Microcalorimetry of Protein Stability and Interactions

By Stephen R. Robertson

A thesis submitted in partial fulfilment for the degree of Ph.D.

University of Glasgow

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

"I owe a lot to my teachers and mean to pay them back some day"
Stephen Leacock.

Acknowledgements

I would like to thank my supervisor Dr Alan Cooper for his help, his advice and his special brand of encouragement at every stage of this work, without which it would not have been possible. I would also like to thank the other members of the Biophysical Chemistry group, Margaret Nutley, Michelle Lovatt, Deborah McPhail, Abdul Wadood Hamid and particularly Drs Steven Ford and Alan McAlpine.

I thank the following people who provided me with samples, and much information and advice: Prof Peter Stockley (Leeds; MS2-viral coat protein and AhrC), Dr Nicola Stonehouse (Leeds; MS2-viral coat protein), Helen Mastico (Leeds; MS2-viral coat protein), James Murray (Leeds, MS2-synthetic RNA 19mer), Coleen Millar (Leeds; AhrC), Prof Geoff Kneale (Portsmouth; Pf1-gV), Dr Ken Davis (Portsmouth; Pf1-gV) and Dr Thomas Creighton (EMBL, Heidelberg; 3SS-α-lactalbumin).

I also thank Prof Nick Price (Stirling) for help concerning fluorescence spectroscopy and Prof Laurence Barron, Dr Steven Ford and Gary Wilson (all Glasgow) for their helpful discussions on α-lactalbumin, including early viewing of their Raman and ROA results.

Finally I'd like to thank all of my family and friends for support, understanding and cups of coffee, especially my wife Lynne.

Abstract

Protein stability and interactions in a number of systems have been examined using a range of calorimetric and spectroscopic methods.

The thermodynamics of unfolding of bovine α-lactalbumin (BLA) was studied using differential scanning calorimetry (DSC) under a variety of pH and calcium ion concentration conditions. Values for ΔG^{o}, ΔH^{o} and ΔS^{o} under each of these conditions were obtained, and these data were used to calculate a value for ΔC_p. The information thus obtained for the wild type (WT) protein was then compared to data obtained under similar conditions for a modified form of the protein (3SS-BLA) which contains only three of the usual four disulfide bonds, the other (C6-C120) having been specifically reduced and blocked. In this way it was possible to gain information on the role of this disulfide bond for the stability of the protein. In line with current theory it was found that the modified form of the protein was less stable and this was accountable for in terms of entropy. However, the difference in unfolding entropy between the two forms ($\Delta\Delta S$) was less than may be anticipated under the current theory. Fluorescence spectroscopy was employed to show whether or not removal of the disulfide bond caused any significant alteration in the dynamics of the native state. It was found that 3SS-BLA was significantly more flexible than its wild type counterpart, and this may account for the unexpectedly low $\Delta\Delta S$.

DSC was also used to to study the effect of heavy and light water on the stability of hen egg white lysozyme (HEWL). It was found in heavy water the protein was enthalpically

destabilised, but this was more than offset by an entropic stabilisation, a finding which is in line with model compound studies. Through this work it was also noticed that thermal aggregation of proteins increases with concentration.

DSC was further employed to study partially proteolyzed Pf1 gene five protein (Pf1-gVp) and complex (Pf1-gVc). It was found that removal of the flexible N-terminal domain had no apparent effect on thermal stability of the complex and therefore is unlikely to be involved in protein-protein interactions in the complex. The T_m of this complex was found to be concentration dependent which shows that the system undergoes cooperative dissociation and unfolding.

Isothermal titration calorimetry (ITC) was used alongside DSC to study the coat protein of the bacteriophage MS2. ITC measurements on the binding of a specific RNA 19-mer to the coat protein were made, forming a basis for future work. The T_m measured in DSC was found to be comparable to literature estimates.

ITC was also employed to study the binding of arginine (Arg) to the protein AhrC. It was found that three molecules of Arg bind per protein hexamer to a strong site and that there is a second, much weaker site or sites which bind a large number of molecules.

Contents

Acknowledgements..... I

Abstract..... II

Table of Contents..... IV

Chapter One Introduction..... 1

1.1 Proteins..... 1

1.1.1 Introduction..... 1

1.1.2 What are Proteins?..... 2

1.2 Factors influencing protein stability..... 3

1.2.1 Hydrogen bonding..... 4

1.2.2 Hydrophobic effect..... 5

1.2.3 Electrostatic interactions..... 7

1.2.4 Van der Waals interactions..... 9

1.2.5 Disulfide bridges..... 10

1.3 Factors influencing protein interactions..... 13

1.4 Protein dynamics..... 15

1.5 Some basic thermodynamics..... 15

1.6 Techniques to be used..... 22

1.6.1 Microcalorimetry..... 22

1.6.2 Protein intrinsic fluorescence..... 23

1.6.3 Protein fluorescence quenching..... 26

1.7 Aims of this thesis..... 30

Chapter Two Microcalorimetry..... 31

2.1 Introduction..... 31

2.2 Differential scanning calorimetry..... 32

2.2.1 The DSC instrument..... 32

2.2.2 The information available through DSC..... 34

2.2.3 Analysis of DSC data..... 42

2.2.4 Models used in DSC analysis..... 46

2.3 Isothermal titration calorimetry..... 51

2.3.1 The ITC instrument..... 51

2.3.2 The information available through ITC..... 54

2.3.3 Control experiments..... 57

2.3.4 Analysis of ITC data..... 59

2.3.5 Models used in ITC analysis..... 61

Chapter Three Materials and methods..... 65

3.1 Proteins..... 65

3.1.1 Preparation of 3SS-α-lactalbumin..... 66

3.1.2 Preparation of Pf1-gV protein and complex..... 66

3.1.3 MS2 preparation and purification..... 67

3.1.4 AhrC preparation and purification..... 69

3.2 Chemicals..... 71

3.2.1 General chemicals..... 71
3.2.2 Chemicals used in protein preparations..... 71
3.3 Differential Scanning Calorimetry..... 72
3.3.1 Procedure for obtaining a thermogram..... 72
3.3.2 Analysis of DSC data..... 74
3.3.3 Temperature axis calibration..... 75
3.3.4 Excess heat capacity axis calibration..... 75
3.4 Isothermal titration Calorimetry..... 76
3.4.1 Procedure for obtaining a binding isotherm..... 76
3.4.2 Analysis of ITC data..... 77
3.4.3 Calibration..... 77
3.5 Fluorescence spectroscopy..... 78
3.5.1 Intrinsic fluorescence..... 78
3.5.2 Fluorescence quenching..... 79
3.6 Protein concentration estimation..... 80
3.6.1 UV absorbtion of proteins..... 80
3.6.2 Estimation of protein extinction coefficients..... 82
3.6.3 Coomassie brilliant blue assay..... 83
3.6.4 Warburg and Christian method..... 84

Chapter Four α-Lactalbumin..... 86
4.1 Introduction..... 86
4.1.1 Previous work on HEWL..... 92
4.1.2 Previous work on BLA..... 94

4.2 Thermodynamics of the unfolding of WT- and 3SS-BLA.....95

4.3 Fluorescence spectroscopy.....116

4.4 Discussion of BLA results.....126

Chapter Five Solvent effects on lysozyme folding stability.....128

5.1 Introduction.....128

5.2 pH versus pD.....132

5.3 Hydrogen exchange in protein interior.....134

5.4 Thermodynamics of lysozyme unfolding in H_2O and D_2O.....140

5.5 Discussion of lysozyme results.....142

Chapter Six Pf1 gene five protein.....144

6.1 Introduction.....144

6.2 Thermodynamics of the unfolding of proteolyzed Pf1-gVc.....150

6.2.1 Thermal stability of the proteolyzed Pf1-gVc.....150

6.2.2 Concentration dependence of thermal transitions.....150

6.2.3 Analysis of Pf1-gVc data.....152

6.3 Effect of aggregation on thermal transitions.....158

Chapter Seven MS2 and AhrC.....161

7.1 MS2 viral coat protein.....161

7.1.1 Introduction.....161

7.1.2 Thermodynamics of the unfolding of MS2 viral coat protein.....166

7.1.3 Thermodynamics of RNA binding by MS2 viral coat protein.....170

"Of making many books there is no end; and much study is a weariness of the flesh."

The Bible, Ecclesiastes.

"Outside of a dog a book is a man's best friend. Inside of a dog it's too dark to read."

Groucho Marx.

