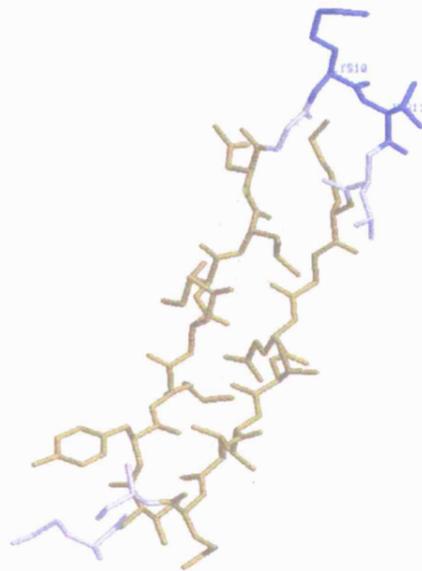


Figure 132. The hairpin turn in Protein G B2 domain.

The blue part of the structure shows the residues G14 to T16.

Table 15 shows the thermodynamic data for 15Ahx16 and 15BTD16.

Result	15Ahx16	15BTD16	WT
m value at pH7.6	9.1±0.1	8.9±0.3	8.9±0.3
m value at pH4.7	6.7±0.5	6.2±0.5	6.9±0.3
$\Delta\Delta G$ at pH4.7	-0.34±0.02	0.1±0.1	
hA33 Fab/ $\mu\text{M}$	None detected	132	35

The m values agree with the value for wildtype. This may mean that the unfolded states of the analogues closely resemble that for the wildtype. There is very little change in the stability of the analogues relative to wildtype. The inserted molecules appear to be well tolerated in the loop. There is a hydrogen bond between G14 and L17. The  $\Delta\Delta G$  values suggest that it may not be broken in the analogue sequences. If it is not, then the loop must be very flexible. The BTD analogue still can bind to the Fab but the Ahx insert has abolished binding. This suggests that the side chain contacts in the loop are important for binding and have been lost by the alkyl chain in the loop. NMR studies on the analogues would be needed to assess this. These results suggest that the turn in the hairpin loop is a good place for performing mutations because the insertions are so well tolerated. The result from the binding experiments suggests that the loop may also be a satisfactory probe for the contact points that affect binding.

## D-Proline N-Me Alanine Analogue and Q37Aib

D-Proline and N-methyl alanine were used to replace K15 and T16 respectively as shown in Figure 132. Aib was used to replace Q37 as shown in Figure 131.

Table 16 shows the thermodynamic data for these domains.

Result	D-P,N-MeA	Q37Aib	WT
m value at pH7.6	15±8	11.7±0.5	8.9±0.3
m value at pH4.7	7.7±0.7	8±1	6.9±0.3
$\Delta\Delta G$ at pH4.7	2.2±0.2	-1.0±0.1	
hA33 Fab/ $\mu$ M	280	251	35

The D-proline, N-Methyl alanine analogue is important because, to date, no protein analogue containing these residues has been published. The Aib analogue is important because this is the first time this residue had been introduced into a protein domain by chemical synthesis. The m values at pH 7.6 and 4.7 suggest that the denatured states of the analogues are similar to wildtype. The D-proline, N-Methyl alanine analogue is more stable than wildtype at pH 4.7 and still retains a binding affinity for the Fab. This is a very interesting result if a more thorough thermodynamic analysis involving CD and calorimetry is found to be in agreement. The 3-D structure should be determined in solution, both on its own and in the presence of the Fab, if the affinity will allow. The D-proline, N-Methyl alanine analogue must still have some favourable contact points with the Fab. This molecule would be a good starting point to examine in detail the conformation of the D-proline,

N-Methyl alanine dipeptide in the context of the structure of the B2 domain and how this relates to the binding to Fab.

It is also interesting to note that the Q37 was not explicitly identified as a residue affected by binding to Fab in Lian's study [114]. Q37Aib is marginally destabilized with respect to wildtype and its binding to Fab is the same as G14K15BTD and D-proline, N-Methyl alanine analogue. Again, NMR studies would be needed to fully assess the effect of Aib on binding.

#### G14K15Ahx and G14K15BTD

In these analogues, amino hexanoic acid and BTD were substituted for G14 and K15 (see Figure 132). Table 17 shows the thermodynamic data.

Table 17.

Result	G14K15Ahx	G14K15BTD	WT
m value at pH7.6	37±3	6.4±0.1	8.9±0.3
m value at pH4.7	5±5	8±2	6.9±0.3
m value from CD	15±7	10±2	10±1
ΔΔG from calor.	-1.67±0.01	-4.31±0.02	
ΔΔG from CD	4±2	-2.7±0.5	
ΔΔG from guan.	3±3	-1±0.25	
hA33 Fab/μM	-	290	35

The m values for the two analogues agree reasonably well with those for the wildtype except for G14K15Ahx at pH 7.6. The m value is four times higher than the value for

wildtype. Assuming the same argument concerning the changes in  $m$  values for staph nuclease are applicable in this case, this result implies that there is less structure in the denatured state of the analogue than for the wildtype. There may be more accessible residues at this pH because the amino hexanoic acid is so much more flexible than the amide backbone. Interestingly, the K15T16Ahx analogue also has the same  $m$  value at this pH as G14K15Ahx. The  $m$  value for G14K15BTD is lower than the value for wildtype. This may imply that there is more structure in the denatured state of the analogue compared to wildtype, because the BTD's rigidity in the denatured state means that parts of the protein are now less accessible to denaturant.

The  $\Delta\Delta G$  values for the Ahx analogue differ widely. Guanidine denaturation is monitored by tryptophan fluorescence. This residue is in the hydrophobic core of the domain and the fluorescence monitors the tertiary structure. Calorimetry is measuring changes in tertiary structure. CD is monitoring changes in secondary structure, mainly helical content. If the denatured state of the analogue is very different in character to the wildtype's, then the differences in the values observed may result from a difference in the end points that are observed in the three experiments. The domain may have different denatured states as a result of heat denaturation compared to guanidine.

The Ahx mutation also abolishes binding to Fab, although the same mutation moved one residue along the chain in K15T16Ahx has an affinity for Fab at  $102\mu\text{M}$ .

The BTD analogue is destabilized with respect to wildtype as determined by all three methods. Both mutations remove the possibility of hydrogen bonding between G14 and L17 as shown in Figure 132. In the Ahx analogue the alkyl chain may be flexible

enough to allow other interactions to compensate. This is probably not possible in the BTM mutation because of the rigidity of the bicyclic ring. This may be the source of the instability of the BTM analogue. The values for a single hydrogen bond from the literature are of the order of  $8\text{kJmol}^{-1}$ . The BTM analogue still binds to Fab with an affinity of  $290\mu\text{M}$ . As for the other analogues, NMR studies would be needed to evaluate the source of the binding to Fab.

### K15T16Ahx and K15T16BTM

In these analogues, amino hexanoic acid and BTM were substituted for K15 and T16 (see Figure 132). Table 18 shows the thermodynamic data.

Table 18.

Result	K15T16Ahx	K15T16BTM	WT
m value at pH7.6	35.5	$27\pm 8$	$8.9\pm 0.3$
m value at pH4.7	$7.4\pm 0.7$	$7\pm 1$	$6.9\pm 0.3$
m value from CD	$4\pm 2$	$11\pm 4$	$10\pm 1$
$\Delta\Delta\text{G}$ from calor.	$-2.96\pm 0.01$	$-2.53\pm 0.02$	
$\Delta\Delta\text{G}$ from CD	$0.4\pm 0.2$	$-1.5\pm 0.5$	
$\Delta\Delta\text{G}$ from guan.	$3.4\pm 0.3$	$-0.1\pm 0.01$	
hA33 Fab/ $\mu\text{M}$	102	-	35

The m values at pH 7.6 are very different to that for the wildtype. Using the same arguments as for the G14K15 analogues, there may be less structure in the denatured

state compared to wildtype. With the BTD moved to the crown of the loop, the peptide chain at the N-terminal may be less restricted than BTD at G14K15.

For the Ahx analogue there are differences in sign from the thermodynamic results similar to the results for G14K15Ahx. The  $m$  value at pH4.7 is in good agreement with the value for the wildtype. This may suggest that the denatured states of both are similar. The differences in sign may be accounted for by the difference in endpoints for the denaturant unfolding and the heat unfolding. Interestingly, in this case it is the Ahx analogue that still binds to Fab and the BTD mutation that abolishes it. NMR experiments are needed to determine the contact areas that result in binding or the lack of binding in these analogues.

## 4.2 Conclusions

It is now possible to make analogue protein domains with any amino acid residue, no matter how hindered, anywhere in the sequence, and so extend the scope of protein stability studies.

The most effective way of achieving the levels of homogeneity required for thermodynamic studies is to use methods of analysis that are orthogonal to the purification methods. It is not acceptable, for example, to use ion-exchange methods to analyze a protein produced from an ion-exchange purification protocol.

There is a lack of correlation in the  $\Delta\Delta G$  values for the analogues and a variation of up to  $\pm 50\%$  in the value of  $m$  compared to wildtype for some analogues. There are several ways of viewing this preliminary stability data.

Firstly, the linear extrapolation to zero denaturant concentration should not be applied to all of the analogues because some of the substitutions are significantly altering the analogue's interaction with guanidine. The large changes in  $m$  values can reasonably be interpreted as a reflection of altered interactions.

Shortle [145] made fragments of staphylococcal nuclease and showed that these fragments and the corresponding mutations, in terms of their stability to denaturants, can be considered to form two types of unfolded states. Mutations, which showed an increase in  $m$ , showed evidence of less structure in the fragments, while those that showed a decrease in  $m$  showed evidence of more structure. Assuming we can apply the principle to the protein G analogues, it may be concluded that the substituents are changing the denatured state as well as the native state. This implies that there is residual structure in the denatured states of the analogues. The aminohexanoic acid

substitutions give a 50% increase in  $m$  value for the G14K15 analogue and a 50% decrease for the K15T16 analogue. It is concluded that the former analogue has less structure in the denatured state than wildtype and the latter has more structure. In Shortle's analysis [145], he observed that there was a significant correlation between changes in stability and changes in the value of  $m$ . This strengthens the conclusion that the putative structural changes in the denatured state which lead to increases or decreases in  $m$  are responsible for a significant fraction of the stability loss for an average mutant. There is no such relation in the protein G analogues as the  $\Delta\Delta G$  values derived from denaturant analysis show an increase in stability for both analogues.

It is also concluded, that because of the variation in the  $\Delta\Delta G$  values for the analogues, that differential scanning calorimetry should be the main source of thermodynamic results. It is evident that these analogues are not unfolding to a single reference state. This makes it difficult to answer the question: "Is the change in stability resulting from a change in the free energy of the folded state or a change in the free energy of the unfolded state or both?" Secondly, the analyses of the data from CD and guanidine experiments generally assume a two-state model. The accuracy of the thermodynamic results and the validity of their interpretation are critically dependent on the validity of this assumption. Differential scanning calorimetry measures energy changes directly and so is much less ambiguous. Uniquely to DSC, a second and independent estimate of the unfolding enthalpy may be made from van't Hoff analysis of the shape of the peak in the thermogram. This value may be compared to the calorimetric enthalpy to determine whether the simple two-state assumption of unfolding has broken down.

It is also concluded that the solution affinity method of measuring binding affinities to Fab fragments is acceptable for measuring how well the analogues bind to antibodies. The method requires micromolar quantities of protein and can measure millimolar equilibrium constants. Most steps can be automated and measurements can be made against a range of antibodies and their fragments.

**Suggestions for further work.**

Use DSC on all the domains to get the most complete thermodynamic data possible.

Use other peptidomimetics from peptide chemistry to make more domains with substitutions in the surface loops, helix and hydrophobic core. Use HATU for the synthesis where hindered residues are to be introduced. Use an orthogonal means of analysis ,for example, CZE, on all new domains, to be assured of homogeneity.

Obtain NMR structures on all new domains and build up a database of thermodynamic and structural data. Finally use NMR to obtain the contact points between the domains and Fab to build more stable proteins.

## Chapter 5: Experimental

### 5.1 Materials and Methods

#### Reagents

The term “water” throughout refers to MilliQ<sup>®</sup> water. Fisons HPLC grade acetonitrile was used. Diethyl ether was from BDH. All peptide syntheses were carried out under argon. Removal of protecting groups and detachment from the solid support was carried out under nitrogen. Piperidine, N-Methylpyrrolidone, dichloromethane, O- (Hydroxybenzotriaz-1-yl)-1,1,3,3, -tetra methyl uronium Hexafluorophosphate and diisopropylethylamine were, obtained from Applied Biosystems as peptide synthesis grade and used without further purification: Trifluoroacetic acid, thioanisole, thiophenol, dimethyl sulphide, ethanedithiol, collidine, triisopropylsilane and phenol were obtained from Aldrich. Guanidine and urea were obtained from Sigma.

Fmoc amino acid derivatives were obtained from Applied Biosystems Ltd. Side - chain protection was t-butyl ethers and esters for Ser, Thr, Tyr, Asp and Glu; Boc for Lys; Pmc for Arg; trityl for Asn, Gln, and Cys. Fmoc-cys(acm)-OH. 7-Azahydroxybenzotriazole and O- (7-Azahydroxybenzotriaz-1-yl)-1,1,3,3, -tetramethyluronium hexafluorophosphate were obtained from Millipore (UK).

Gel filtration was performed on pre-packed PD-10 columns containing pre-equilibrated Sephadex<sup>®</sup> G-25. The columns were equilibrated with 25 ml of buffer. Samples were applied in 2.5 ml of buffer and eluted with 3.5 ml of buffer.

Amino acid analyses were carried out at Alta Biosciences, University of Birmingham. Edman degradations were carried out in house on an Applied Biosystems 470 protein sequencer using the Prospin<sup>®</sup> sample preparation cartridge technique. This uses a

Problott PVDF membrane attached to an insert in contact with a 3000 Da cut off filter which concentrates the protein and increases the amount absorbed into the hydrophobic membrane. Contaminants and salts are removed without sample loss by simple washing. The PVDF membrane is wetted with 25 $\mu$ l of methanol. The excess methanol is removed before adding the sample. After adding the sample  $\approx$ 200 pmol, the insert cap is closed and the sample is centrifuged to dryness. This is important for optimal protein binding. The membrane is removed by punching it out and putting it into a clean Eppendorf containing 20% methanol in water. The disc is carefully inserted into the sequencing Blot Cartridge and leak tested. The sample is sequenced for the required number of cycles of Edman degradation using BGN470-1 programs for the initial Reaction and Conversion cycles followed by BLOTT1 and RUN470-1, respectively. The HPLC separation of the PTH amino acids was carried out using Program RUN470-1. The HPLC buffer contains 5% N,N-diisopropylethylamine, 10% water and 85% methanol. CZE was performed on an Applied Biosystems 270A-HT system, using a silica capillary and 20mM citrate pH 2.5 as running buffer, in free solution mode. The standard conditions used were as follows: a one minute flush of the capillary with 0.1M sodium hydroxide, a two minute equilibration with running buffer; a vacuum injection of sample of between three and ten seconds (depending on the concentration of the sample); ten to twenty minutes with running buffer at 25 kV, temperature 30°C with UV wavelength detection at 200nm.

HPLC was carried out using an Applied Biosystems 151A system, incorporating two 1406A solvent delivery systems and a 783A programmable absorbance detector. The injection system was a Rheodyne<sup>®</sup> 7125 valve with 5.0-ml sample loop. The columns used were preparative scale reverse phase C<sub>4</sub>, C<sub>8</sub>, and C<sub>18</sub> Dynamax 3000 nm pore

size, run at 10 ml/min. C<sub>4</sub>, C<sub>8</sub> Aquapore<sup>®</sup>RP-300 and Aquapore<sup>®</sup> weak cation exchange columns were run at 8 ml/min. The wavelength for detection of the peptide bond was 225 nm. Gradients for reverse phase separations were made from increasing percentages of HPLC acetonitrile/0.05% TFA in 0.1% TFA/water. Wavelengths for detection were 225 nm for the peptide bond or 280 nm for tyrosine and tryptophan. Gradients for cation exchange chromatography were made from increasing percentages of 1M sodium chloride/ 20 mM sodium phosphate pH 6.0 in 20mM sodium phosphate pH 6.0. Anion exchange chromatography was carried out using Mono Q, Resource Q, Synchronpak AX-100, and Poros HQ/M columns. Buffer A was 20 mM TRIS pH 7.6. Buffer B was 1M NaCl in 20 mM TRIS pH 7.6.

#### Chain Assembly

Peptide acids were assembled from the C- to N-terminal on scales from 0.1 mmol. to 0.25 mmol using a crosslinked polystyrene support containing the C-terminal residue linked to the resin through a HMP ester linkage (the so-called "Wang" resin). Peptide amides were assembled on the MBHA resin through a 4-(2', 4'-dimethoxyphenyl-Fmoc-aminomethyl)-phenoxyacetamido-norleucyl ester. The chain assembly was carried out in a fully automated fashion on an Applied Biosystems 430A synthesizer in an argon atmosphere. An n optimised set of synthetic cycles was developed. These were based on the standard protocols supplied with the instrument. Removal of the N-terminal Fmoc group was achieved with between two to a maximum of four treatments with 20% piperidine in NMP for times varying between three and ten minutes depending on the difficulty of coupling. The resin was then washed with NMP (5x10s). A routine single coupling protocol was used with a coupling time of at

least thirty minutes up to a maximum of one hour. A second coupling step was incorporated in cases of difficult sequences. Protected amino acids were incorporated as active esters in two ways. The first method derived the esters from HBTU in the presence of two equivalents of DIEA and one equivalent of HOBT. The reaction mixture was added to the resin after seven minutes of activation. The second method derived the esters from HATU in the presence of two equivalents of DIEA. The reaction mixture was added to the resin after two and a half minutes of activation. At the end of the assembly the N-terminal Fmoc group was removed with 20% piperidine in NMP and the resin dried under argon.

#### General Deblock for Peptides attached to Acid Labile Handles

The three types of resins used were the HMP also known as the Wang resin; the p-hydroxymethylphoxymethyl grafted polystyrene-polyoxyethylene (Novasyn<sup>®</sup>TGA); and the [4-(2', 4'-dimethoxyphenyl-Fmoc-aminomethyl)-phenoxyacetamido-norleucylaminomethyl resin] or Rink amide AM resin.

The dried resin was placed in a round-bottomed flask equipped with magnetic stirring bar. The appropriate cleavage solution was added slowly with stirring. After an elapsed time that was dependent on the particular sequence, the resin was removed by filtration under reduced pressure. The resin was washed with a solution of 95% TFA, 5% water. The filtrate was partially evaporated *in vacuo* to an oil using a rotary evaporator equipped with a dry-ice trap in a fume hood. The residue was co-evaporated with ether three times to remove residual TFA. Approximately 500 to 900 ml. of diethyl ether at -20°C was added with swirling. The resulting white precipitate

was allowed to settle. The ethereal layer was decanted and the residue evaporated to dryness. The ether addition, decantation and evaporation were repeated a further three times. Spotting on to a TLC plate and spraying with ninhydrin checked the ether layer. The latter gives a vivid blue colour when a free amine group is present.

The powder was evaporated *in vacuo* to dryness. The powder was then dissolved in a solution of 0.2M TRIS-HCL pH 8.6 containing 8M guanidine and the pH readjusted to 8.6. It was left to stand for thirty minutes to reverse any N to O shift of serine or threonine. If the peptide contained cysteine residues 80 mg. of DTT per cys was added and the mixture allowed to stand for a further 30 minutes. The mixture was then acidified with concentrated acetic acid to pH 4, filtered under reduced pressure and chromatographed on Sephadex<sup>®</sup>G-25 packed PD-10 columns, eluting with 20% acetic acid, 2.5 ml sample and 3.5 ml. eluant. The fractions containing peptide were pooled and lyophilised to a white amorphous powder and stored at -20°C.

<sup>1</sup>H NMR spectra were obtained using Bruker AX 300 and AX600 spectrometers. MS analyses were carried out at the following locations: the University of Nottingham on a Kratos plasma desorption machine; the University of Aberdeen on a Vestec<sup>®</sup> laser desorption spectrometer and in house on an electrospray mode spectrometer.

#### Matrix-assisted laser desorption

For MALD-TOF, the samples were dissolved in 0.1% TFA and diluted by a factor of between 10 and 1000. The matrix used was  $\alpha$ -cyano-4-hydroxycinnamic acid as a saturated solution in 35% acetonitrile/65% 0.1%TFA. 1.5 $\mu$ l of the matrix solution and 1.0 $\mu$ l of the peptide solution were deposited on the x, y plate and air-dried at room temperature. The measurements were acquired in the linear mode on a VESTEC<sup>®</sup>

instrument. The acceleration voltage used was 30kV. A nitrogen laser (337 nm, 3ns pulse) was used with a laser spot size of approx. 30  $\mu\text{m}$ . Several shots were accumulated for each spectrum.

## Electrospray

Spectra were acquired using a Micromass Quattro I, triple quadrupole mass spectrometer. Samples were analysed by infusing the sample in a 50:50 MeOH:H<sub>2</sub>O, 4% formic acid solution, via a syringe pump at 5  $\mu\text{l}/\text{min}$ , and acquiring by positive mode electrospray ionisation. Data was acquired from 600 to 1800 amu in 4s at a cone voltage of 40 - 60 volts.

The multiply charged raw data were deconvoluted using the Micromass MaxEnt software (Vn.2.1)

## Circular Dichroism.

Circular dichroism spectra were measured with an Aviv Associates Model 61DS spectropolarimeter containing a thermostable cell holder at the University of Stirling under the direction of Prof. Nick Price. All measurements were made at pH 7.0 and are expressed as mean residue ellipticity ( $q$ ), in units of  $^{\circ}\text{cm}^2\text{dm}^{-1}$ . Peptide concentration was obtained by UV absorbance using the relation  $A_{280\text{nm}}(1\text{mg/ml})=1.32$ . The sample concentrations used are given on the relevant traces but are typically about 20 $\mu\text{M}$ . The cell path length varied between 0.2 and 2 cm. Single scans were obtained by taking points every 0.5 nm with a 4s time constant and a bandwidth of 1.5 nm. Baseline corrected spectra were smoothed using a third-order least squares

polynomial with a window size between 6 and 15 points. The absolute ellipticity of the instrument was calibrated using d-10-camphorsulphonic acid at 290.5 nm and the wavelength calibrated to within 0.1nm using benzene vapour.

### Guanidine denaturation

A Perkin-Elmer LS-50 luminescence spectrometer was used, equipped with a plate reader accessory and interfaced to a personal computer running plate scan software. For protein G B2 domain analogues, a five fold change in fluorescence was obtained with an excitation wavelength of 280 nm and an emission wavelength of 350 nm. Some experiments were run with an excitation wavelength of 295 nm and an emission wavelength of 350 nm. All experiments were performed at 25°C. The emission bandpass was 10 nm. The guanidine hydrochloride solutions were prepared by aliquoting appropriate aliquots using Gilson pipettes of a solution of 4M guanidine containing 20 mM TRIS for pH 7.6 solutions and 50 mM citrate, 100 mM phosphate for pH 4.7 solutions. The 4M guanidine solution had been prepared by diluting an 8M solution (Pierce Sequanal grade) with the appropriate buffer. For each data point in the experiment, 50  $\mu$ l of protein in 20 mM TRIS for pH 7.6 or 50 mM citrate, 100 mM phosphate for pH 4.7, were diluted into 200 $\mu$ l of the appropriate denaturant concentration using a Gilson pipette, into the wells of a fluorescence assay plate. This produced final concentrations of protein of 30 $\mu$ M. Concentrations of guanidine were in the range 0 to 6.5M. The solutions were equilibrated on a shaker for one hour before measuring their fluorescence. Fifty-five data points per experiment were collected for analysis. Analysis of the denaturation curve was performed using the

two-state unfolding model and the linear extrapolation method. The two methods were combined into a single equation to describe the shape of the denaturation curve:

$$F = \frac{F_n(0) + m_n[D] + (F_u(0) + m_u[D]) \exp \frac{m([D] - [D_{1/2}]}{RT}}{1 + \exp \frac{m([D] - [D_{1/2}]}{RT}}$$

where F is the observed fluorescence (after allowing for the fluorescence of guanidine and buffer),  $F_n$  and  $m_n$  are the intercept and slope, respectively, of the pre-transition baseline,  $F_u$  and  $m_u$  are the intercept and slope of the post-transition baseline, m is the dependence of the free energy of unfolding on guanidine concentration, and  $D_{1/2}$  is the midpoint of the denaturation curve. The experimental points were fitted using Kaleidagraph™ software. The free energy of unfolding  $\Delta G_{\text{water}}$  (defined as the conformational stability of the protein) is equal to  $m[D_{1/2}]$ . By calculating  $\Delta\Delta G$  from the difference in midpoints, a long extrapolation back to 0M guanidine is avoided.

### Differential Scanning Calorimetry

DSC measurements were performed on a Microcal VP-DSC at the University of Glasgow under the direction of Prof. Alan Cooper. The temperature was increased at a rate of 60 K/h on all samples. The power input from an upward scan of buffer vs. buffer was subtracted from the first upward scan to obtain the excess power input for the unfolding transitions. The excess power thermal scans were converted to excess

heat capacity vs. T scans by dividing by the scan rate. The scans were carried out in 0.1M acetate buffer pH 4.7. Sample concentrations were determined using the relation  $A_{280\text{nm}}(1\text{mg/ml}) = 1.32$ . Protein concentrations were between 0.25mM and 0.45mM, corresponding to 1.5 to 3 mg/ml.

## NMR

NMR experiments were performed on Bruker AMX600 and AM500 spectrometers at the NMR centre, University of Leicester, under the direction of Dr. Lu-Yun Lian. The proton resonance frequencies were 600.13 and 500.13 MHz, respectively. The NMR measurements were made at 20°C and 25°C. Homonuclear Hartmann-Hahn spectroscopy (HO-HAHA) and NOe spectroscopy (NOESY) experiments were performed in water and DMSO at 295K, 305K, 310K and 315K. Spectra were acquired in the phase sensitive mode by time-proportional phase incrementation of the first pulse (TPPI). Mixing times for the HO-HAHA and NOESY experiments were 100 ms and 300 ms respectively. Mixing times for the TOCSY experiments were 80 ms. Data were processed on-line on a Bruker Aspect 3000 data station and the delineation of the spin systems and NOESY signals was accomplished according to the method of Wurtrich.

## Binding Studies

Binding experiments were performed on a BIAcore 2000 (Pharmacia Biotech AB, Uppsala, Sweden) under the direction of Dr. Lesley Chaplin and Dr. Shauna West at Celltech. Experiments were carried out at 298K in 10mM HEPES buffered saline

with 150mM sodium chloride, 3.4mM EDTA and 0.005% Surfactant P20 (Pharmacia). Protein binding to human A33 Fab was analysed by preincubating dilutions of analogue with a constant concentration of hA33 Fab. The mixture was passed over a sensor chip to which protein G had been immobilized. The initial on rate of uncomplexed Fab to solid phase protein G was then measured. This on rate is proportional to the free concentration of Fab. The concentrations of free hA33 Fab were plotted against the total concentration of the analogue using the BIAcore solution affinity model to calculate the affinity.

## 5.2 Experimental

V3 Loop of gp120.

CTRPNNNTRKSIRIQRGPGRAFVTIGKIGNMRQAHC.

The sequence was assembled on an Applied Biosystems 430A automated peptide synthesizer by the Fmoc/t-butyl strategy. The machine cycles are shown in Table 19. 4-(2',4'-Dimethoxyphenyl-Fmoc-aminomethyl)-phenoxy polystyrene resin(1.21g, 0.41 mmol/g, 0.5 mmol) was used in reaction column A. The N-terminal Fmoc group was removed as shown in Table 19. Chain assembly was achieved by reacting the protected amino acid with one equivalent of HOBt and DIPCDI in DMF, waiting 34 minutes for activation at room temperature and adding the mixture to the resin. Side chain protection was as follows: Trt (Cys, Asn and Gln); t-Bu (Ser, Thr); Mtr and PMC for Arginine; Boc (Lys); and none (Ile, Pro, Ala, Gly, Phe, Val). Each residue was triple coupled for thirty minutes per coupling. After completion of assembly, the N-terminal Fmoc group was removed and the resin dried under argon. The resin beads were red in picric sulphonic acid solution.