Chapter one

Introduction

1.1 Proteins

1.1.1 Introduction

Proteins play a central role in the existence and function of all biological organisms. There are proteins for nearly every job, with proteins being responsible for almost all of the chemistry done in or by the organism (enzymes), for transport of molecules to the required site (e.g. hemoglobin), for regulation of gene transcription and for structural elements (e.g. hair, skin, muscle etc), while others act as gateways, permitting or preventing substances from entering or leaving cells (membrane proteins). Proteins are used in a remarkable number of ways, it is the way proteins are constructed which make this diversity possible.

1.1.2 What are Proteins?

Proteins are macromolecules consisting of long chains of amino acids in an order dictated by the gene. There are twenty common amino acids and the sequence in which they occur is called the protein's primary structure, with different proteins having different primary structures. Proteins fall into three main categories: fibrous, membrane and globular, of which the latter type will be discussed in this thesis.

In globular proteins the polypeptide chains collapse upon themselves in very specific ways, folding up to form the protein's native or folded state. The native state of a protein is its three dimensional conformation, and is commonly described by reference to a further three levels of structure (in addition to the primary structure) within the protein. Secondary structure elements, such as α-helix, β-sheet and certain turns and loop structures fit together to form tertiary structure which gives monomeric proteins their overall shape, finally association of monomers into larger aggregates occurs in some proteins resulting in what is known as quaternary structure.

The overall shape of the protein and hence its native state is extremely important for its function. The native state of a protein depends on its primary structure, and hence each protein forms a unique tertiary fold (although there are families where, despite differences in primary structure, the native fold is more or less the same, with only minor differences). The native state is generally only marginally more stable than the unfolded state, and relatively small perturbations of the environment lead to the protein becoming denatured, or unfolded.

1.2 Factors influencing protein stability

The factors which result in the stability of the native state of a protein are still only poorly understood. The subtle interplay of the different forces creates a picture so complex that it is only slowly being revealed.

Folding of the protein chain is, in principle, entropically unfavourable. Restriction of the inherently flexible polypeptide chain to a more limited region of conformational space involves a reduction in conformational entropy (negative ΔS, positive contribution to ΔG) which must be offset by other more favourable free energy contributions if folding is to be stable. These additional contributions, arising from the multiplicity of interactions between side-chain and/or backbone groups, may be either enthalpic or entropic depending on the nature of the interaction, and will also contain major contributions from solvent (water) effects.

However, the balance is even more subtle than that. The stability of the native state of a protein must be measured relative to the stability of the unfolded state, it is the magnitude of the difference in stability which is significant. Therefore any effect which stabilises the unfolded state reduces the magnitude of the difference and consequently destabilises the native state, similarly any effect which destabilises the unfolded state effectively stabilises the native state. So all of the forces exerted in the native state must be sufficient to counter the possible interactions in the unfolded state.

A number of non-covalent interactions can be identified as possibly playing a part in stabilising the native fold. These are hydrogen bonding, hydrophobic effect, electrostatic interactions and van der Waals interactions. In addition to these one type of

covalent interaction is extremely important, the disulfide bridge, which forms a crosslink between remote parts of the polypeptide chain, and may have a strong influence on structure and stability.

For details on many of these forces the following reviews are useful: Kauzmann, 1959; Dill, 1990; Rose & Wolfenden, 1993.

1.2.1 Hydrogen bonding

Hydrogen bonds are formed when two electronegative atoms (usually oxygen or nitrogen in proteins) share a common hydrogen atom. The hydrogen atom is covalently bound to one atom, which acts as an electron acceptor, and the H bond is formed to the other atom, which acts as an electron donor. The strength of the bond depends on the electronegativity of the atoms and the bonding length and orientation, the more the three atoms deviate from co-linearity the weaker the bond. Typically the H bond has a free energy of 25-80 $kJmol^{-1}$ in the gas phase, where clearly an ideal geometry can be maintained (Weiner, 1984; Meot-Nir & Sieck, 1986).

H bonds have long been believed to be important in protein folding (Mirsky & Pauling, 1936; Pauling, 1939; Pimental & McClellan 1960) and some have concluded that H bonding is the principal source of stabilisation in proteins (Privalov & Gill, 1988). However, competition between intra-molecular H bonds within proteins and inter-molecular H bonds between protein and water reduces the effective strength of the H bond markedly, so much so that H bonds between polar groups in water form with extremely low enthalpies (Klotz & Franzen, 1962; Schellman, 1955). The reported

values for H bond in proteins have been favourable, but small, ranging from -2.0 to -18.8 kJmol-1 (Fersht et al, 1985; Doig & Williams, 1992; Shirley et al, 1992).

The strength of the intra-molecular interaction may be enhanced by entropic considerations. Consider a group, with H bonding potential, exposed in the unfolded state which forms an H bond to a water molecule from the bulk solvent. This bound water molecule is released upon folding when the group responsible for binding it becomes involved in an intra-molecular H bond. The entropy of the protein is decreased by the intra-molecular bond, however the entropy of the solvent is increased which may lead to an overall favourable entropy change in the system. Again the system has to be defined to include the solvent.

In balance, however, it would appear that hydrogen bonding gives only marginal stability to the folded structure of a protein. While the largest role played may be one of favouring particular structures over other possibilities which may arise given the other possible forces on their own, α-helix and β-sheet structures depend heavily on H bonding.

1.2.2 Hydrophobic effect

The hydrophobic effect has exercised much discussion among chemists for some considerable time (Traube, 1891). The effect can be seen most clearly when considering the transfer of non-polar solute molecules into the aqueous phase. Small hydrocarbon molecules tend to associate into oil drops when introduced into water. The reasons

behind this are complex, and depend on the nature of the solvent as much as, or perhaps more, than on the nature of the solute.

The hydrophobic effect is not really a bond between the two interacting groups in the conventional sense (and so is non directional). When a hydrophobic molecule is introduced into bulk water it is unable to form hydrogen bonds to the water. In order to compensate for this water molecules in the neighbourhood of the non-polar molecule form as many H bonds to each other as possible, resulting in a more ordered structure than in bulk water (Frank & Evans, 1945) and therefore the entropy is decreased. The entropy is further decreased as the non-polar molecule itself loses much of its original rotational and translational entropy (Aronow & Witten, 1960; Howarth, 1975).

In order to minimise these effects non-polar molecules are excluded from the water as much as is possible, forming the classical oil drop, and decreasing the hydrophobic surface exposed to the solvent. The hydrophobic effect therefore results, not from attractive forces between the non-polar molecules but rather it results from the mutual repulsion of solute and solvent molecules which is driven, in part at least, by solvent entropy requirements.

What this means for proteins is that they tend to fold in such a way that the non-polar side chains, such as those of valine, leucine, iso-leucine and phenylalanine etc. form a hydrophobic core, from which water is largely excluded.

The evidence for the involvement of the hydrophobic effect is good. Transfer of hydrocarbons into polar environments is accompanied by large positive changes in heat capacity, which has been explained in terms of gradual melting of the structured water

layers formed around the solute. Similar large changes in heat capacity are observed upon unfolding of protein molecules. Secondly addition of cyclodextrin compounds have an adverse affect on stability (Cooper, 1992), this has been interpreted in terms of hydrophobic side chains on the unfolded protein being solvated by the cyclodextrin molecule, removing it from contact with water, and therefore reducing the magnitude of the hydrophobic effect experienced.

That the hydrophobic effect could drive folding seems likely, that it would necessarily lead to a unique folded state seems less so. It may well prove to be the case that while the hydrophobic effect drives the overall tendency to fold it is other interactions, such as H bonds, which enforce specific configurations.

1.2.3 Electrostatic interactions

As every schoolchild knows like charges repel, unlike charges attract, the potential between any two point electric charges is given by Coulomb's law:

$$F = \frac{q_1 q_2 k}{\varepsilon r} \qquad (1.1)$$

Where q_1 and q_2 are the charges on the two points, k is a constant, ε is the dielectric constant and r is the distance between the two points. As can be readily seen from equation 1.1 the magnitude of the potential between the two groups depends on the charge; so in proteins, groups which carry a complete, permanent charge are most strongly affected, however those which carry a partial charge either as a result of a permanent or induced dipole are also affected to a degree depending on the net charge.

Proteins are known to be denatured on addition of either acid or base, since this effects the overall charge on the protein it was assumed (Mirsky & Pauling, 1936) that the driving force behind folding was electrostatic interactions. As the extremes of pH are approached the net charge on the protein increases, and since the charge density on a folded protein molecule is greater than on an unfolded protein this results in denaturation.