Table 19. Diisopropylcarbodiimide, Hydroxybenzotriazole, DMF Coupling Cycle for the Solid Phase Synthesis of the V3 loop.

Step	Reagents	Time/min.
(i)	20% piperidine in DMF	3, 15
(ii)	DMF wash	10 x 0.5
iii	2 equivalents of Fmoc-AA-HOBt ester from DIPCDI	30
iv	DMF wash	10 x 0.5
v	2 equivalents of Fmoc-AA-HOBt ester from DIPCDI	30
vi	DMF wash	10 x 0.5
vii	2 equivalents of Fmoc-AA-HOBt ester from DIPCDI	30
viii	DMF wash	10 x 0.5

The dried peptide-resin was packed into a small glass sintered column and eluted with a mixture of thiophenol (5 ml), dimethyl sulphide (5 ml), water (4 ml), thioanisole (5 ml), ethanedithiol (20 ml) and trifluoroacetic acid (76 ml). The red mixture was stirred at room temperature for ninety minutes under nitrogen, then evaporated *in vacuo* to an oil. The mixture was co-evaporated with ether three times to remove residual acid and finally, chilled ether (500ml) was added and the white precipitate allowed to settle. The ether layer was decanted and the residue evaporated *in vacuo*. The ether washing steps were repeated a further four times. The precipitate was then evaporated to dryness. 20 mM TRIS/ 4M guanidine pH 8.3 (50ml) was added to the white powder in an ice bath and the mixture stirred for thirty minutes. DTT (100mg) was then added and the mixture left for a further thirty minutes. Glacial acetic acid was then added to pH 4 and the mixture partially evaporated *in vacuo*. The residue was taken up in 10% acetic acid, filtered to remove insoluble gel-like material and gel filtered on Sephadex G-25 with 10% acetic acid as eluant. 3.5 ml fractions were collected, water added and the mixture lyophilised to a white powder.

#### Disulphide bridge formation

Water (1L), ammonium bicarbonate (7.6g) and EDTA (840mg) were added to the powder (200mg). Oxygen was then bubbled through the cloudy mixture for two hours and the mixture allowed to stand at 4°C overnight. The mixture was then carefully acidified to pH 4 with glacial acetic acid and partially evaporated *in vacuo*. The water bath temperature did not exceed 40°C. The residue was taken up in 10% acetic acid and gel-filtered on Sephadex G-25 columns using 10% acetic acid as eluant. 3.5 ml fractions were lyophilised to a white powder.

The crude material was purified on HPLC ion exchange and reverse phase columns with CZE monitoring of the reverse phase fractions according to the gradients set out in Tables 20, 21 and 22.

Table 20. HPLC Ion-Exchange Conditions

Column	30 nm Aquapore Carboxy Methyl 10 x 100mm
Buffer 1	24 mM phosphate/ 10% acetonitrile pH 6.2
Buffer 2	24 mM phosphate/ 10% acetonitrile/ 1M sodium chloride pH 6.2
Flow Rate	8 ml/min.
Detection	225 nm UV
Gradient	0 to 50% buffer 2 over 40 minutes

Table 21. HPLC Reverse-Phase Conditions

Column	30 nm Dynamax Cg 2.5 x 25cm
Buffer 1	0.05% TFA in water
Buffer 2	0.05% TFA in acetonitrile
Flow Rate	12 ml/min.
Detection	225 nm UV
Gradient	10 to 30% buffer 2 over 40 minutes

Table 22. CZE Conditions

Column	Silica
Buffer 1	20 mM citrate pH 2.5
Detection	200nm UV
Gradient	Isocratic

After lyophilisation of the fractions the yield was 20 mg of homogeneous material.

The percentage of free thiol in the preparation was determined to be <0.2%, by reaction with dithiodipyridine.

**Amino acid analysis:** Ala 2.1, Arg 5.93, Asx 4.10, Glx 2.05, Gly 4.26, His 0.94, Ile 3.96, Phe 0.89, Pro 3.04, Thr 2.89, Ser, 0.98, Val 0.98, Met 1.00, Lys 1.98.

**Edman degradation:** Sequence analysis confirmed the first twenty-five residues.

**Laser Desorption MS:** (M + H)<sup>+</sup> 4049.4 (Calc. 4049 from sequence)

Trypstatin

IAACN LPIVQ GPCRA FAELL AFDAQ QGKCI QFIYG GCKGN NNKFY  
SEPKC KWCYCG VPGDG Y

The chain assembly was carried out on an Applied Biosystems 430A automated peptide synthesizer by the Fmoc/t-Bu strategy. The machine cycles are shown in the Table 23. The starting resin used in column A was 4-hydroxymethyl-phenoxy methyl polystyrene (128mg, 0.12 mmol of 0.94 mmol g<sup>-1</sup>). The first amino acid was coupled to the resin as a symmetrical anhydride formed from DIPCDI with a 1% DMAP as catalyst for one hour. The N-terminal Fmoc group was removed as shown in Table 23. Chain assembly was achieved by reacting the protected amino acid with

one equivalent of HBTU/DIEA in NMP, waiting 5 minutes for activation at room temperature and adding the mixture to the resin, final concentration 0.3M. Side chain protection was as follows: Trt (Cys, Asn and Gln); t-Bu (Ser, Thr, Tyr); PMC for Arginine; Boc (Lys); and none (Ile, Pro, Ala, Gly, Phe, Val, Trp). Each residue was single coupled for thirty minutes per coupling. After completion of assembly, the N-terminal Fmoc group was removed and the resin dried under argon. The resin beads were red in picric sulphonic acid solution.

Table 23. HBTU/ DIEA /NMP Coupling Cycle for the Solid Phase Synthesis of Trypstatin.

Step	Reagents	Time/min.
i	20% piperidine in NMP <sup>a</sup>	3, 15
ii	NMP wash	10 x 0.5
iii	8.3 equivalents of Fmoc-AA-HOBt ester from HBTU	30
iv	NMP wash	10 x 0.5

<sup>a</sup> The final N-terminal Fmoc group is removed by treatments of 40% piperidine in NMP for 7 and twenty minutes.

Phenol (4 ml), p-cresol (4 ml) and dimethylsulphide (4 ml) were added to the resin in a 1.0l roundbottomed flask under nitrogen. A mixture of thiophenol (6 ml), trifluoroacetic acid (76 ml), ethanedithiol (20 ml) and water (4 ml) was added and the mixture stirred at RT for 2 hours. The resin was filtered and washed with TFA. The filtrate was evaporated *in vacuo* and co-evaporated with ether. The mixture was co-evaporated with ether three times to remove residual acid and finally, chilled ether (500ml) was added and the white precipitate allowed to settle. The ether layer was decanted and the residue evaporated *in vacuo*. The ether washing steps were repeated a further four times. The precipitate was then evaporated to dryness. 50 mM TRIS/6M guanidine pH 8.3 (50ml) was added to the white powder in an ice bath and the mixture stirred for thirty minutes. Dithiothreitol (100mg) was then added and the mixture left for a further thirty minutes. Glacial acetic acid was then added to pH 4 and the mixture partially evaporated *in vacuo*. The residue was taken up in 20% acetic acid, filtered to remove insoluble gel-like material then gel filtered on Sephadex G-25 with 20% acetic acid as eluant. 3.5 ml fractions were collected, water added and the mixture lyophilised to a white powder.

Table 24. HPLC Reverse-Phase Conditions

---

Column	30 nm Dynamax C8 2.5 x 25cm
Buffer 1	0.05% TFA in water
Buffer 2	0.05% TFA in acetonitrile
Flow Rate	12 ml/min.
Detection	225 nm UV
Gradient	20 to 50% buffer 2 over 30 minutes

---

The crude material was purified on HPLC reverse phase conditions under the conditions given in Table 24. The yield of homogeneous material was 264 mg (22% overall, 71% from crude)

**Amino acid analysis:** Ala 6.9, Arg 1.1, Asx 6.10, Glx 5.0, Gly 8.2, Ile 3.0, Phe 4.0, Pro 4.0, Thr 0.97, Ser, 0.0, Val 1.5, Lys 5.1, Leu 3.0, Tyr 3.4, Cys 0.0, Trp 0.0.

**Edman degradation:** Sequence analysis confirmed the first thirty-nine residues.

**Plasma Desorption MS:** (M + H)<sup>+</sup> 6610.8 (Calc. 6610.4 from sequence).

**MALD-TOF:** (M + H)<sup>+</sup> 6619 (Calc. 6610.4 from sequence).

**ES- MS:** 6610.4 (Calc. 6610.4 from sequence).

## Trypstatin-74

IAACN LPIVQ GPCRA FAELL AFDAQ QGKCI QFIYG GCKGN NNKFY  
SEPKC KWYCG VPGDG YSGGS GGSGG CSGG.

The chain assembly was carried out on an Applied Biosystems 430A automated peptide synthesizer by the 9-fluorenylmethoxycarbonyl/t-butyl strategy. The machine cycles are shown in Table 25. The starting resin used in reaction column A was 4-hydroxymethyl-phenoxy methyl polystyrene (128mg, 0.12 mmol of 0.94 mmol g<sup>-1</sup>) was used. The first amino acid was coupled to the resin as a symmetrical anhydride formed from DIPCDI with a 1% DMAP as catalyst for one hour. The N-terminal Fmoc group was removed as shown in Table 25. Chain assembly was achieved by reacting the protected amino acid with one equivalent of HBTU/DIEA in NMP, waiting 5 min. for activation at room temperature and adding the mixture to the resin, final concentration 0.3M. Side chain protection was as follows: Trt (Cys, Asn and Gln); t-Bu (Ser, Thr, Tyr); PMC for Arginine; Boc (Lys); and none (Ile, Pro, Ala, Gly, Phe, Val, Trp). Each residue was single coupled for thirty minutes per coupling. After completion of assembly, the N-terminal Fmoc group was removed and the resin dried under argon. The resin beads were red in PSA solution.

Table 25. HBTU/ DIEA /NMP Coupling Cycle for the Solid Phase Synthesis of Trypstatin-74.

Step	Reagents	Time/min.
1	20% piperidine in NMP <sup>a</sup>	3, 15
2	NMP wash	10 x 0.5
3	8.3 equivalents of Fmoc-AA-HOBt ester from HBTU	30
4	NMP wash	10 x 0.5

<sup>a</sup> The final N-terminal Fmoc group is removed by treatments of 40% piperidine in NMP for 7 and twenty minutes.

Phenol (4 ml), p-cresol (4 ml) and dimethylsulphide (4 ml) were added to the resin in a 1.0l round bottomed flask under nitrogen. A mixture of thiophenol (6 ml), TFA (76 ml), ethanedithiol (20 ml) and water (4 ml) was added and the reaction mixture stirred at RT for 2 hours. The resin was filtered and washed with TFA. The filtrate was evaporated *in vacuo* and co-evaporated with ether. The mixture was co-evaporated with ether three times to remove residual acid and finally, chilled ether (500ml) was added and the white precipitate allowed to settle. The ether layer was decanted and the residue evaporated *in vacuo*. The ether washing steps were repeated a further four times. The precipitate was then evaporated to dryness. 50 mM TRIS/ 6M guanidine pH 8.3 (50ml) was added to the white powder in an ice bath and the mixture stirred for thirty minutes. DTT (100mg) was then added and the mixture left for a further thirty minutes. Glacial acetic acid was then added to pH 4 and the mixture partially evaporated *in vacuo*. The residue was taken up in 20% acetic acid, filtered to remove insoluble gel-like material and gel filtered on Sephadex G-25 with 20% acetic acid as eluant. 3.5 ml fractions were collected, water added and the mixture lyophilised to a

white powder. The crude material was purified on HPLC reverse phase conditions under the conditions given in Table 26.

Table 26. HPLC Reverse-Phase Conditions

Column	30 nm Dynamax Cg 2.5 x 25cm
Buffer 1	0.05% TFA in water
Buffer 2	0.05% TFA in acetonitrile
Flow Rate	12 ml/min.
Detection	225 nm UV
Gradient	20 to 50% buffer 2 over 30 minutes

**ES- MS:** 7520. (Calc. 7517 from sequence).

## Refolding Methods

### Air Oxidation

Trypstatin was dissolved in 1% acetic acid at a concentration of 0.1 mg/ml. The pH was adjusted to 8.4 with concentrated ammonia and the cloudy solution stirred for 50 hours vigorously in a one litre flask. The solution was filtered and the filtrate loaded on to a Dynamax C<sub>8</sub> column and eluted with a gradient of 0 to 60% acetonitrile over 40 minutes. The main peak was collected, lyophilised to yield approx. 0.1 mg of reduced material.

### Procedure for human transforming growth factor – alpha

Crude peptide was partially dissolved in 1.0M guanidine/0.25M tris pH2 at a concentration of 0.095 mg/ml. The pH was adjusted to 8.5 and the mixture stirred vigorously in an open flask for 50 hours. The sample was filtered and loaded on to a Dynamax C<sub>8</sub> column and eluted with a gradient of 20 to 50% acetonitrile over 40 minutes. The material returned had mass of reduced material 6610 D

### Procedure for Echistatin

Crude peptide (63mg) was dissolved in 10% acetic acid (20ml) and this solution added to 4L of 0.5% acetic acid with stirring. The pH was adjusted to 8.2 with concentrated ammonia and the mixture left standing in the cold room for 10 days. The material returned had mass of reduced material 6610 D.