Ion pairing is where specific charged groups interact to form a salt bridge. This was first advanced as the dominant force in protein folding in the 1930's, when the belief was that ion pairing between carboxyl and amino on side chains was responsible for stabilising the native state (Cohn et al, 1933; Mirsky & Pauling, 1936; Eyring & Stearn, 1939). The situation can most readily be envisaged by considering two charged groups which are separated in water. Water molecules have a significant, permanent dipole which results in the charged groups becoming surrounded by water molecules. The water molecules in the vicinity therefore become compressed and oriented in particular ways. When the two charges are allowed to come together their electric fields no longer have such a strong influence on the behaviour of the water molecules and so the solvent becomes much less compressed and structured and so a large increase in entropy accompanies the association of the charged groups. This can also be seen from the temperature dependence of the dielectric constant (ε) of water. Reducing the temperature forces the water molecules into a more structured arrangement resulting in a smaller entropy difference and therefore a smaller force, i.e. the dielectric constant of water is seen to increase with decreasing temperature.

In proteins the magnitude of the electrostatic force is increased where charged groups are interacting in environments from which water is excluded. Water is very polar, and so has a relatively high dielectric constant of 80, while the interior of the protein has a significantly smaller dielectric constant. As can be seen from equation 1.1 the magnitude of the effect falls off with distance and with increasing dielectric constant of the solvent, so the influence of charged groups depends on where those groups are situated and how polar the environment around them is. However, it is known that few ion pairs become buried upon protein folding (Barlow & Thornton, 1983) and that only surface ion pairs contribute to protein stability. This is likely to be due to the unfavourable energy associated with burying charged groups in the hydrophobic interior of the protein.

Salts attenuate the attraction between ion pairs by shielding the charge, whereas pH directly alters the charge on acidic or basic groups. That addition of salt, and variation of pH, close to the isoelectric point (the point where the protein has zero net charge and so variation around this point leads to a change from a net positive to a net negative charge on the protein) has little effect on the stability of the protein (Tanford, 1968; Hermans & Scheraga, 1961) leads to the conclusion that ion pairing is not the dominant force in protein folding.

1.2.4 Van der Waals interactions

Van der Waals interactions are a particular case of electrostatic interactions arising when transient dipoles in one group of atoms within the molecule induce dipoles in groups in the vicinity. Van der Waals interactions are extremely short range varying

as the inverse of distance to the power six. However, packing in the protein interior is extremely close, in fact it has been noted that packing in a protein's interior is as close as within crystals of small organic compounds (Richards, 1977) allowing these forces to come into play.

Since van der Waals interactions are constantly present, whether groups have a permanent dipole or not, it is impossible to distinguish these forces from the other non-covalent interactions and as a result values measured for other interactions probably contain a contribution from van der Waals interactions.

1.2.5 Disulfide bridges

Disulfide bridges are formed when the sulfhydryls of two cysteine side chains in a protein are oxidised to form a covalent crosslink between remote parts of the polypeptide chain. Unlike the non-covalent interactions described above the covalent nature of the disulfide bridge confers much more extreme limitations on the geometry adopted by the protein in its locality. The sulfur atoms must be 2.05 ± 0.03 Å apart, the angle between the sulfur atoms and the β-carbon atom must be approximately 103^{o} and the C-S-S-C dihedral angle must be close to $\pm 90^{o}$.

The thermodynamic role played by disulfide bridges and other crosslinks in stabilising the native state of proteins is the subject of much discussion (See Betz, 1993 for recent review). The generally held view is that the disulfide bridge has little direct effect on the native state of the protein, but rather increases the free energy gap between native and unfolded states by decreasing the stability of the denatured state. When a

disulfide bridge is formed two remote parts of the polypeptide chain are locked together thus reducing the region of conformational space which the unfolded chain can sample (Flory, 1956). Much experimental evidence supports this hypothesis (Anfinsen & Scheraga, 1975; Ikeguchi et al, 1992; Cooper et al, 1992).

The magnitude of this effect is predicted to depend on the number of residues in the loop enclosed by the disulfide bridge (Schellman, 1955; Poland & Scheraga, 1965; Pace et al, 1988), and is calculated by relating the entropy change upon removal of a single disulfide bridge in the unfolded state to the probability that both of the free cysteines could occupy a certain volume of space at the same time in the absence of the constraints of the crosslink. This is described by equation 1.2

$$\Delta S = -R \ln\left(\frac{3}{(2pl^2N)^{3/2}}\right)\nu_s \qquad (1.2)$$

where ΔS is the change in entropy, R is the gas constant, N is the number of statistical segments in the loop (i.e. the number of amino acid residues), l is the average length of the segments (i.e. 3.8 Å) and ν_s is the volume element. The magnitude of the volume element which should be used has been variously given as $\nu_s = 2.83$ Å^3 (Schellman, 1955) $\nu_s = 5.17$ Å^3 (Poland & Scheraga, 1965) and $\nu_s = 57.9$ Å^3 (Pace et al, 1988).

It is assumed that the difference in entropy between the unfolded state with and without the disulfide bridge intact is equal to the difference between the unfolding entropy of native state with and without the disulfide bridge intact (see figure 1.1). This is based on the assumption that the absolute entropy of the native state with disulfide removed is not significantly different from that with disulfide intact. This assumption is challenged by Karplus et al. (1987).

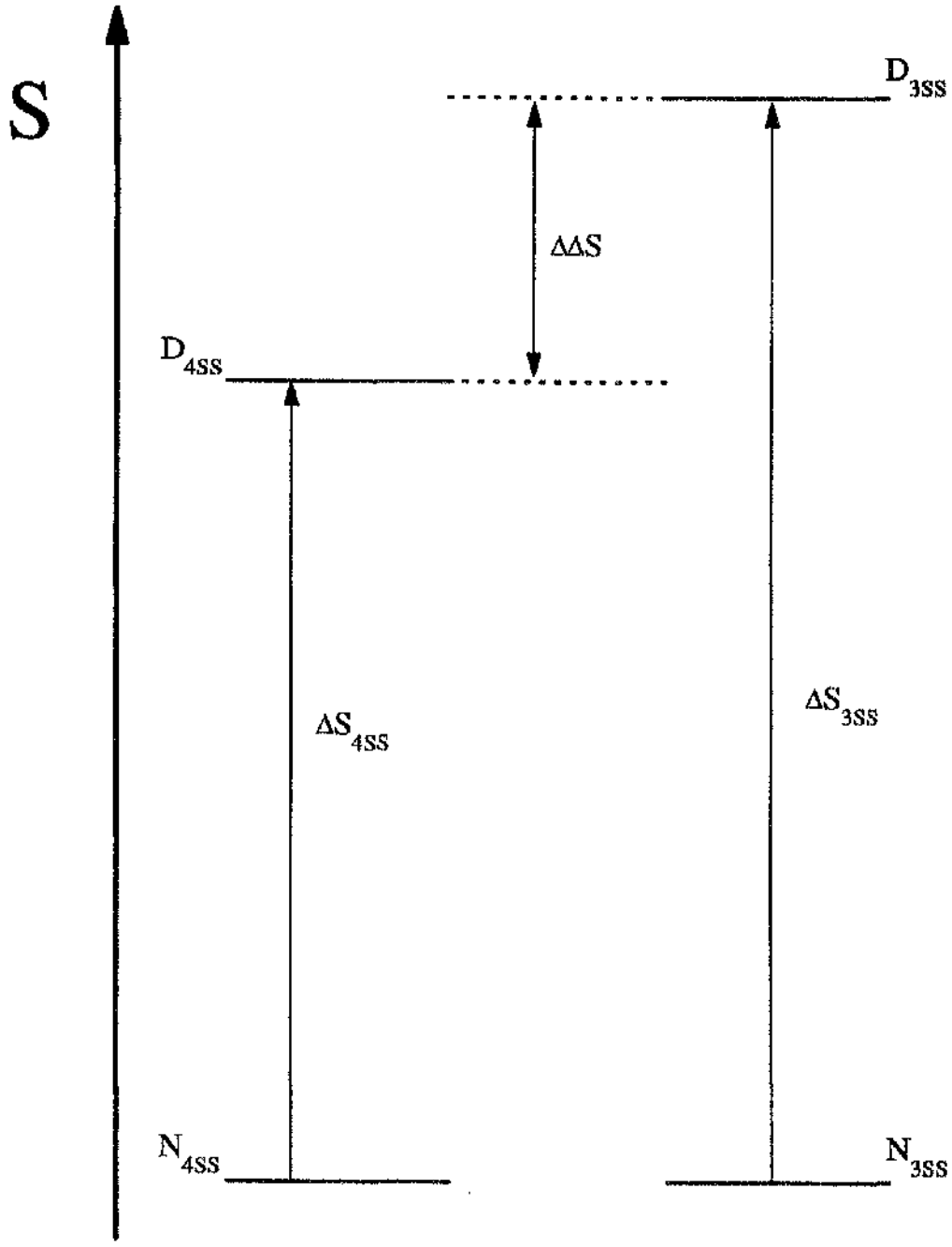

Figure 1.1 Schematic representation of the entropy changes involved in the denaturation of WT & 3SS-BLA.