### Procedure for Interleukin-3

10 mg of peptide (1.6  $\mu$ M) was dissolved in 200 $\mu$ l of 6M guanidine/ 0.25 mM TRIS and the solution added to 100ml of 2M guanidine pH 8.5 containing 10 $\mu$ M of

oxidized glutathione and 100 $\mu$ M of reduced glutathione. The mixture was left standing overnight at room temperature. HPLC showed no change in the chromatogram of the reduced species. The mixture was diluted to ~0.67M guanidine and left for a further 66 hours at 4°C. HPLC showed no major species.

#### Creighton's procedure for BPTI

7.6  $\mu$ M trypstatin, 100mM Tris, 200mM KCl, 10mM GSH, 1mM GSSG, pH 8.8. A lot of precipitation occurred overnight even when peptide concentrations of 1 $\mu$  M were used. HPLC indicated very little material in the solution.

#### Bowman-Birk Inhibitor

The Bowman-Birk inhibitor contains a trypsin-binding domain. A column of Trypsin-Sepharose was made up as follows: 10g of CN-Br sepharose, 350 mg trypsin in 0.1M sodium bicarbonate pH 8.3 and left for four hours at room temperature. The gel was then blocked with 1M ethanolamine for twenty minutes at room temperature, and washed with 0.1M acetate buffer pH 4 and 0.5M sodium chloride and stored in coupling buffer. 5mg of peptide was then dissolved in 100ml of Creighton's buffer, and the mixture left at 37°C for one hour. Trypsin sepharose was added and the mixture gently swirled for a further 90 minutes. The matrix was transferred to a sintered funnel and eluted with 50 ml fractions of 0.1M HCl and aliquots of the fractions were run on the HPLC. No protein material was eluted from the column.

#### Method according to Cleary

The crude powder (20mg) was suspended in 20 ml of 25mM Tris-HCl/ 6M guanidine hydrochloride, pH 8.0, and shaken for 10 hours. This solution was diluted 5-fold, in three incremental additions, with 25 mM Tris-HCl/ 1.25 mM oxidised and 1.25 mM reduced glutathione, pH 8.0, with 2 hour intervals between additions. The supernatant was then acidified with acetic acid, gel-filtered on Sephadex G-25, and the eluant lyophilised. HPLC showed a broad unresolved peak characteristic of aggregated soluble material.

#### DMSO oxidation [146]

Peptide aliquots of trypstatin at 5mg/ml in 6M guanidine/ 25mM tris pH 8 were adjusted to pH 3 and pH 6 by addition of 1M NaOH solution. DMSO was added to each aliquot to give final percentages between 20 and 60%. The reaction mixtures were monitored by reverse phase HPLC after 1, 4 and 18 hours. In all the experiments, although precipitation was not observed, RP-HPLC profiles showed the presence of aggregated material.

## Synthesis of Fmoc-L-BTD-OH

**S-3-Carbobenzoxy-5-oxo-1, 3-oxazolidione-4-propionic acid.** A solution of Cbz-glutamic acid (25 g, 0.089 mol) paraformaldehyde (5.39 g, 0.178 mol) and *p*-toluene sulphonic acid (1.02 g, 0.0053 mol) was refluxed in toluene (700ml) for one hour with removal of water (1.6 ml) in a Dean-Stark trap. Ethyl acetate (100 ml) was added and the solution washed with 0.3M potassium carbonate (20 ml) and H<sub>2</sub>O (3x20 ml). The solution was dried over MgSO<sub>4</sub> and evaporated *in vacuo* to give 21.0 g of an oil which was carried on without further purification: <sup>1</sup>H NMR CDCl<sub>3</sub>, 300 MHz δ 2.2-2.3 (m, 2H); δ 2.5 (m, 2H); δ 4.4 (m, 1H); δ 5.2 (m, 3H); δ 5.55 (br 1H); δ 7.2-7.4 (m, 5H).

**S-3-Carbobenzoxy-5-oxo-1, 3-oxazolidione-4-propanal.** A solution of S-3-Carbobenzoxy-5-oxo-1, 3-oxazolidione-4-propionic acid (20.0g, 70 mmol) in dry dichloromethane (200ml) was treated with (COCl)<sub>2</sub> (13ml, 2 equiv.) in the presence of a catalytic amount of DMF (0.2 ml, 4 mol%) at room temperature for 18 h. The reaction mixture was evaporated *in vacuo* to a yellow oil, and the crude acid chloride was redissolved in dry EtOAc (600 ml) and stirred in an ice bath under N<sub>2</sub>. Bu<sub>3</sub>SnH (20 g, 1 equiv.) was added by syringe pump over a 4 h period. The solution was then allowed to warm to room temperature and stirred for 18 h. The solution was washed with H<sub>2</sub>O (2x400 ml) and saturated NaCl (200 ml), dried (MgSO<sub>4</sub>) and evaporated *in vacuo*. The oil was dissolved in CH<sub>3</sub>CN (500 ml) and washed with petroleum ether (5x100 ml). Evaporation of the CH<sub>3</sub>CN layer gave 16.2 g of a yellow oil which was carried through without further purification: <sup>1</sup>H NMR CDCl<sub>3</sub>, 300 MHz δ 9.6 (s, 1H, CHO).

**Cbz-L-BTD-OMe.** L-cys-OMe.HCl (8.0 g, 0.8 equiv.) was added in one portion to a stirred solution of the crude aldehyde (16.2 g, 60 mmol) in pyridine (500 ml) at room temperature under N<sub>2</sub>, and the clear solution left to stand for 5 d and then evaporated *in vacuo* to an oil. Chromatography (silica gel, EtOAc/ hexane 1:1 to 6:4 gradient) gave 7.0 g of condensation product, N-hydroxymethyl-Cbz-L-BTD-OMe as a white foamy solid.

A solution of the bicyclic dipeptide (7.0 g, 17.5 mmol) in dry MeOH (200 ml) was stirred with solid Na<sub>2</sub>CO<sub>3</sub> (140 mg, 0.1 equiv.) at room temperature for 18 h. Evaporation *in vacuo* gave 7.0 g of product which was carried through without further purification.

**H-L-BTD-OH.** A solution of the methyl ester (7.0 g) in MeOH (50 ml) was treated with 1 M NaOH (41 ml, 2.2 equiv.) at 4°C for 4 h. The reaction mixture was evaporated *in vacuo* diluted with 40 ml of H<sub>2</sub>O, adjusted to pH 1 with 1 M HCl, and then extracted with EtOAc (3x200 ml). The organic layer was washed with saturated NaCl, dried over MgSO<sub>4</sub> and evaporated *in vacuo* to a white foam.

The crude acid was stirred with 30% HBr/AcOH (20 ml). at room temperature for 18 h.. There was vigorous evolution of gases for the first 30 min. The mixture was co-evaporated with EtOAc and Et<sub>2</sub>O and evaporated *in vacuo* to give 9.0 g of a clear oil.

**Fmoc-L-BTD-OH.** A cloudy solution of the free acid (9.0 g) in 10% Na<sub>2</sub>CO<sub>3</sub> (100 ml, 2.5 equiv.) was added over 15 min. to a vigorously stirred solution of N- (Fmoc-O-succinimide (18.5 g) in DME (180 ml) and the suspension stirred for 18 h. The reaction mixture was partially evaporated *in vacuo* and diluted with H<sub>2</sub>O (300 ml) to give a clear solution which was extracted with Et<sub>2</sub>O (100 ml) and EtOAc (2x200 ml). The aqueous phase was cooled in an ice bath and acidified with concentrated HCl to pH 1.5. The cloudy solution was extracted with EtOAc (3x500 ml), and the organic

layer washed with saturated NaCl (400 ml), dried over MgSO<sub>4</sub> and evaporated to give 8.5 g of a white foamy solid. Chromatography (silica gel, CH<sub>2</sub>Cl<sub>2</sub>/MeOH/AcOH 95:4.9:0.1) gave 5.0 g of Fmoc protected acid.

#### Wild-type Protein G B2 Domain

NH<sub>2</sub>-TTYKL VINGK TLKGE TTTEA VDAAT AEKVF KQTAN DNGVD  
GEWTY DDATK TFTVT EKPE-OH

The chain assembly was carried out on an Applied Biosystems 430A peptide synthesizer using Fastmoc chemistry with an 8-fold excess of amino acid for each 60 min coupling step. The polypeptide was made as a C-terminal acid starting with 0.21 g of Fmoc-Glu (OtBu)-4-hydroxymethyl-phenoxy methyl polystyrene resin(0.55 mmol/g). All amino acids were coupled as Fmoc derivatives. The following protecting groups were used: Boc for Lys and Trp; t-Bu groups for Asp, Glu, Tyr and Thr; Trt for His, Asn and Gln. Chain assembly was achieved with HBTU/HOBt/DIEA (1:1:2) activation in a mixture of NMP/DMF (2:1) at a concentration of 0.3M. Removal of the N-terminal Fmoc group was achieved with four treatments of four minutes each with 20% piperidine in NMP. The first twenty amino acids were coupled with a double coupling cycle of 60 min each without capping. The last 39 were coupled with a single cycle of 60 min and the resin was then capped for 8 min with 20% acetic anhydride in NMP. At the end of the assembly the N-terminal Fmoc group was removed with 40% piperidine in NMP and the resin dried under argon. The PSA positive resin was then treated with 2.5% triisopropylsilane, 2.5% H<sub>2</sub>O in TFA (200 ml) for 2 h at RT. Filtration, evaporation *in vacuo*, and precipitation with ether (500 ml) gave 0.8 g of white powder which was dissolved in 6M guanidine, 0.1M TRIS, 2 mM EDTA pH 8.7, left to stand for 30 min at RT, and desalted by loading onto a C<sub>8</sub> RP-HPLC Hyperprep column (300 Å, 250x

21.5 mm) and running a linear gradient of 20-40% 0.05% TFA in CH<sub>3</sub>CN at a flow rate of 10 ml/min over 80 min. Lyophilisation of the broad peak gave 90mg of white powder. 25 mg fractions of the above solution were loaded onto a Synchronpak AX-100 anion exchange column (250x 21.5 mm) and the products eluted with a linear gradient of 0-40% 20 mM TRIS pH 7.4/1M sodium chloride in 20 mM TRIS pH 7.4 at a flow rate of 8 ml/min over 40 min. Elution was monitored at 225 nm. The four major fractions were desalted by loading onto a C<sub>8</sub> RP-HPLC Hyperprep column (300 Å, 250x 21.5 mm) and eluted with a linear gradient of 0-50% 0.05% TFA in CH<sub>3</sub>CN at a flow rate of 10 ml/min over 30 min. The partially purified material was then refolded by heating a solution of the protein in 50 mM TRIS pH 7.6 (50ml) for 15 minutes at 75°C. The mixture was then cooled rapidly in an ice bath and filtered. The filtrate was kept at 4°C throughout the purification process. The pH was adjusted to 3 with acetic acid and tris base added until the pH was 7. This material was loaded onto a Poros HQ/M anion exchange column. The products were eluted with a linear gradient of 0-30% 20 mM TRIS pH 7.5/0.5M sodium chloride in 20 mM TRIS pH 7.5 at a flow rate of 10 ml/min over 10 min. Elution was monitored at 225 nm. The eluted peak was gel-filtered into water using PD-10 columns. Lyophilisation gave 29 mg of wild type Protein G B2 domain in 4% yield: ES-MS 6444.5 (Calc. 6443 from sequence); MALD-TOF: 6443.3. Amino acid Analysis: Ala 10 AsX 14.5, Glx 12.2, Ile 1, Phe 3.4, Lys 11.4, Thr 24.3, Pro 1.9, Gly 6.9, Val 7.4, Leu 3.4, Tyr 3.5, Trp N.D.

## K15, T16 BTB B2 Domain

NH<sub>2</sub>-TTYKL VING BTB LKGE TTTEA VDAAT AEKVF KQTAN DNGVD  
GEWTY DDATK TFTVT EKPE-OH

The chain assembly and removal of protecting groups were carried out in exactly the same manner as for the wildtype synthesis. Refolding and purification were carried out as follows. The powder from ether precipitation was dissolved in 6M guanidine, 0.1M TRIS, 2 mM EDTA pH 8.7 (20 ml) and left to stand for 20 min at RT. This material was desalted by loading onto a C<sub>8</sub> RP-HPLC Hyperprep column (300 Å, 250x 21.5 mm) and running a linear gradient of 20-40% 0.05% TFA in CH<sub>3</sub>CN at a flow rate of 10 ml/min over 80 min. Lyophilisation of the broad peak gave 90mg of white powder. The partially purified material was then refolded by heating a solution of the protein in 50 mM TRIS pH 7.6 (50ml) for 15 minutes at 75°C. The mixture was then cooled rapidly in an ice bath and filtered. The filtrate was kept at 4°C throughout the purification process. The pH was adjusted to 3 with acetic acid and tris base added until the pH was 7. This material was loaded onto a Poros HQ/M anion exchange column. The products were eluted with a linear gradient of 0-30% 20 mM TRIS pH 7.5/0.5M sodium chloride in 20 mM TRIS pH 7.5 at a flow rate of 10 ml/min over 10 min. Elution was monitored at 225 nm. The eluted peak was gel-filtered into water using PD-10 columns Lyophilisation gave 27 mg of analogue Protein G B2 Domain in 4% yield: ES-MS 6413 (Calc. 6412 from sequence).

G14, K15 BTB B2 Domain.

NH<sub>2</sub>-LTPAV TTYKL VIN BTB TLKGE TTTEA VDAAT AEKVF KQTAN  
DNGVD GEWTY DDATK TFTVT EKPE-OH

The chain assembly and removal of protecting groups were carried out in exactly the same manner as for the wildtype synthesis. Refolding and purification were carried out as for K15T16BTB. Lyophilisation gave 6 mg of analogue Protein G B2 domain in 1% yield: ES-MS 6938 (Calc. 6934 from sequence).