It has been asserted by Creighton (Creighton, 1975; Creighton, 1983; Creighton & Goldenberg 1984; States et al, 1984; Goldenberg & Creighton, 1985; States et al, 1987; Creighton, 1988;) that this theory is incomplete, and he cites as an example the protein bovine pancreatic trypsin inhibitor (BPTI), which is stabilised to a much greater extent by its three disulfide bridges than the theory would predict.

A completely different hypothesis has been advanced by Doig & Williams (Doig & Williams, 1991), with some empirical backing (Kuroki et al, 1992a). In this proposal the increase in conformational entropy of the unfolded state upon removal of the disulfide bridge is more than compensated for by a concomitant decrease in the entropy of the solvent due to solvent ordering around the hydrophobic residues exposed when the constraint is released. They propose that the stabilisation experienced is due to an enthalpic effect caused by a more favourable hydrogen bonding network being set up in the disulfide intact protein.

1.3 Factors influencing protein interactions

For proteins to fulfil their purpose they must interact with the appropriate target, which may be one or more of RNA, DNA, other proteins, small molecules or metal ions. The factors which influence the association of the protein with these ligands are the same list of non-covalent interactions which are responsible for protein stability. Occasionally ligand and protein become covalently bound, but even so initial recognition is driven by the non-covalent forces already described.

Clearly the situation is somewhat clearer when small molecules and ions are the ligands. Metal ion binding is driven by electrostatic interactions. Specificity in these interactions is controlled by the charge on the ion and the size of the binding site. Ions with significantly larger or smaller ionic radii bind less tightly than the ions specific to that protein. The forces involved in small molecule binding depend on the nature of the molecule in question, and possible conformational changes within either the protein or the ligand (indeed such conformational changes within ligands may enhance the function of the protein).

Where the ligand is a nucleic acid molecule, proteins are able to recognise and bind to specific sites along the length of the chain. The basis for this specificity is still only poorly understood. Stretches of RNA can adopt particular conformations, which may help in recognition, but many proteins are able to recognise particular sections of the DNA duplex (Luisi & Sigler, 1990; Sauer et al, 1990; Pabo & Sauer, 1992). In DNA it is common for the recognition site to be repeated, for example in the methionine repressor (MetJ) system in *Escherichia coli* the binding site consists of multiple tandem repeats of an eight base pair sequence (Phillips et al, 1989; Somers & Phillips, 1992) which is able to bind two dimers of the protein next to each other. Cooperative interactions between the first and second dimer bound means that the second dimer to bind does so with greater affinity (Phillips et al, 1989; Cooper et al, 1994; He et al, 1994; Mao et al, 1994).

1.4 Protein dynamics

Of equal importance for the function of proteins is their flexibility. It is tempting to think of proteins as being like the pictures given by X-ray crystallography which are extremely useful, but the temptation to think of them as absolutes is strong. However these are stills, allowing us to see averaged structures. If X-ray movies were available these would tell us a lot more of interest; that is, what the protein is actually doing in solution. The protein is not stuck, frozen into one particular conformation but rather it is dynamic, moving around, sampling a large number of very similar conformations. This flexibility contributes significantly to the absolute entropy of the native protein and probably in many instances to its function also (Cooper, 1976; Sturtevant, 1977; Cooper, 1980; Cooper, 1984; McCammon & Harvey, 1987; Brooks et al, 1988) and so it is of interest, when studying the thermodynamics of proteins, to consider their dynamics too.

1.5 Some basic thermodynamics

Presented here are some of the basic thermodynamic concepts which will be used throughout this thesis. It is not my intention to insult the reader, however the start is invariably the best place to start and so I make no excuse for presenting some very basic material here. The text by E. Brian Smith is very helpful (Smith, 1992).

Every substance has an internal energy, this is the sum of all the kinetic and potential energy within the substance. When a physical or chemical change occurs, resulting in transfer of energy from or to the system's surroundings the law of conservation of energy (first law of thermodynamics) tells us that this must be accompanied by an equal and

opposite change to the internal energy of the system. The internal energy change (ΔU) is given by:

$$\Delta U = q + w \tag{1.3}$$

where q is the heat transferred to the system and w is the work done on the system. The term work includes all forms of work, even that done on (or by) the atmosphere when expansion (or contraction) takes place, so called PV work. This is generally the only kind of work chemical reactions are allowed to do, so we find:

$$w = -P\Delta V \tag{1.4}$$

If no volume change is allowed ΔV is zero and the internal energy change is simply:

$$\Delta U = (q)_v \tag{1.5}$$

where the subscript v denotes constant volume.

If on the other hand the reaction occurs at constant pressure then the internal energy change is given by:

$$\Delta U = (q)_p - P\Delta V \tag{1.6}$$

which can be written as:

$$U_B - U_A = (q)_p - P(V_B - V_A) \tag{1.7}$$

where the subscripts denote the states before and after the reaction has taken place. Rearranging equation 1.7 gives:

$$(q)_p = (U_B + PV_B) - (U_A + PV_A) \tag{1.8}$$

The enthalpy of a substance is defined as:

$$H = U + PV \tag{1.9}$$

So at constant pressure we get:

$$H_B - H_A = \Delta H = (q)_p \tag{1.10}$$

Entropy (S) is a measure of the disorder of a substance, and is defined in the following way:

$$\Delta S = S_B - S_A = \int_A^B \frac{dq_{rev}}{T} \tag{1.11}$$

where q_{rev} is the heat transferred when the change is reversible i.e. at equilibrium. The molecular basis of entropy can be found from statistical thermodynamics to depend on the number of ways (W) a system can be arranged without influencing the overall energy of the system, and is given by:

$$S = k \ln W \tag{1.12}$$

where k is the Boltzmann constant. This arises from considerations of probability and looking at the simple game of rolling two dice helps us to understand what is meant by the term number of ways. When we roll two dice there is only one way to get a total of twelve (two sixes) however there are two ways to get eleven (five and six or six and five) so the probability of finding we've rolled eleven is twice the probability of rolling twelve, and so on.

A change to the system will occur in such a way as to obtain the lowest energy, however there is likely to be a larger number of ways of obtaining a higher energy, and so the system compromises these two requirements and resides in the most probable energy level.

In calorimetric experiments it is the changes in heat, at constant pressure, which accompany the process under study (e.g. binding, unfolding, etc) and therefore the enthalpy change which we measure. However, in any equilibrium process the events which take place must satisfy the two (generally) mutually contradictory requirements of enthalpy and entropy at the same time (see previous paragraph). Enthalpy requirements mean that the system will change in such a way as to attain as low an energy as possible, while entropy requirements mean that the system changes in such a way as to maximise the disorder, and therefore the number of ways of arranging the energy. The result is a balance between the two; a reaction is spontaneous if and only if the direction of change is such as to minimise the free energy (ΔG) i.e. ΔG is negative (by convention). ΔG depends on the enthalpy and entropy according to the Gibbs-Helmholtz equation:

$$\Delta G^{o} = \Delta H^{o} - T\Delta S^{o} \qquad (1.13)$$

where ΔG^0 is the free energy under standard conditions (pressure = 1atm and all concentrations = 1M), ΔH^0 is the standard enthalpy change, ΔS^0 is the standard entropy change and T is the absolute temperature. The same equation holds for conditions other than standard, replacing ΔG^0, ΔH^0 and ΔS^0 with ΔG, ΔH and ΔS respectively.

The heat capacity of a substance is defined as being the amount of energy required to raise 1kg of the substance through 1 °C, i.e.

$$C = \frac{dq}{dT} \tag{1.14}$$

where C is the heat capacity, q is the heat and T is the temperature. If the heat capacity is measured at constant volume then combination of equations 1.5 and 1.14 gives:

$$C_v = \left(\frac{\partial U}{\partial T}\right)_v \tag{1.15}$$

If on the other hand the heat capacity is measured at constant pressure combination of equations 1.10 and 1.14 gives:

$$C_p = \left(\frac{\partial H}{\partial T}\right)_p \tag{1.16}$$

From here all discussion of heat capacity will be limited to heat capacity at constant pressure, C_p, since it is under these conditions that most experiments are carried out, including those presented here. The heat capacity at constant pressure is generally

somewhat larger than at constant volume due to the extra work required for the expansion of the system against the atmosphere, although the two are numerically very similar for liquids and solids where the expansion is small. For ideal gases the two are related as follows:

$$C_p = C_v + nR \tag{1.17}$$

where n is the number of moles of the gas and R is the gas constant.

The other way to look at heat capacities is to say that both entropy and enthalpy and therefore free energy are related to the heat capacity of the system. The relationships between ΔH and ΔC_p and between ΔS and ΔC_p are given by the equations:

$$\Delta H = \int_0^T \Delta C_p(T)dT \tag{1.18}$$

$$\Delta S = \int_0^T \frac{\Delta C_p(T)}{T} dT \tag{1.19}$$

where $\Delta C_p(T)$ is the change in heat capacity at constant pressure at the temperature T (the same equations hold for the absolute enthalpy, entropy and heat capacity, substituting H, S and C_p where appropriate).

The equilibrium constant for a reaction is a measure of how near to completion the reaction will go. For example, in the reaction where a ligand molecule (L) is bound by a protein (N) according to the following equation:

$$N + L \rightleftharpoons NL \tag{1.20}$$

where NL is the complex thus formed, the equilibrium constant (otherwise known as the binding constant, the affinity constant or the association constant) is given by:

$$K = \frac{[NL]}{[N][L]} \tag{1.21}$$

where K is the equilibrium constant, [NL], [N] and [L] are the equilibrium concentrations of the complex, the free protein and the free ligand respectively (in fact the activities of the species involved should be employed, however the activity coefficients are extremely difficult to calculate and in any case tend to be close to one). The inverse of this is the dissociation constant, and is simply the equilibrium constant for the back reaction.

The equilibrium constant for a reaction is related to the free energy change in the following way:

$$\Delta G^{o} = -RT \ln K \tag{1.22}$$

where R is the gas constant. Combining equations 1.21 and 1.13 gives:

$$\ln K = -\frac{\Delta H^{o}}{RT} + \frac{\Delta S^{o}}{R} \tag{1.23}$$

Differentiation with respect to inverse temperature allows the van't Hoff equation to be obtained:

$$\Delta H_{vH} = \frac{-Rd(\ln K)}{d\,1/T} \tag{1.24}$$

If it is assumed that ΔH^O and ΔS^O do not vary with temperature (in other words assuming that ΔC_p is zero, which is often a fair assumption when considering a fairly narrow temperature range) then a plot of lnK versus inverse T gives a straight line with slope equal to $-\Delta H_{vH}/R$.

While the calorimeter measures only enthalpy changes it is a simple matter to obtain the entropy change if it is recalled that for any process which has reached equilibrium ΔG is equal to zero, i.e.

$$\Delta S = \frac{\Delta H}{T} \tag{1.25}$$

1.6 Techniques to be used

1.6.1 Microcalorimetry

The primary means of investigation to be employed in this thesis are the microcalorimetric techniques differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC). A full discussion of these is postponed until Chapter 2.

1.6.2 Protein intrinsic fluorescence

When light of an appropriate wavelength is shone onto a substance it can promote an electron from the highest occupied energy level of the ground state to a higher energy level, an excited state. If the substance relaxes by emitting a photon this is known as luminescence (figure 1.2). There are two types of luminescence; the term phosphorescence is used to describe the delayed emission when the excited state is of a different multiplicity to the ground state (generally ground states are singlet states, so the excited state may be a triplet state for example). This can take up to a matter of seconds. However if both ground and excited state are of the same multiplicity then the photon will be emitted in a matter of nanoseconds, this process is termed fluorescence. (For useful discussion of all aspects of fluorescence see Lakowicz, 1983 and Campbell & Dwek, 1984)

The side chains of the amino acids tryptophan (Trp or W), tyrosine (Tyr or Y) and, to a much smaller extent, phenylalanine (Phe or F) are fluorescent, so any protein containing these residues is also fluorescent. It is difficult to resolve the fluorescence of a protein into portions resulting from each of these residues and so it is common practice when studying protein fluorescence to use an excitation wavelength which minimises the fluorescence from Tyr and Phe residues. This is accomplished by exciting at 295 nm where Trp residues absorb strongly while Tyr and Phe residues are only weakly absorbing. Trp is chosen due to its high quantum yield, and its sensitivity to environmental effects.

One of the most frequently employed aspects of the intrinsic fluorescence phenomenon is the sensitivity of Trp fluorescence to the polarity of the solvent. The

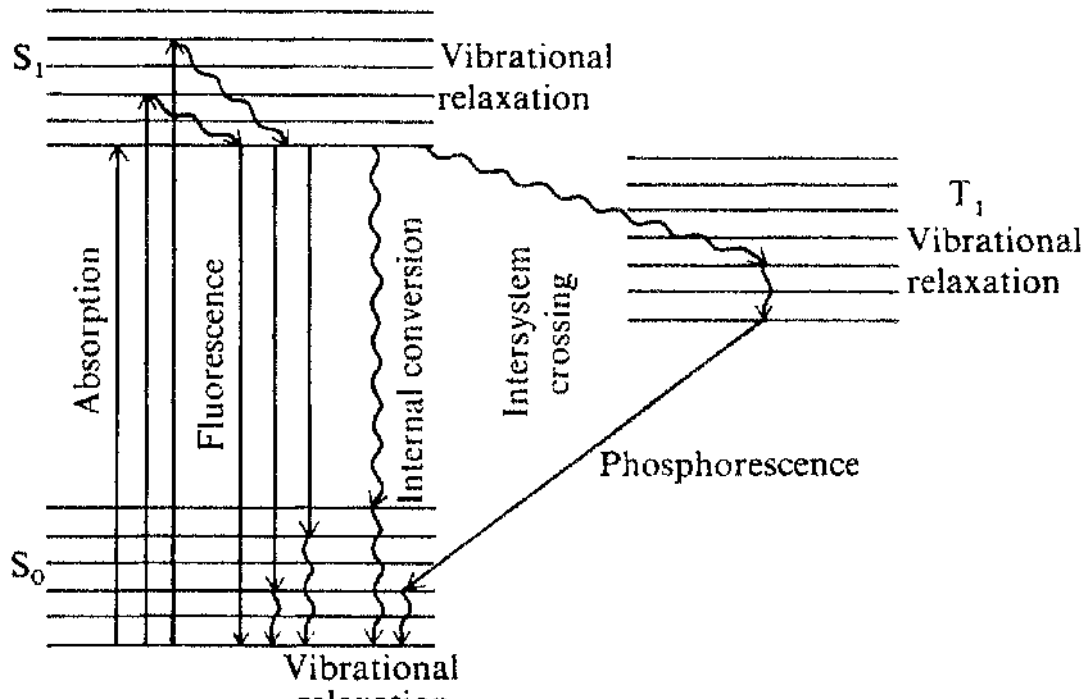

Figure 1.2 Jablonski diagram showing excitation and possible routes of deexcitation.

wavelength of maximal intensity (λ_{max}) of free Trp in solution depends on the nature of the solvent and varies from 297 nm in cyclohexane to 347 nm in water (Sun & Song, 1977). In general the Trp residues of a protein are buried in the polar interior of the molecule, so in practice this means that the observed λ_{max} will reflect the average extent to which the Trp residues are exposed to the water solvent. Any perturbation of the structure which results in greater penetration of water into the hydrophobic core will result in a red shift (i.e shift the fluorescence to a higher wavelength), while those which tend to exclude water from the core will induce a blue shift.

Not all relaxation processes result in the emission of a photon, those that don't are termed radiationless decay. The quantum yield (Φ) is the ratio of the number of photons emitted to the number absorbed (Campbell & Dwek, 1984), and can be calculated by considering the rates of the possible modes of relaxation in the following way:

$$\Phi = \frac{k_f}{\left(k_f + \sum_i k_i\right)} \tag{1.26}$$

where k_f is the rate of fluorescence and k_i is the rate of decay by the i'th mode of radiationless decay. This can also be stated in terms of the fluorescence lifetimes:

$$\Phi = \frac{\tau}{\tau_f} \tag{1.27}$$

where τ_f is the fluorescence lifetime in the absence of competing non-radiative decay processes and τ is the lifetime in their presence. τ_f and τ are given by the following equations:

$$\tau = \frac{1}{\left(k_f + \sum_i k_i\right)} \tag{1.28}$$

$$\tau_f = \frac{1}{k_f} \tag{1.29}$$

The measured fluorescence intensity (I) depends on the quantum yield in the following way:

$$I = P^* \Phi \tag{1.30}$$

where P^* is the population of the excited state. One cause of radiationless decay which can be used to give additional useful information is collisional quenching.

1.6.3 Protein fluorescence quenching

Static quenching, which results from complex formation between the fluorophore and some suitable quenching agent, is of less interest and utility than dynamic or collisional quenching. Static quenching can occur alongside dynamic quenching, causing analysis of the quenching process to become complicated.