Biotin 0, G14, K15 BTB B2 Domain

Biotin-LTPAV TTYKL VIN BTB TLKGE TTTEA VDAAT AEKVF KQTAN  
DNGVD GEWTY DDATK TFTVT EKPE-OH

The chain assembly and removal of protecting groups were carried out in exactly the same manner as for the wildtype synthesis. Filtration, evaporation *in vacuo*, and precipitation with ether (500 ml) gave a white powder which was dissolved in 6M guanidine, 0.1M TRIS, 2 mM EDTA pH 8.7 (20 ml) and left to stand for 20 min at RT. The solution was added by syringe pump over a 3h period to 20 mM TRIS pH 7.4 (1 L). The mixture was allowed to stand at 4°C for 3d and loaded onto a Synchronpak AX-100 anion exchange column (250x 21.5 mm) and the products eluted with a linear gradient of 0-30% 20 mM TRIS pH 7.5/1M sodium chloride in 20 mM TRIS pH 7.5 at a flow rate of 8 ml/min over 90 min. Elution was monitored at 225 nm. The major fraction was desalted by loading onto a C<sub>8</sub> RP-HPLC Hyperprep column (300 Å, 250x 21.5 mm) and eluted with a linear gradient of 10-50% 0.05% TFA in CH<sub>3</sub>CN at a flow rate of 10 ml/min over 40 min. Lyophilisation gave 6 mg of analogue Protein G Domain 2 in 1% yield: ES-MS 7164 (Calc. 7160.7 from sequence).

## Q37Aib B2 Domain

NH<sub>2</sub>-TTYKL VINGK\_TLKGE TTTEA VDAAT AEKVF KAibTAN DNGVD  
GEWTY DDATK TFTVT EKPE-OH

The chain assembly and removal of protecting groups were carried out in the same manner as for the wildtype synthesis with the following exceptions. A single coupling of 40 minutes was used. Activation of amino acids was achieved with HATU/HOAt/DIEA (1:1:2) in NMP at a concentration of 0.3M. Removal of the N-terminal Fmoc group was achieved with four treatments of four minutes each with 20% piperidine in NMP. At the end of the assembly the N-terminal Fmoc group was removed with 40% piperidine in NMP and the resin dried under argon. The PSA positive resin was then treated with 2.5% triisopropylsilane, 2.5% H<sub>2</sub>O in TFA (100 ml) for 90 min at RT. Filtration, evaporation *in vacuo*, and precipitation with ether (500 ml) gave 700 mg of white powder which was dissolved in 6M guanidine, 0.1M TRIS, 2 mM EDTA pH 8.7 (20 ml) left to stand for 20 min at RT, and dialysed (by Dispodialyser) into 20 mM TRIS pH 7.7. This material was desalted by loading onto a C<sub>8</sub> RP-HPLC Hyperprep column (300 Å, 250x 21.5 mm) and eluted with a linear gradient of 10-80% 0.05% TFA in CH<sub>3</sub>CN at a flow rate of 10 ml/min over 40 min. Lyophilisation gave 269 mg of white powder. This material was loaded onto a MonoQ HR 5/5 anion exchange column and the products eluted with a linear gradient of 0-40% 20 mM TRIS pH 7.5/1M sodium chloride in 20 mM TRIS pH 7.5 at a flow rate of 1 ml/min over 40 min. Elution was monitored at 225 nm. Three major fractions were desalted by loading onto a C<sub>8</sub> RP-HPLC Hyperprep column (300 Å, 250x 21.5 mm). The partially purified material was then refolded by heating a solution of the protein in 50 mM TRIS pH 7.6 (50ml) for 15 minutes at 75°C. The mixture was then cooled rapidly in an ice bath and filtered. The filtrate was kept at 4°C throughout the purification process. The pH was adjusted to 3 with acetic acid and tris base added until the pH was 7. This material was loaded onto a Poros HQ/M

anion exchange column. The products were eluted with a linear gradient of 0-30% 20 mM TRIS pH 7.5/0.5M sodium chloride in 20 mM TRIS pH 7.5 at a flow rate of 10 ml/min over 10 min. Elution was monitored at 225 nm. The eluted peak was gel-filtered into water using PD-10 columns. Lyophilisation gave 30 mg of material. ES-MS 6400 (Calc. 6401 from sequence).

#### K15T16Ahx B2Domain

NH<sub>2</sub>-TTYKL VINGAhx LKGE TTTEA VDAAT AEKVF KQTAN DNGVD  
GEWTY DDATK TFTVT EKPE-OH

The chain assembly and removal of protecting groups were carried out in the same manner as for the wildtype synthesis with the following exception.. A single coupling step of 60 minutes was used. Refolding and purification were carried out as for K15T16BTD with the following exception. The mixture was applied to a Resource Q ion-exchange column (1ml) in 5ml aliquots. The domain was eluted with a linear gradient of 0-30% 20 mM TRIS pH 7.5/0.5M sodium chloride in 20 mM TRIS pH 7.5 at a flow rate of 4 ml/min over 10 min. Elution was monitored at 225 nm. The eluted peak was desalted by gel-filtration in to water. Lyophilisation gave 35mg. ES-MS 6328 (Calc. 6326 from sequence).

#### G14K15Ahx B2 Domain

NH<sub>2</sub>-TTYKL VINAhx TLKGE TTTEA VDAAT AEKVF KQTAN DNGVD  
GEWTY DDATK TFTVT EKPE-OH

The chain assembly and removal of protecting groups were carried out in the same manner as for the wildtype synthesis with the following exception. A single coupling step of 60 minutes was used. Refolding and purification were carried out as for

K15T16BTD with the following exception. After refolding, the mixture was applied to a Resource Q ion-exchange column (1ml) in 5ml aliquots. The domain was eluted with a linear gradient of 0-30% 20 mM TRIS pH 7.5/0.5M sodium chloride in 20 mM TRIS pH 7.5 at a flow rate of 4 ml/min over 10 min. Elution was monitored at 225 nm. The eluted peak was desalted by gel-filtration into water. Lyophilisation gave 33mg. ES-MS 6371 (Calc. 6372 from sequence).

#### K15dPT16N-MeA B2 Domain

NH<sub>2</sub>-TTYKL VINGdP N-MeALKGE TTTEA VDAAT AEKVF KQTAN DNGVD  
GEWTY DDATK TFTVT EKPE-OH

The chain assembly and removal of protecting groups was carried out in the same manner as for the wildtype synthesis with the following exceptions. A single coupling of 40 minutes was used. Activation of amino acids was achieved with HATU/HOAt/DIEA (1:1:2) in NMP at a concentration of 0.3M. Refolding and purification were carried out as for K15T16BTD. Lyophilisation gave 20 mg of material. ES-MS 6396 (Calc. 6397 from sequence).

#### G14K15BTD B2 Domain

NH<sub>2</sub>-TTYKL VINBTD TLKGE TTTEA VDAAT AEKVF KQTAN DNGVD  
GEWTY DDATK TFTVT EKPE-OH

The chain assembly and removal of protecting groups were carried out in the same manner as for the wildtype synthesis with the following exceptions. A single coupling of 40 minutes was used. Activation of amino acids was achieved with HATU/HOAt/DIEA (1:1:2) in NMP at a concentration of 0.3M. Refolding and

purification were carried out as for K15T16BTD. Lyophilisation gave 21 mg of material. ES-MS 6456 (Calc. 6457 from sequence).

#### 15BTD16 B2 Domain

NH<sub>2</sub>-TTYKL VINGK BTD TLKGE TTTEA VDAAT AEKVF KQTAN DNGVD  
GEWTY DDA TK TFTVT EKPE-OH

The chain assembly and removal of protecting groups were carried out in the same manner as for the wildtype synthesis with the following exceptions. A single coupling of 40 minutes was used. Activation of amino acids was achieved with HATU/HOAt/DIEA (1:1:2) in NMP at a concentration of 0.3M. Refolding and purification were carried out as for K15T16BTD. Lyophilisation gave 14 mg of material. ES-MS 6642 (Calc. 6641 from sequence).

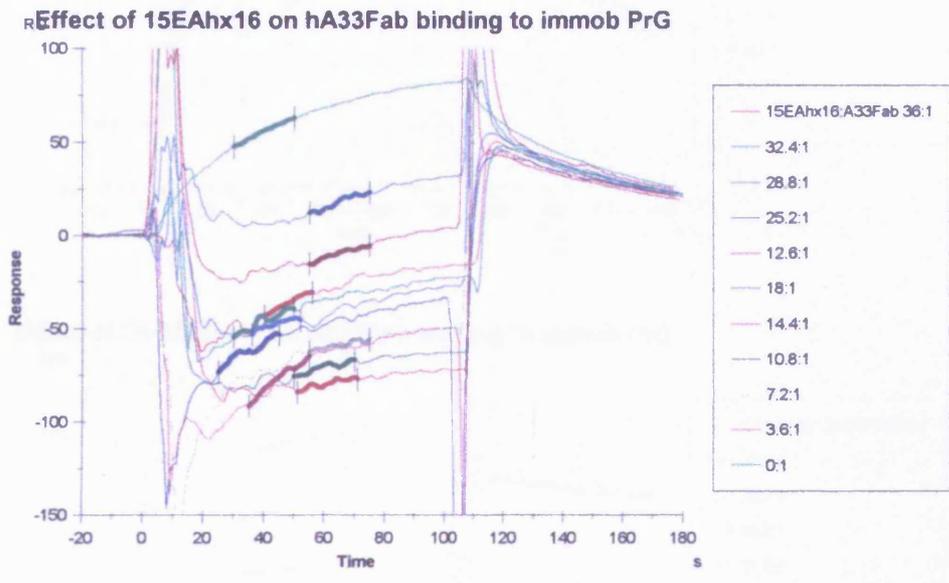
#### 15Ahx16 B2 Domain

NH<sub>2</sub>-TTYKL VINGK Ahx TLKGE TTTEA VDAAT AEKVF KQTAN DNGVD  
GEWTY DDA TK TFTVT EKPE-OH

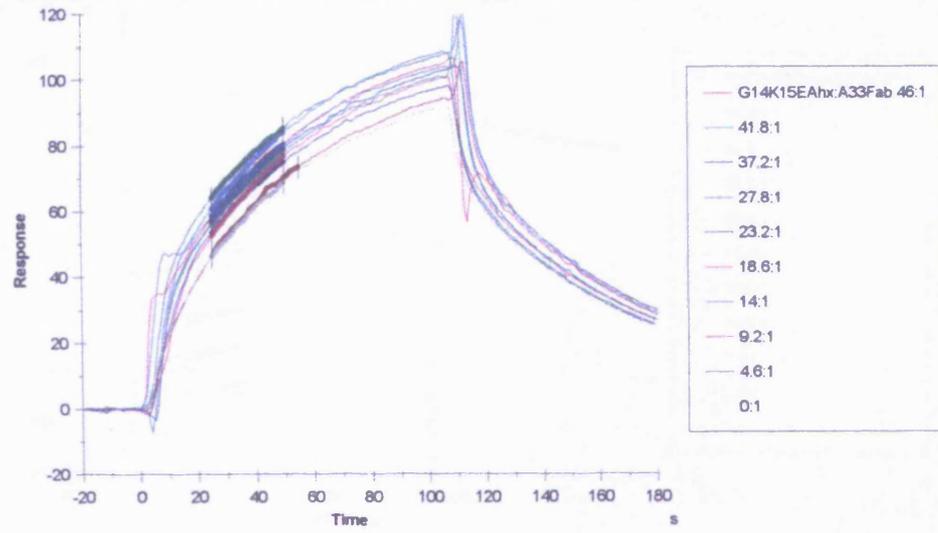
The chain assembly and removal of protecting groups were carried out in the same manner as for the wildtype synthesis with the following exceptions. A single coupling of 40 minutes was used. Activation of amino acids was achieved with HATU/HOAt/DIEA (1:1:2) in NMP at a concentration of 0.3M. Refolding and purification were carried out as for K15T16BTD. Lyophilisation gave 7mg of material. ES-MS 6557 (Calc. 6556 from sequence).

# Appendices

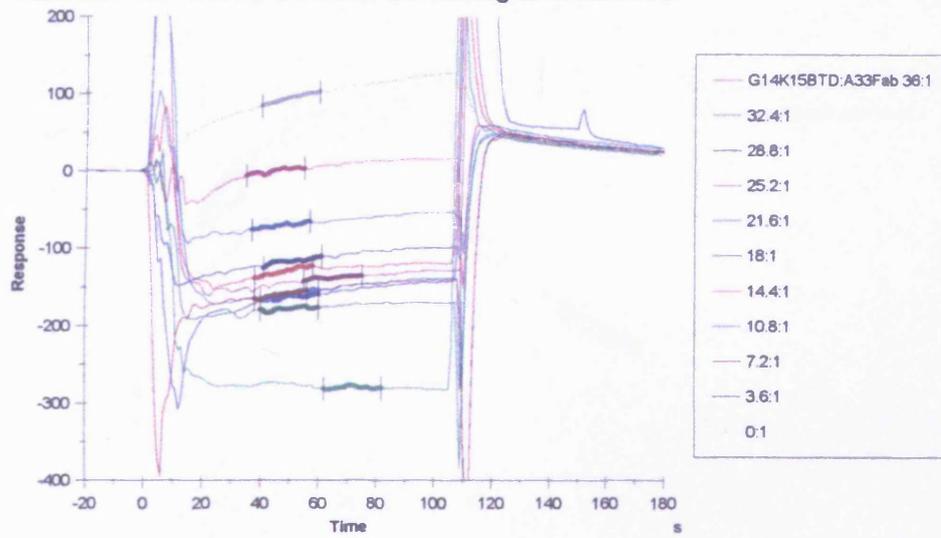
## Appendix 1



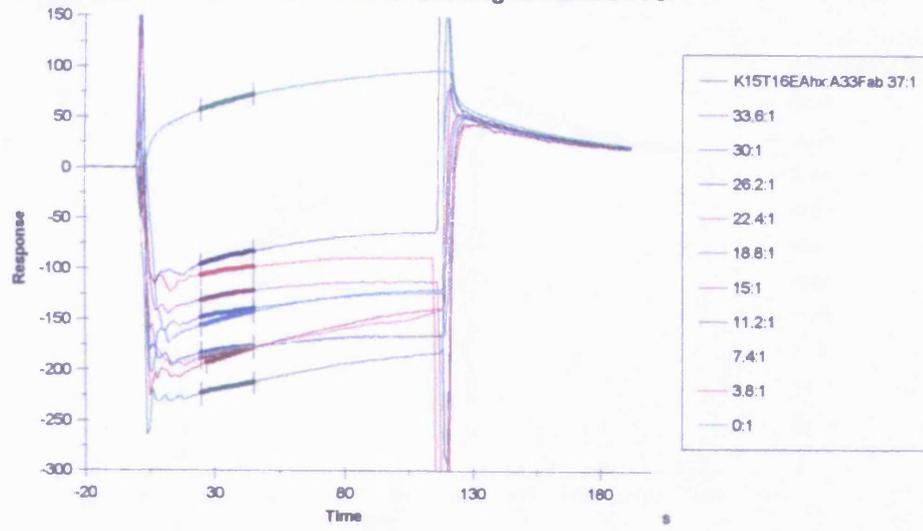
**Effect of G14K15EAhx on hA33Fab binding to immob PrG**