Dynamic or collisional quenching, as the name suggests, requires that the quencher and the fluorophore come into contact. When the collision occurs the excited state is relaxed non-radiatively to the ground state. Since the process requires that the quencher encounters the fluorophore in its excited state it depends on the length of time

the molecule remains in that state in the absence of quencher (the fluorescence lifetime; t_f), and on how long it takes for the quencher to diffuse to its site of action. The rate of collisions with quencher is often assumed to be pseudo first order with a rate constant k[Q] ([Q] is the concentration of the quencher). Recall equation 1.26

$$\Phi = \frac{k_f}{\left(k_f + \sum_i k_i\right)} \tag{1.26}$$

Separating the rate of depopulation by collisional quenching from the other radiationless decay processes gives:

$$\Phi_Q = \frac{k_f}{k_f + \sum_i k_i + k[Q]} \tag{1.31}$$

Dividing equation 1.31 into equation 1.26 gives us:

$$\frac{\Phi}{\Phi_Q} = \frac{k_f + \sum_i k_i + k[Q]}{k_f + \sum_i k_i} \tag{1.32}$$

This can be simplified, using equations 1.28 and 1.30, to the following:

$$\frac{I_o}{I} = 1 + k[Q]\tau \tag{1.33}$$

where I_o and I are fluorescence intensities in the absence and presence of quencher respectively. Making the substitution of K for $k\tau$ gives the Stern-Volmer equation:

$$\frac{I_o}{I} = 1 + K[Q] \qquad (1.34)$$

where K is the Stern-Volmer quenching constant (not to be confused with the equilibrium constant). K can be calculated by plotting a graph of I_o/I versus [Q] (a Stern-Volmer plot) which yields a straight line with slope equal to K.

In proteins the fluorophores are generally Trp residues which are buried in the molecular interior this requires that the structure permits passage of the quenching molecule to the site of action. The flexibility of the protein's native structure is such that even relatively large quenching molecules, such as succinimide, (Badley, 1975; Eftink & Ghiron, 1981; Eftink & Ghiron, 1984) are able to permeate into the interior.

In proteins where there is more than one Trp residue the rate of quenching of each of these will depend on their exposure. A modification of the Stern-Volmer equation (Lehrer, 1971), in which it is assumed that there exist two populations of Trp residues, a and b, one of which (a say) is accessible to the quencher and is quenched at a certain rate K_a, the other (b) which is inaccessible and hence is not quenched at all. Under these circumstances the total fluorescence intensity in the absence of quencher (I_o) is given by:

$$I_o = I_{o,a} + I_{o,b} \qquad (1.35)$$

where $I_{o,a}$ and $I_{o,b}$ are intensities deriving from population a and b respectively. Combining equation 1.35 with the Stern-Volmer equation (equation 1.34) gives us:

$$I = \left(\frac{I_{o,a}}{1 + K_a[Q]} \right) + I_{o,b} \tag{1.36}$$

where I is the intensity in the presence of quencher at a concentration of [Q]. Subtraction of equation 1.36 from equation 1.35 gives:

$$\Delta I = I_o - I = I_{o,a} \left(\frac{K_a[Q]}{1 + K_a[Q]} \right) \tag{1.37}$$

Inversion of 1.37 followed by division into 1.35 gives:

$$\frac{I_o}{\Delta I} = \frac{1}{f_a K_a[Q]} + \frac{1}{f_a} \tag{1.38}$$

where f_a is the fraction of the fluorescence intensity deriving from the accessible Trp residues:

$$f_a = \frac{I_{o,a}}{I_{o,a} + I_{o,b}} \tag{1.39}$$

This modified form of the Stern-Volmer equation (equation 1.38) is limited in its usefulness to situations where one population of Trp's is completely inaccessible. Where there exists more than one population of Trp's, each with different non-zero quenching constants the result is a Stern-Volmer plot which curves downwards (Teale & Badley, 1970; Lehrer, 1971). Here at low quencher concentration the initial slope reflects the quenching constant of the more readily accessible Trp's while at higher concentrations the slope reflects the quenching constant of the more buried Trp's (Eftink & Ghiron,

1976a). This assumption allows fluorescence quenching data for multi-Trp containing proteins, such as α-lactalbuminto be usefully analysed.

1.7 Aims of this thesis

The aims of this thesis are to investigate the thermodynamic basis of protein interactions (for which the following systems will be employed: the bacteriophage MS2 viral coat protein and specific RNA sequence, the gene five product of filamentous bacteriophage Pf1 and the Arg repressor of *Bacillus subtilis*, AhrC) and stability (for which the proteins bovine α-lactalbumin and hen egg white lysozyme will be used). Background details of each of these systems will be given in the relevant chapters.

The calorimetric technique of ITC is to be used to measure the enthalpies associated with the binding of Arg to AhrC and a specific RNA 19mer to MS2 viral coat protein. DSC will be used to measure the stability of the Pf1 nucleoprotein complex with and without the flexible C-terminal domain, to measure the effect of solvent on the stability of HEWL and to measure the effect of the C6-C120 disulfide bridge in BLA.

Measurements of protein intrinsic fluorescence and fluorescence quenching will enable the dynamics of 3SS-BLA, WT-BLA and apo-BLA to be investigated, and related to the observed thermodynamics.

"Science is for those who learn; poetry is for those who know."

Joseph Roux.

Chapter two

Microcalorimetry

2.1 Introduction

Thanks to modern electronics it is now possible to build ultra sensitive microcalorimeters, capable of measuring the very small heat effects associated with processes of biochemical interest, such as protein denaturation and ligand binding. In fact these machines are now sensitive enough to measure accurately effective temperature changes of the order of a millionth of a degree. There are two main types of calorimeter, differential scanning calorimeters (Privalov, 1974; Sturtevant, 1974; Privalov & Potekhin, 1986; Sturtevant, 1987; Chowdhry & Cole, 1989; Cooper & Johnson 1992) and isothermal reaction calorimeters (Sturtevant, 1974; Cooper & Johnson 1992). Microcal Inc. make examples of both types and these are the machines which have been used for all calorimetric experiments in this thesis and therefore the discussions in the next few sections are based on them.

2.2 Differential scanning calorimetry

In differential scanning calorimetry (DSC) processes which are initiated by raising the temperature are investigated. These are usually endothermic processes, though non-equilibrium, kinetically controlled, exothermic reactions may also be observed. These however rarely, if ever, give useful information and often obscure the endotherm being collected. In particular DSC is useful for the study of the thermal denaturation of proteins. The DSC instrument measures the difference in heat uptake between the sample and the solvent as the temperature is increased.

2.2.1 The DSC instrument

The MC-2D DSC instrument produced by Microcal Inc. was used exclusively (figure 2.1). This machine consists of two tantalum cells which are, as close as manufacturing tolerances allow, identical. They resemble lollipops in shape and are filled via a capillary tube. In the instrument used here the sample and reference cells have volumes of approximately 1.3 ml, and are always totally filled, with a little excess in a reservoir at the top of the capillary loading tube.

The cells are suspended in an adiabatic chamber, each is supplied with heaters and thermopiles to measure differences in temperature between the sample and reference cells and between the cells and the adiabatic jacket. The jacket is also supplied with heaters. The whole of this construction is thermally insulated from the environment. The adiabatic chamber may be evacuated, or, as in the case of the instrument actually used,

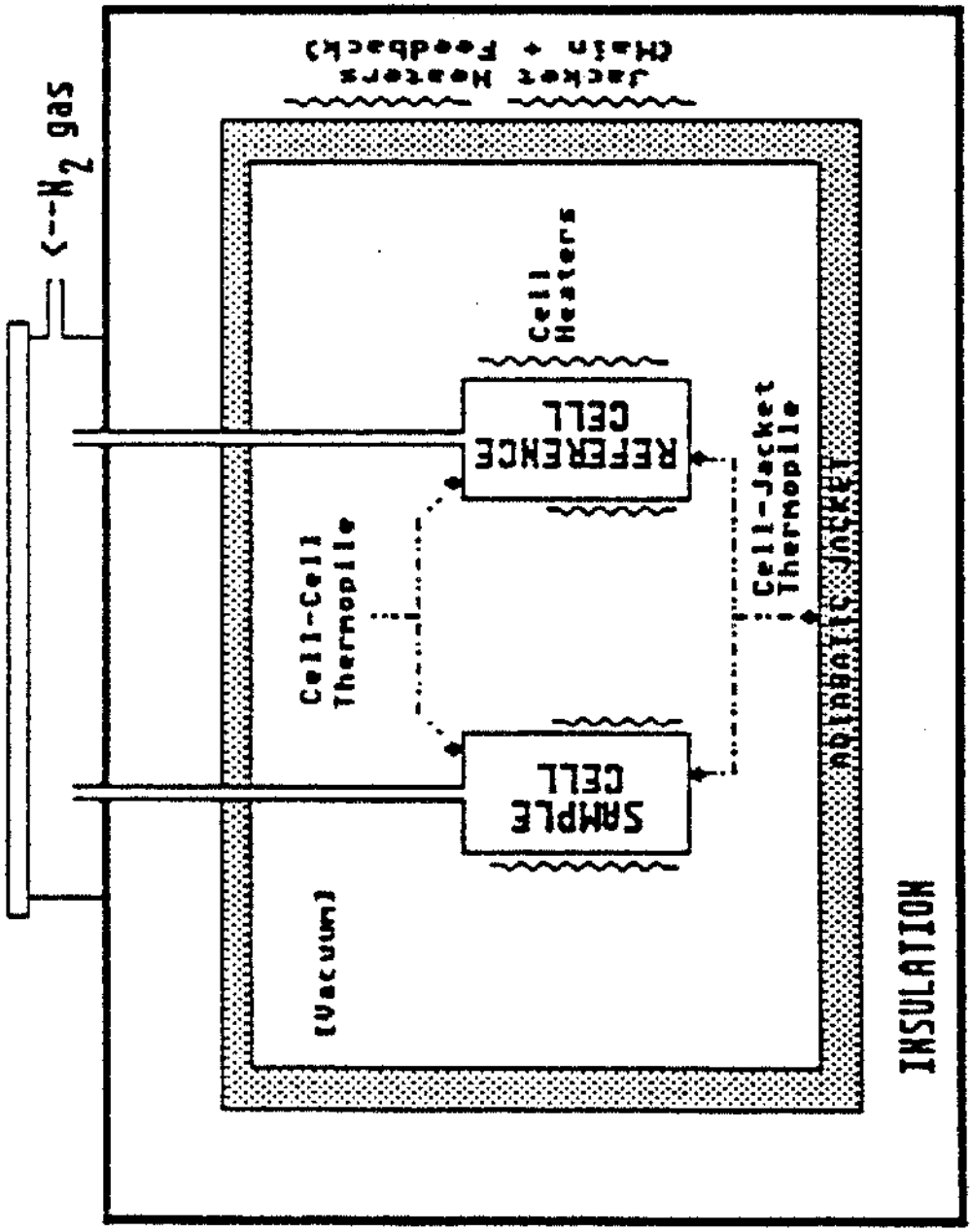

Figure 2.1 The Microcal MC-2D differential scanning calorimeter.

this chamber may be filled with dry nitrogen gas at around 1atm pressure. During the experiment the cells are maintained under a nitrogen pressure of approximately 2atm, in order to inhibit bubble formation. The instrument used is also fitted with an EM electronics N2a DC nanovoltmeter as a pre-amplifier, and a circulating water bath for rapid cooling after termination of the scan.

Cells and jacket each have two sets of heaters, during an experiment power is supplied to the main heaters in order to raise the temperature at the pre-set rate. An active feedback system monitors the difference in temperature between the sample and reference cells and between the cells and the jacket, and heat is supplied where necessary, via the feedback heaters, in order to ensure that their temperatures are maintained equal. The difference in the amount of energy supplied to the sample and reference cells is recorded as a function of temperature.

The DSC instrument is connected to a PC, both for instrument control and data recording, and origin software for this purpose is available from Microcal Inc..

2.2.2 The information available through DSC

A large amount of information is available through DSC experiments. All heat transfers either into or out of the sample or reference cells are recorded, no matter what their source and these may serve to obscure the signal resulting from the thermal transition in the protein under study. One particularly common problem of this kind arises when the protein undergoes exothermic protein-protein aggregation concomitant

with the unfolding transition. This tends to artificially sharpen the shape of the transition, causing the high temperature side of the melting curve to drop off more rapidly than expected. This can be seen in the thermograms of MS2 in later, results sections of the thesis (e.g. Section 7.1.2, figure 7.4)

The most mundane, but in some ways most important, piece of information made available through DSC is the temperature at which a particular protein unfolds under the conditions used in the experiment (figure 2.2). For a two state unfolding process in which the equilibrium is between just two possible states, native and denatured:

$$N \rightleftharpoons D \qquad (2.1)$$

the so called transition mid-point temperature (T_m) of the unfolding curve is the temperature at which exactly half of the molecules will be in the folded, N state and the other half unfolded, D state (or more realistically T_m is the temperature at which each molecule spends exactly half of its time in each state). This is the temperature which halves the area under the curve (often the temperature of maximum heat capacity). The T_m is a useful point of reference when discussing the thermal stability of the protein.

Changing solvent conditions can have a marked effect on the thermal stability of a protein, and therefore on the T_m. It is, therefore, a simple task to investigate which solvent conditions, and additives are either beneficial or detrimental to protein stability.

In the DSC experiment we measure the difference in heat uptake between sample and reference. In other words we are measuring the difference in their heat capacities under constant pressure (C_p) as a function of temperature. Recalling equation 1.18:

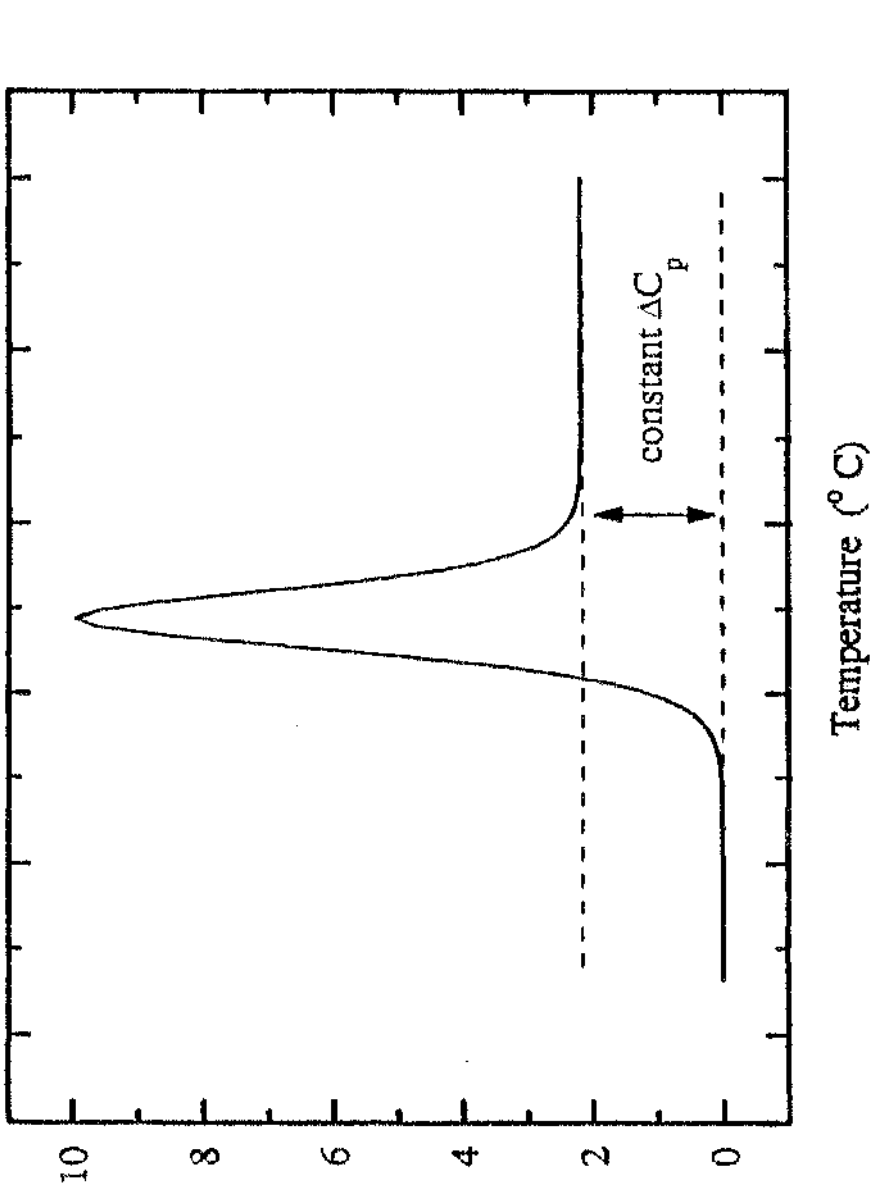

Figure 2.2 Idealised DSC trace (simulated data).

$$\Delta H = \int_0^T \Delta C_p(T)dT \tag{1.18}$$

allows us to see that the calorimetric enthalpy (Δh_{cal}) of the transition can be obtained by integration of the DSC trace with respect to temperature. It is usual to express this as enthalpy per mole of protein (ΔH_{cal}):

$$\Delta H_{cal} = \Delta h_{cal} M \tag{2.2}$$

where M is the molecular weight of the protein used. This estimate does not depend on any models chosen to fit the experimental data, it is an absolute measure of the extra energy supplied to the sample cell during the transition. However, care must be exercised in the choice of the baseline used for integration.