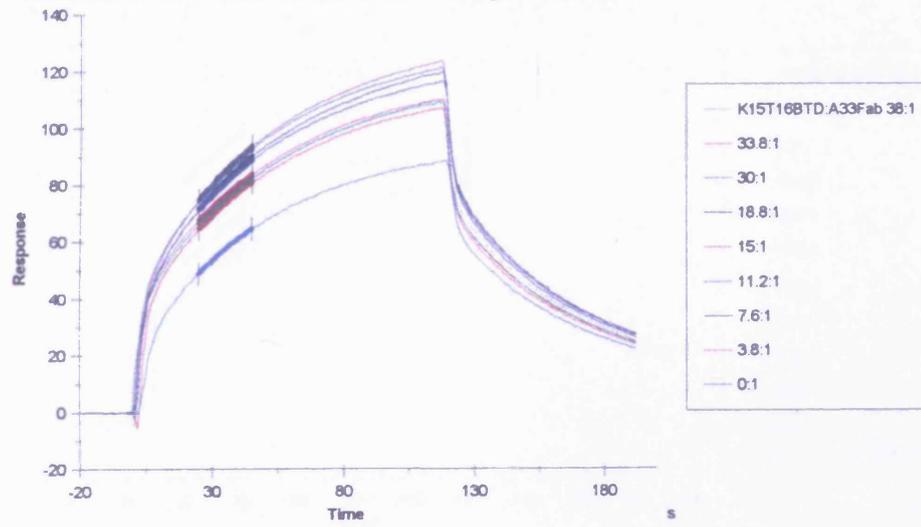
**Effect of G14K15BTD on hA33Fab binding to immob PrG**



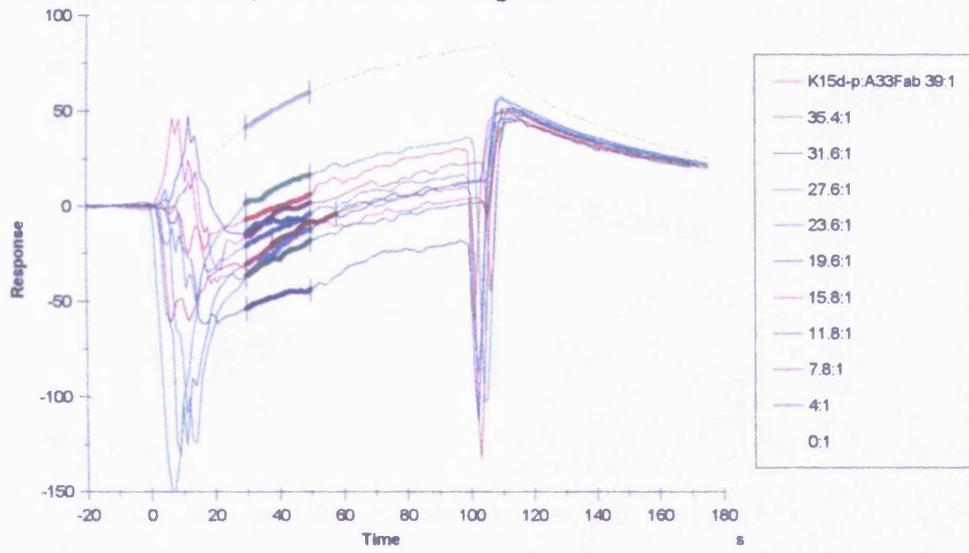
### Effect of K15T16EAhx on hA33Fab binding to immob PrG



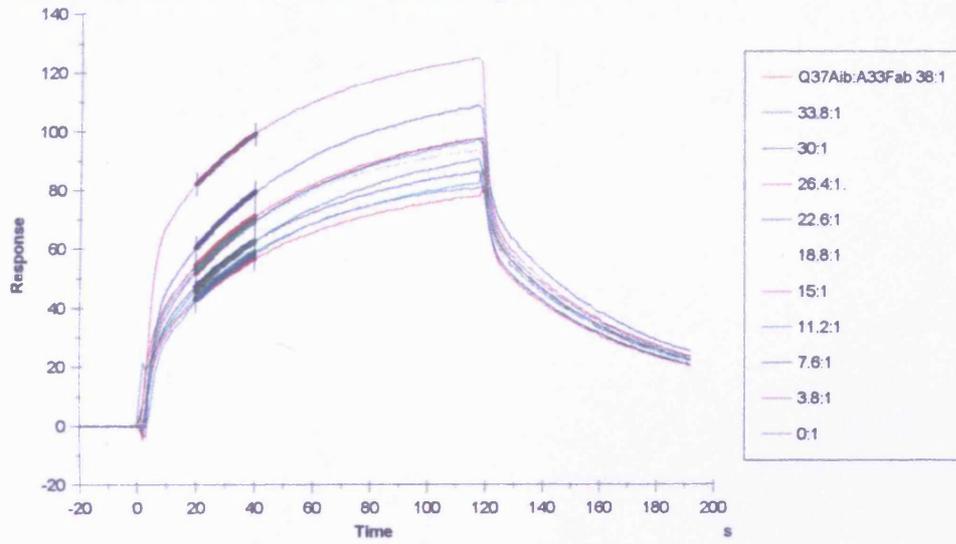
### Effect of K15T16BTD on hA33Fab binding to immob PrG



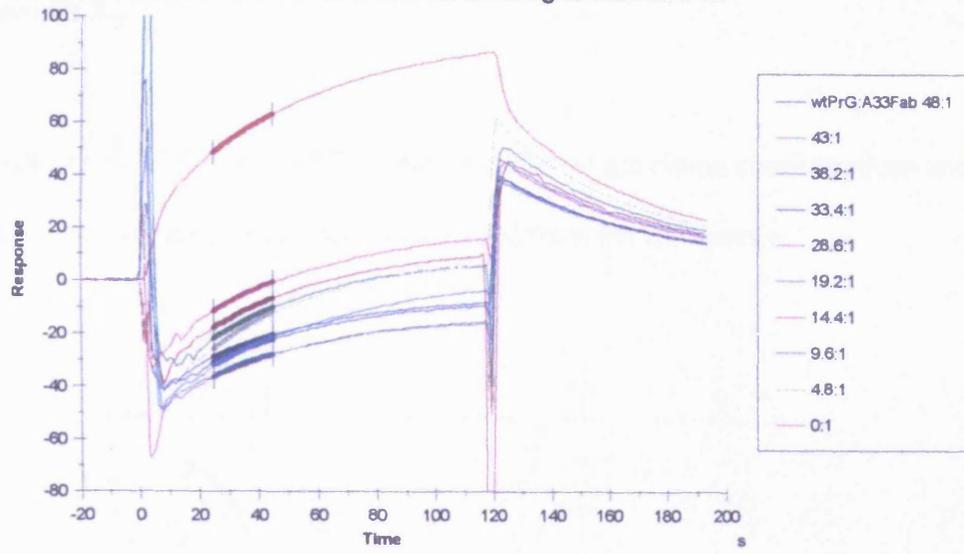
RU Effect of K15d-p on hA33fab binding to immob PrG



RU Effect of Q37Aib on hA33Fab binding to immob PrG

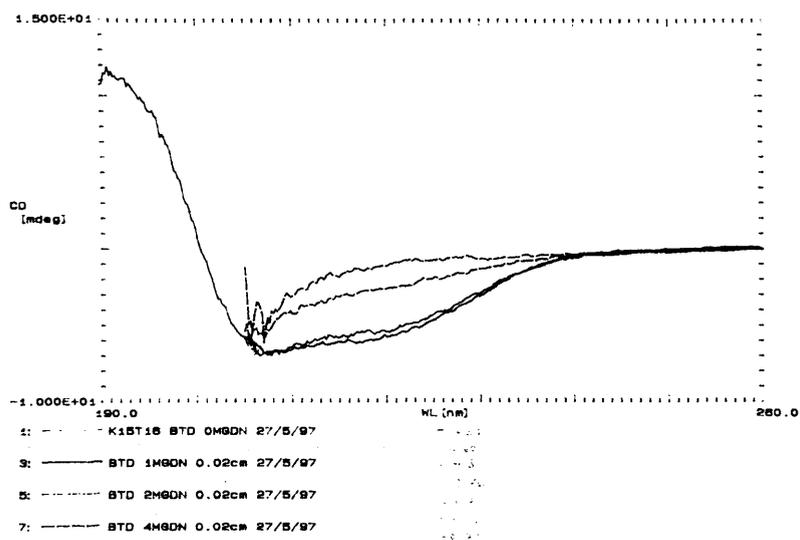


### Effect of soluble wtPrG on hA33Fab binding to immob PrG



## Appendix 2.

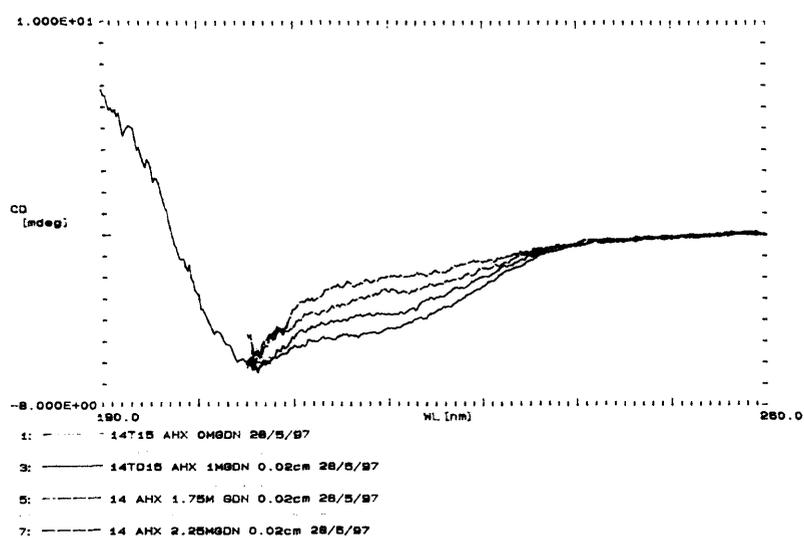
CD spectra of K15T16BTD B2 domain at different guanidine concentrations and the data for the ellipticity at 225 nm as measured from the CD spectra.



⑤

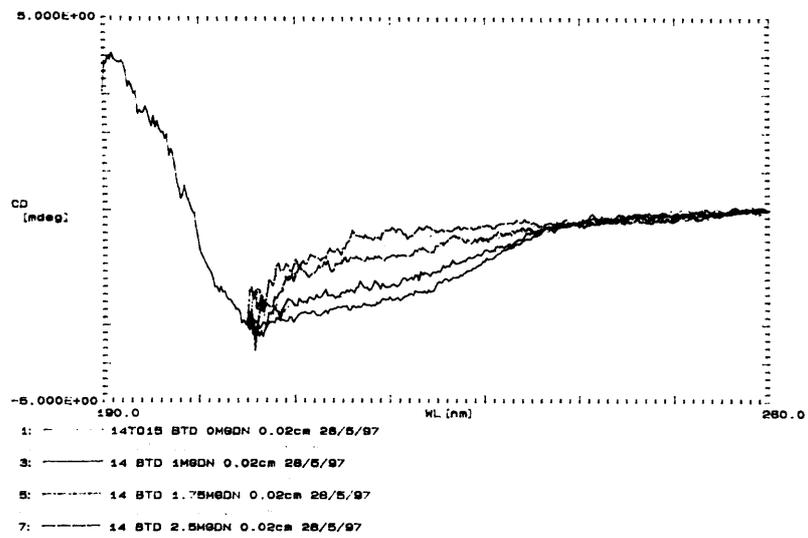
[GdnHCl]/M	Ellipticity at 225nm/mdeg	% change
0	-4.63	0
0.5	-4.60	0.8
1.0	-4.30	8.4
1.5	-3.76	22.2
1.6	-3.06	40.1
1.8	-2.64	50.8
2.0	-1.89	69.9
2.5	-1.19	87.8
4	-0.71	100

CD spectra of G14K15Ahx B2 domain at different guanidine concentrations and the data for the ellipticity at 225 nm as measured from the CD spectra.



[GdnHCl]/M	Ellipticity at 225nm/mdeg	% change
0	-3.58	0
0.5	-3.33	8.7
1.0	-2.92	23.0
1.5	-2.75	28.9
1.75	-2.35	42.9
2.0	-2.17	49.3
2.25	-1.68	66.2
2.5	-1.39	76.3
3.0	-1.13	85.4
4	-0.71	100

CD spectra of G14K15BTD B2 domain at different guanidine concentrations and the data for the ellipticity at 225 nm as measured from the CD spectra.



(α)

[GdnHCl]/M	Ellipticity at 225nm/mdeg	% change
0	-1.89	0
0.5	-1.74	8.7
1.0	-1.59	23.0
1.5	-1.15	28.9
1.75	-0.93	42.9
2.0	-0.70	49.3
2.5	-0.58	76.3
4	-0.59	100

## References

1. Zale, S.E., Klibanov A.M., *Biochemistry*, 1986, **25**, 5432-5444
2. Haynie, D.T. and E. Freire, *Structural energetics of the molten globule state. [Review] [114 refs]*. *Proteins*, 1993. **16**(2): p. 115-40.
3. Shortle, D., *The denatured state (the other half of the folding equation) and its role in protein stability. [Review] [39 refs]*. *Faseb Journal*, 1996. **10**(1): p. 27-34.
4. Fersht, A.R., *The sixth Datta Lecture. Protein folding and stability: the pathway of folding of barnase. [Review]*. *Febs Letters*, 1993. **325**(1-2): p. 5-16.
5. Marqusee, S. and R.T. Sauer, *Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in lambda repressor*. *Protein Science*, 1994. **3**(12): p. 2217-25.
6. Green, S.M. and D. Shortle, *Patterns of nonadditivity between pairs of stability mutations in staphylococcal nuclease*. *Biochemistry*, 1993. **32**(38): p. 10131-9.
7. Baase, W.A., et al., *Dissection of protein structure and folding by directed mutagenesis*. *Faraday Discuss.*, 1992. **93**(Structure and Activity of Enzymes): p. 173-81.
8. Serrano, L., et al., *Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles*. *Biochemistry*, 1990. **29**(40): p. 9343-52.
9. Green, S.M., A.K. Meeker, and D. Shortle, *Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state*. *Biochemistry*, 1992. **31**(25): p. 5717-28.
10. Lim, W.A., D.C. Farruggio, and R.T. Sauer, *Structural and energetic consequences of disruptive mutations in a protein core*. *Biochemistry*, 1992. **31**(17): p. 4324-33.
11. Eriksson, A.E., W.A. Baase, and B.W. Matthews, *Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences*. *J. Mol. Biol.*, 1993. **229**(3): p. 747-69.
12. Brunet, A.P., et al., *The role of turns in the structure of an alpha-helical protein*. *Nature*, 1993. **364**(6435): p. 355-8.
13. Kauzmann, *PNAS*, 1996. **93**: p. 4448.
14. M, R.F., *Journal of Molecular Biology*, 1974. **82**: p. 1.
15. Berndt, K.D., et al., *Designed replacement of an internal hydration water molecule in BPTI: structural and functional implications of a glycine-to-serine mutation*. *Biochemistry*, 1993. **32**(17): p. 4564-70.
16. Walter S, Hubner B., Hahn U., Schmid F. X., *Destabilization of a protein helix by electrostatic interactions*. *Journal of Molecular Biology*, 1995. **252**:133- 143.
17. Sali D., Bycroft M, and A.R. Fersht, *Stabilization of protein Structure by interaction of alpha-helix dipole with a charged side-chain*. *Nature*, 1988. **335** 740-743.

18. Clarke, J., K. Henrick, and A.R. Fersht, *Disulfide mutants of barnase. I: Changes in stability and structure assessed by biophysical methods and X-ray crystallography*. Journal of Molecular Biology, 1995. **253**(3): p. 493-504.
19. Jaenicke, R. *Protein folding: local structures, domains and assemblies*. in *Appl. Enzyme Biotechnol., [Proc. Tex. A&M Univ., IUCCP Symp.]*, 9th. 1991. Inst. Biophys. Phys. Biochem., Univ. Regensburg, Regensburg, D-8400, Germany: Plenum, New York, N.
20. Shortle, D., H.S. Chan, and K.A. Dill, *Modeling the effects of mutations on the denatured states of proteins*. Protein Science, 1992. **1**(2): p. 201-15.
21. Dill, K.A., et al., *Principles of protein folding--a perspective from simple exact models. [Review] [335 refs]*. Protein Science, 1995. **4**(4): p. 561-602.
22. Bryan, P.N., *Site-directed mutagenesis to study protein folding and stability*. Methods Mol. Biol. (Totowa, N. J.), 1995. **40**,: p. 271-89.
23. Anthony-Cahill, S.J., et al., *Site-specific mutagenesis with unnatural amino acids*. Trends-Biochem-Sci, 1989. **14**(10): p. 400-403.
24. Noren, C.J., et al., *A general method for site-specific incorporation of unnatural amino acids into proteins*. Science, 1989 Apr. **244**(4901): p. 182-188.
25. Ellman, J., et al., *Biosynthetic method for introducing unnatural amino acids site-specifically into proteins*. Methods-Enzymol, 1991. **202**: p. 301-336.
26. Ellman, J.A., D. Mendel, and P.G. Schultz, *Site-specific incorporation of novel backbone structures into proteins*. Science, 1992. **255**(5041): p. 197-200.
27. Dawson, P.E., et al., *Synthesis of proteins by native chemical ligation*. Science, 1994. **266**(5186): p. 776-9.
28. Schnolzer, M. and S.B. Kent, *Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease*. Science, 1992. **256**(5054): p. 221-5.
29. Chang, T.K., et al., *Subtiligase: a tool for semisynthesis of proteins*. Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(26): p. 12544-8.
30. Jackson, D.Y., et al., *A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues*. Science, 1994. **266**(5183): p. 243-7.
31. Schnolzer, M., et al., *In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences*. International Journal of Peptide & Protein Research, 1992. **40**(3-4): p. 180-93.
32. Kent, S.B., *Chemical synthesis of peptides and proteins. [Review]*. Annual Review of Biochemistry, 1988. **57**: p. 957-89.
33. Carpino, L.A. and G.Y. Han, *9-Fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group*. J. Amer. Chem. Soc, 1970. **92**(19): p. 5748-9.
34. Carpino, L.A. and G.Y. Han, *9-Fluorenylmethoxycarbonyl amino-protecting group*. J. Org. Chem, 1972. **37**(22): p. 3404-9.
35. Choi, H. and J.V. Aldrich, *Comparison of methods for the Fmoc solid-phase synthesis and cleavage of a peptide containing both tryptophan and arginine*. International Journal of Peptide & Protein Research, 1993. **42**(1): p. 58-63.

36. Ellman, J.A., et al., *Site-specific isotopic labeling of proteins for NMR studies*. J-Am-Chem-Soc, 1992 Sep. **114**(20): p. 7959-7961.
37. Karle I. L., B., B, *Structural Characteristics of helical peptide molecules containing Aib residues*. Biochemistry, 1990. **29**(29): p. 6747-6756.
38. Milton, R.C., S.C. Milton, and S.B. Kent, *Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity [corrected] [published erratum appears in Science 1992 Jul 10;257(5067):147] [see comments]*. Science, 1992. **256**(5062): p. 1445-8.
39. Chalmers D. K., Marshall.G.R., *Pro-D-NMe-Amino acid and D-Pro-NMe-Amino acid: Simple, Efficient Reverse-Turn Constraints*. JACS, 1995. **117**: p. 5927-5937.
40. Saragovi, H.U., et al., *Design and synthesis of a mimetic from an antibody complementarity-determining region*. Science, 1991 Aug. **253**(5021): p. 792-795.
41. Baca, M., P.F. Alewood, and S.B. Kent, *Structural engineering of the HIV-1 protease molecule with a beta-turn mimic of fixed geometry*. Protein Science, 1993. **2**(7): p. 1085-91.
42. Carter, P. and J.A. Wells, *Dissecting the catalytic triad of a serine protease*. Nature, 1988. **332**(6164): p. 564-8.
43. Soriano-Garcia, M., et al., *Structure of calcium prothrombin fragment 1 including the conformation of the Gla domain*. Biochemistry, 1989. **28**(17): p. 6805-10.
44. Stubbs, M.T., et al., *The interaction of thrombin with fibrinogen. A structural basis for its specificity*. European Journal of Biochemistry, 1992. **206**(1): p. 187-95.
45. Folkers, P.J., et al., *Solution structure of recombinant hirudin and the Lys-47---Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study*. Biochemistry, 1989. **28**(6): p. 2601-17.
46. Otto, A. and R. Seckler, *Characterization, stability and refolding of recombinant hirudin*. Eur-J-Biochem, 1991 Nov. **202**(1): p. 67-73.
47. Bode, W. and R. Huber, *Proteinase-protein inhibitor interaction*. Biomed-Biochim-Acta, 1991. **50**(4-6): p. 437-446.
48. Mittman, C., ed. *Pulmonary Emphysema and Proteolysis*. 1972, Academic Press.
49. Taylor, J.C.a.M., C, ed. *Pulmonary Emphysema and Proteolysis* 1986. 1987, Academic Press.
50. *Proteinase inhibitors*, ed. A.J. Barrett, and Salveson, G. 1986: Elsevier.
51. Baici, A., Salgam, P., Cohen, G., Fehr, K., and Boni, A., *Rheumatology International*, 1982. **2**: p. 11-16.
52. Burkhardt, H., Kasten, M., Rauls, S., and Rehkopf, E., *Rheumatology International*, 1987. **7**: p. 133-138.
53. Mizuguchi K., D.C.M., Blundell T. L., Overington J. P., *HOMSTRAD: a database of protein structure alignments for homologous families*. Protein Science, 1998. **7**: p. 2469-2471.
54. Liepinsh, E., et al., *Solution structure and dynamics of PEC-60, a protein of the Kazal type inhibitor family, determined by nuclear magnetic resonance spectroscopy*. J. Mol. Biol., 1994. **239**(1): p. 137-53.

55. Bode, W. and R. Huber, *Natural protein proteinase inhibitors and their interaction with proteinases. [Review]*. European Journal of Biochemistry, 1992. **204**(2): p. 433-51.
56. Salier, J.P., *Inter-alpha-trypsin inhibitor: emergence of a family within the Kunitz-type protease inhibitor superfamily*. Trends-Biochem-Sci, 1990. **15**(11): p. 435-439.
57. Weissman, J.S. and P.S. Kim, *Reexamination of the folding of BPTI: predominance of native intermediates [see comments]*. Science, 1991. **253**(5026): p. 1386-93.
58. Kido, H., Y. Yokogoshi, and N. Katunuma, *Kunitz-type protease inhibitor found in rat mast cells. Purification, properties, and amino acid sequence*. Journal of Biological Chemistry, 1988. **263**(34): p. 18104-7.
59. Berndt, K.D., et al., *Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures*. Journal of Molecular Biology, 1992. **227**(3): p. 757-75.
60. Darby, N.J., et al., *Kinetic roles and conformational properties of the non-native two-disulphide intermediates in the refolding of bovine pancreatic trypsin inhibitor*. Journal of Molecular Biology, 1992. **224**(4): p. 905-11.
61. Darby, N.J. and T.E. Creighton, *Dissecting the disulphide-coupled folding pathway of bovine pancreatic trypsin inhibitor. Forming the first disulphide bonds in analogues of the reduced protein*. Journal of Molecular Biology, 1993. **232**(3): p. 873-96.
62. van Mierlo, C.P., et al., *Partially folded conformation of the (30-51) intermediate in the disulphide folding pathway of bovine pancreatic trypsin inhibitor. 1H and 15N resonance assignments and determination of backbone dynamics from 15N relaxation measurements*. Journal of Molecular Biology, 1993. **229**(4): p. 1125-46.
63. van Mierlo, C.P., et al., *1H NMR analysis of the partly-folded non-native two-disulphide intermediates (30-51,5-14) and (30-51,5-38) in the folding pathway of bovine pancreatic trypsin inhibitor*. Journal of Molecular Biology, 1994. **235**(3): p. 1044-61.
64. Ferrer, M., C. Woodward, and G. Barany, *Solid-phase synthesis of bovine pancreatic trypsin inhibitor and analogs designed to determine the role of disulfide bridges in protein folding and stability*. Pept, 1992(Publisher).
65. Groeger, C., H.R. Wenzel, and H. Tschesche, *Chemical semisynthesis of aprotinin homologues and derivatives mutated in P' positions*. J-Protein-Chem, 1991. **10**(5): p. 527-533.
66. Wagner, G., et al., *Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN*. Journal of Molecular Biology, 1987. **196**(3): p. 611-39.
67. Creighton, T.E., et al., *On the biosynthesis of bovine pancreatic trypsin inhibitor (BPTI). Structure, processing, folding and disulphide bond formation of the precursor in vitro and in microsomes*. Journal of Molecular Biology, 1993. **232**(4): p. 1176-96.

68. States, D.J., et al., *Conformations of intermediates in the folding of the pancreatic trypsin inhibitor*. Journal of Molecular Biology, 1987. **195**(3): p. 731-9.
69. Blattner, W.A., *HIV epidemiology: past, present, and future*. FASEB-J, 1991. **5**(10): p. 2340-2348.
70. Rossi, P. and V. Moschese, *Mother-to-child transmission of human immunodeficiency virus*. FASEB-J, 1991. **5**(10): p. 2419-2426.
71. Jameson, B.A., et al., *Location and chemical synthesis of a binding site for HIV-1 on the CD4 protein*. Science, 1988. **240**(4857): p. 1335-9.
72. Becerra, S.P., et al., *Purification and characterization of the RNase H domain of HIV-1 reverse transcriptase expressed in recombinant Escherichia coli*. Febs Letters, 1990. **270**(1-2): p. 76-80.
73. Wlodawer, A., et al., *Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease*. Science, 1989 Aug. **245**(4918): p. 616-621.
74. Hui, K.Y., et al., *A rational approach in the search for potent inhibitors against HIV proteinase*. FASEB-J, 1991. **5**(11): p. 2606-2610.
75. Miller, M., et al., *Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution*. Science, 1989 Dec. **246**(4934): p. 1149-1152.
76. Stephens, P., Clements, G., Yarranton, G., and Moore, J., *A chink in HIV's armour*. Nature, 1990. **343**: p. 219.
77. Javaherian, K., Langlois A., McDanal C., Ross K., Eckler L., Jellis C., Profy A., Rusche J., Bolognesi D., Putney S., Matthews T. ., *Principal neutralising domain of the human immunodeficiency virus type 1 envelope protein*. Proc. Natl. Acad. Sci. (USA), 1989. **86**: p. 6768-6772.
78. Nara, P.L., R.R. Garrity, and J. Goudsmit, *Neutralization of HIV-1: a paradox of humoral proportions*. FASEB-J, 1991. **5**(10): p. 2437-2455.
79. Clements, G.J., et al., *The V3 loops of the HIV-1 and HIV-2 surface glycoproteins contain proteolytic cleavage sites: a possible function in viral fusion?* AIDS-Res-Hum-Retroviruses, 1991. **7**(1): p. 3-16.
80. Catasti, P., et al., *Local and global structural properties of the HIV-MN V3 loop*. Journal of Biological Chemistry, 1995. **270**(5): p. 2224-32.
81. Kido, H., A. Fukutomi, and N. Katunuma, *Tryptase TL2 in the membrane of human T4+ lymphocytes is a novel binding protein of the V3 domain of HIV-1 envelope glycoprotein gp 120*. Febs Letters, 1991. **286**(1-2): p. 233-6.
82. Katunuma, N.K., Hiroshi, *New physiological functions of intracellular proteases and their inhibitors*. Seikagaku, 62(1), 18-31, 1990.
83. Hattori, T.K., Atsushi; Takatsuki, Kiyoshi; Kido, Hiroshi; Katunuma, Nobuhiko, *Involvement of tryptase-related cellular protease(s) in human immunodeficiency virus type 1 infection*. FEBS Lett., 248(1-2), 48-52, 1989.
84. Barany, G. and F. Albericio, *Mild orthogonal solid-phase peptide synthesis*. Pept, 1990.
85. Staley, J.P. and P.S. Kim, *Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond*. Proc-Natl-Acad-Sci-U-S-A, 1992 Mar. **89**(5): p. 1519-1523.

86. Woo, D.D., et al., *Chemical synthesis in protein engineering: total synthesis, purification and covalent structural characterization of a mitogenic protein, human transforming growth factor-alpha*. Protein Engineering, 1989. **3**(1): p. 29-37.
87. Garsky, V.M., et al., *Chemical synthesis of echistatin, a potent inhibitor of platelet aggregation from *Echis carinatus*: synthesis and biological activity of selected analogs*. Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(11): p. 4022-6.
88. Clark-Lewis, I., et al., *Chemical synthesis, purification, and characterization of two inflammatory proteins, neutrophil activating peptide 1 (interleukin-8) and neutrophil activating peptide*. Biochemistry, 1991 Mar. **30**(12): p. 3128-3135.
89. Sueiras-Diaz, J. and J. Horton, *First solid phase synthesis of endothelial interleukin-8 [Ala-IL8]77 using BOC-TBTU chemistry with in situ neutralisation and comparison with synthesis of monocyte interleukin-8 [Ser-IL8]72 using the DCC-HOBt method*. Tetrahedron-Lett, 1992 May. **33**(19): p. 2721-2724.
90. Clark-Lewis, I., et al., *Automated chemical synthesis of a protein growth factor for hemopoietic cells, interleukin-3*. Science, 1986. **231**(4734): p. 134-9.
91. Hurlle, M.R., et al., *Denaturant-dependent folding of bovine pancreatic trypsin inhibitor mutants with two intact disulfide bonds*. Biochemistry, 1990 May. **29**(18): p. 4410-4419.
92. Schnolzer, M. and S.B. Kent, *Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease*. Science, 1992. **256**(5054): p. 221-5.
93. Flecker, P., *Refolding of proteinase inhibitor variants on trypsin-sepharose as a matrix with complementary structure*. Biol-Chem-Hoppe-Seyler, 1990. **371** : p. 53-58.
94. Cleary, S., Mulkerrin, M.G., Kelley, R.F., Biochemistry, 1989. **28**: p. 1884 - 1891.
95. Lu, H.S., et al., *Folding and oxidation of recombinant human granulocyte colony stimulating factor produced in *Escherichia coli*. Characterization of the disulfide-reduced intermediates and cysteine----serine analogs*. Journal of Biological Chemistry, 1992. **267**(13): p. 8770-7.
96. Ohta, Y., et al., *Pro-peptide as an intramolecular chaperone: renaturation of denatured subtilisin E with a synthetic pro-peptide [corrected] [published erratum appears in Mol Microbiol 1991 Dec;5(12):3090]*. Molecular Microbiology, 1991. **5**(6): p. 1507-10.
97. Baker, D., J.L. Silen, and D.A. Agard, *Protease pro region required for folding is a potent inhibitor of the mature enzyme*. Proteins, 1992. **12**(4): p. 339-44.
98. Ramos, C., J.R. Winther, and M.C. Kielland-Brandt, *Requirement of the propeptide for in vivo formation of active yeast carboxypeptidase Y*. Journal of Biological Chemistry, 1994. **269**(9): p. 7006-12.
99. Sorenson, P., et al., *The pro region required for folding of carboxypeptidase Y is a partially folded domain with little regular structural core*. Biochemistry, 1993. **32**(45): p. 12160-6.

100. Winther, J.R. and P. Sorensen, *Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity*. Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(20): p. 9330-4.
101. Winther, J.R., P. Sorensen, and M.C. Kielland-Brandt, *Refolding of a carboxypeptidase Y folding intermediate in vitro by low-affinity binding of the proregion*. Journal of Biological Chemistry, 1994. **269**(35): p. 22007-13.
102. Weissman, J.S. and P.S. Kim, *The pro region of BPTI facilitates folding*. Cell, 1992. **71**(5): p. 841-51.
103. Itoh, H., et al., *Mast cell protease inhibitor, trypstatin, is a fragment of inter-alpha-trypsin inhibitor light chain*. Journal of Biological Chemistry, 1994. **269**(5): p. 3818-22.
104. Stahl, S., et al., *Engineered bacterial receptors in immunology. [Review] [48 refs]*. Current Opinion in Immunology, 1993. **5**(2): p. 272-7.
105. Alexander, P., J. Orban, and P. Bryan, *Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G*. Biochemistry, 1992. **31**(32): p. 7243-8.
106. Alexander, P., et al., *Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures*. Biochemistry, 1992. **31**(14): p. 3597-603.
107. Boutillon, C., et al., *Synthesis, three-dimensional structure, and specific 15N-labelling of the streptococcal protein G B1-domain*. European Journal of Biochemistry, 1995. **231**(1): p. 166-80.
108. Lian, L.Y., et al., *Determination of the solution structures of domains II and III of protein G from Streptococcus by 1H nuclear magnetic resonance*. Journal of Molecular Biology, 1992. **228**(4): p. 1219-34.
109. Osborne, M.J., et al., *Sequential assignments and identification of secondary structure elements of the colicin E9 immunity protein in solution by homonuclear and heteronuclear NMR*. Biochemistry, 1994. **33**(41): p. 12347-55.
110. Derrick, J.P. and D.B. Wigley, *Crystal structure of a streptococcal protein G domain bound to an Fab fragment*. Nature, 1992. **359**(6397): p. 752-4.
111. Derrick, J.P., et al., *Crystallization and preliminary X-ray analysis of the complex between a mouse Fab fragment and a single IgG-binding domain from streptococcal protein G*. Journal of Molecular Biology, 1992. **227**(4): p. 1253-4.
112. Gallagher, T., et al., *Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR*. Biochemistry, 1994. **33**(15): p. 4721-9.
113. Gronenborn, A.M., et al., *A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G [see comments]*. Science, 1991. **253**(5020): p. 657-61.
114. Lian, L.Y., et al., *Mapping the interactions between streptococcal protein G and the Fab fragment of IgG in solution [letter] [published erratum appears in Nat Struct Biol 1994 Nov;1(11):829]*. Nature Structural Biology, 1994. **1**(6): p. 355-7.

115. Orban, J., P. Alexander, and P. Bryan, *Hydrogen-deuterium exchange in the free and immunoglobulin G-bound protein G B-domain*. *Biochemistry*, 1994. **33**(19): p. 5702-10.
116. Orban, J., et al., *Assessment of stability differences in the protein G B1 and B2 domains from hydrogen-deuterium exchange: comparison with calorimetric data*. *Biochemistry*, 1995. **34**(46): p. 15291-300.
117. Ybe, J. A. ., & Hecht, M. H. Sequence replacements in the central .beta.-turn of plastocyanin. *Protein Sci.*, (1996). **5**(5), 814-24.
118. Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J., & Wells, J. A. (1994). *A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues*. *Science*, **266**(5183), 243-7.
119. E. Atherton, R.C.Sheppard, (1989) *Solid Phase Synthesis: A Practical Approach*. IRL Press Oxford.
120. Abdelmoty, I., Albericio, F., Carpino, L. A., Foxman, B. M., & Kates, S. A. (1994). *Structural studies of reagents for peptide bond formation: Crystal and molecular structures of HBTU and HATU*. *Lett. Pept. Sci.*, **1**(2), 57-67.
121. Konig, W., and Geiger, R., (1970). *Chemische Berichte*, **103**, 788-798.
122. Le Nguyen, D., and Castro, B. In *Peptide Chemistry 1987* (ed. T Shiba and S Sakakibara), 231-238. Protein Research Foundation, Osaka.
123. Fields, C. G., Lloyd, D. H., Macdonald, R. L., Otteson, K. M., & Noble, R. L. (1991). *HBTU activation for automated Fmoc solid-phase peptide synthesis*. *Peptide-Res*, **4**(2), 95-101.
124. Coste, J., Le-Nguyen, D., and Castro, B., (1990). *Tetrahedron Letters*, **31**, 205-208.
125. Carpino L A, (1995). *1-Hydroxy-7-azabenzotriazole. An Efficient Peptide Coupling Additive*. *Journal of American Chemical Society*, **115**, 4397-4398
126. Carpino, L., El-Faham, A., Minor, C. A., & Albericio, F. (1994). *Advantageous applications of azabenzotriazole (triazolopyridine)- based coupling reagents to solid-phase peptide synthesis*. *J. Chem. Soc., Chem. Commun*, **2**, 201-3.
127. (a) Carpino, L. A. ., El-Faham, A., & Albericio, F. (1995). *Efficiency in Peptide Coupling: 1-Hydroxy-7-azabenzotriazole vs. 3,4-Dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine*. *J. Org. Chem.*, **60**(11), 3561-4.  
 (b) Carpino, L. A. , Ionescu, D., & El-Faham, A. (1996). *Peptide Coupling in the Presence of Highly Hindered Tertiary Amines*. *J. Org. Chem.*, **61**(7), 2460-5.  
 (c) Carpino, L. A. & El-Faham, A. (1994). *Effect of Tertiary Bases on O-Benzotriazolyluronium Salt-Induced Peptide Segment Coupling*. *J. Org. Chem.*, **59**(4), 695-8.
128. Ehrlich, A., Rothmund, S., Brudel, M., Beyermann, M., & Carpino, L A. (1993). *Synthesis of cyclic peptides via efficient new coupling reagents*. *Tetrahedron Lett*, **34**(30), 4781-4.
129. Carpino, L. A., El-Faham, A., & Albericio, F. (1994). *Racemization studies during solid-phase peptide synthesis using azabenzotriazole-based coupling reagents*. *Tetrahedron Lett.*, **35**(15), 2279-82.

130. Karle I. L., Balamam, B., *Structural Characteristics of alpha-helical Peptide Molecules containing Aib Residues*. Biochemistry. ,1990. **29**(29): p. 6747-6756.
131. Raganathan D., Shah K., Vaish N., (1992). *An Exceptionally Mild and Efficient Route to Dehydroalanine Peptides*. J. Chem. Soc., Chem. Commun, 1145-1147.
132. Karle, I. L., Flippen-Anderson, J. L., Sukumar, M., Uma K., & Balamam, P. (1992). *Modular Design of Synthetic Protein Mimics. Crystal Structure of Two Seven-Residue Helical peptide Segments Linked by Aminocaproic Acid*. (1991) Journal of American Chemical Society, 113, 3953-3956.
133. Shortle, D., & Sondek, J. (1995). *The emerging role of insertions and deletions in protein engineering*. [Review] [40 refs]. Current Opinion in Biotechnology, 6(4), 387-93.
134. Creighton T., *Proteins: Structures and Molecular Properties*. 1993. 183-186. W.H. Freeman.
135. Robinson S.R., Sligar S.G., (1992) *Electrostatic Stabilization in Four-Helix Bundle Proteins*. 2(5): 826-837.
136. O'Neil, K. T., Hoess, R. H., Raleigh, D. P., & DeGrado, W. F. (1995). *Thermodynamic genetics of the folding of the B1 immunoglobulin-binding domain from streptococcal protein G*. Proteins, 21(1), 11-21.
137. Lodder M., Golovine S., and Hecht S.(1997). *Chemical Deprotection Strategy for the Elaboration of Misacylated Transfer RNA's*. J. Org. Chem., 62, 778-779.
138. Hickman, A. B. ., Palmer, I., Engelman, A., Craigie, R., & Wingfield, P.(1994). *Biophysical and enzymic properties of the catalytic domain of HIV-1 integrase*. J. Biol. Chem., 269(46), 29279-87.
139. Chrusciel, R. A., & Romines, K. R.(1997). *Recent developments in HIV protease inhibitor research*. Expert Opin. Ther. Pat., 7(2), 111-121.
140. Etzhorn,F.A., Guo, T., Lipton, M.A., Goldberg A.D., and Bartlett P.A.,(1995). *Cyclic Hexapeptides and Chimeric Peptides as Mimics of Tendamistat..* Journal of American Chemical Society, 115, 4397-4398
141. Sato, K., Nagai, U. J., J. Chem. Soc., Perkin Trans. 1 (1986), 1231-1236.
142. Nicholson H., Becktel W. J., Matthews B.W., *Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles*. Nature, 1988. **336**(6200): 651-656.
143. Huyghues-Despointes, B.M.P., Scholtz J.M., Baldwin R.L. *Helical Peptides with Three Pairs of Asp-Arg and Glu-Arg Residues in Different Orientations and Spacings*. Protein Science, 1993, 2:80-85.
144. Fersht AR, Serrano L., *Principles of Protein Stability Derived from Protein Engineering Experiments*. Curr. Opin. Struct. Biol. (1993). 3: 75-83.
145. Flanagan, J. M., Kataoka, M., Shortle, D., & Engelman, D. M. (1992). *Truncated staphylococcal nuclease is compact but disordered*. Proceedings of the National Academy of Sciences of the United States of America, 89(2), 748-52.
146. Tam, J.P., Guo, T., Wu, C-R., Liu W., and Zhang J-W.,(1991). *Disulphide Bond Formation in Peptides by Dimethyl Sulphoxide*. Journal of American Chemical Society, 113, 6657-6662.