It is also possible to obtain the van't Hoff enthalpy for the transition by using the calorimetric signal as a probe to measure the extent of reaction and hence the equilibrium constant as a function of temperature and employing the van't Hoff equation:

$$\Delta H_{vH} = -\frac{Rd(\ln K)}{d\left(\frac{1}{T}\right)} \tag{1.24}$$

The van't Hoff enthalpy is independent of the calorimetric enthalpy, and nor does it depend on the concentration of protein used in the cell, but rather is calculated per mole. It does however depend on an assumed model (usually 2-state) for the unfolding process. This is a very helpful aspect of the calorimetric technique, because not only

does this supply two independent measures of the enthalpy (hopefully increasing accuracy) it also allows us to calculate the size of the cooperative unit.

In a protein unfolding transition the molecules may act in concert as dimers or trimers etc. when this does occur the cooperative unit is two or three respectively. Obviously monomers unfold with a cooperative unit of at least one (a cooperative unit of less than one would tend to indicate either the existence of more than one domain in the protein with different T_m's or that the unfolding is non-cooperative). Measurements of the van't Hoff enthalpy based on the calorimetric signal are per mole of cooperative unit, while the calorimetric enthalpy is generally calculated per mole of monomer, the cooperative unit is, therefore, simply given by:

$$n = \frac{\Delta H_{vH}}{\Delta H_{cal}} \tag{2.3}$$

where n is the number of monomers in the cooperative unit.

Since the van't Hoff enthalpy derives from the shape of the transition the cooperativity of the transition can be qualitatively observed from the shape of the C_p versus T curve. A sharp transition indicates a highly cooperative process, while broad transitions are indicative of less inter-molecular cooperativity.

One final, crucial piece of information can, in principle, be gleaned directly from the DSC data. As can be seen from figure 2.2 the unfolding transition is accompanied by a large, positive change in heat capacity, ΔC_p. Because of this ΔC_p the enthalpy, entropy and therefore free energy of the transition are temperature dependent. It is necessary to have an accurate estimate of ΔC_p in order that the thermodynamic parameters can be

calculated at suitable reference temperatures. The variation in ΔH and ΔS with temperature are given by the following:

$$\Delta H(T_1) = \Delta H(T_0) + \Delta C_p(T_1 - T_0) \tag{2.4}$$

$$\Delta S(T_1) = \Delta S(T_0) + \Delta C_p \ln\left(\frac{T_1}{T_0}\right) \tag{2.5}$$

In the idealised case presented in figure 2.2 it would be a simple matter to measure ΔC_p from the difference between the pre- and post-transition baselines. Unfortunately "the truth is rarely pure and never simple" (Oscar Wilde). The typical DSC trace (see figure 2.3 for an example) is different in appearance, pre- and post-transition baselines are almost never co-linear, presenting a problem for the estimation of ΔC_p, which is acutely sensitive to the choice of baseline. In practice much greater accuracy is achieved by measuring the enthalpy of the transition over a range of conditions which cause a shift in T_m and plotting the graph of ΔH versus T, the slope of which is equal to ΔC_p.

Using the information from DSC allows calculation of the free energy changes which direct the process of unfolding. An equation which relates the standard free energy change for the reaction to the experimentally derived parameters of ΔH_m, T_m and ΔC_p can be derived as follows. First recall equation 1.18:

$$\Delta H = \int_0^T \Delta C_p(T)dT \tag{1.18}$$

which can also be written:

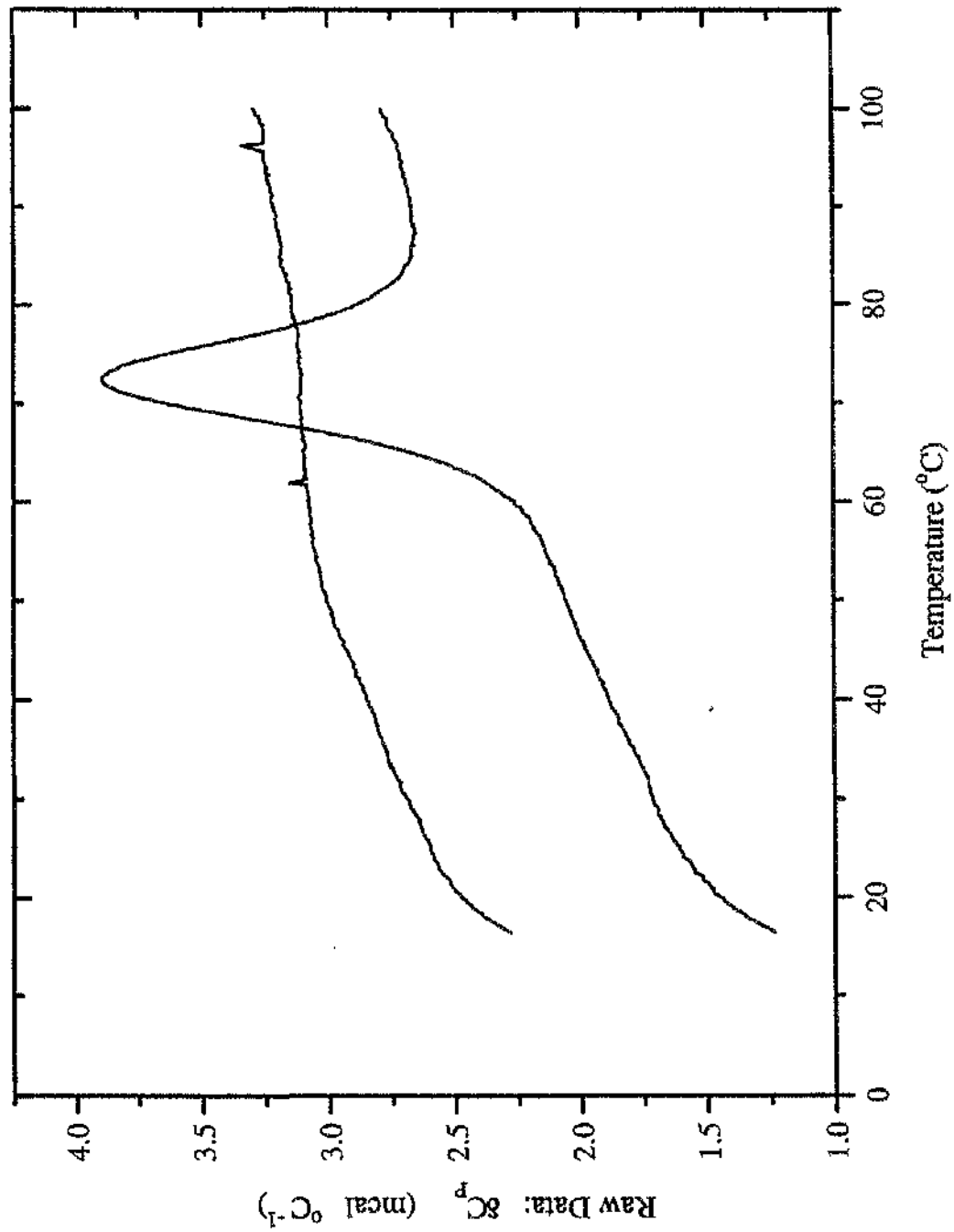

Figure 2.3 Typical DSC trace, showing both the normalised data, prior to baseline subtraction and the buffer base line (DSC obtained with BLA in Tris buffer pH 7.4, 100mM NaCl, 100mM $CaCl_2$.

$$\Delta H_m = \Delta H_{298} + \int_{298}^{T} \Delta C_p(T_m) dT \tag{2.6}$$

Carrying through the integration, assuming that ΔC_p does not vary with temperature and then rearranging gives:

$$\Delta H_{298} = \Delta H_m - \Delta C_p(T_m - 298) \tag{2.7}$$

Notice that rearrangement and substitution of appropriate temperatures gives equation 2.4. Also recall equation 1.19:

$$\Delta S_m = \int_0^{T_m} \frac{\Delta C_p}{T} dT \tag{1.19}$$

which can similarly be written:

$$\Delta S_{298} = \Delta S_m - \Delta C_p \ln\left(\frac{T_m}{298}\right) \tag{2.8}$$

Again, notice the similarity to equation 2.5. However the free energy at T_m, ΔG_m, is equal to zero and therefore rearrangement of the Gibbs-Helmholtz equation (equation 1.13) gives:

$$\Delta S_m = \frac{\Delta H_m}{T_m} \tag{2.9}$$

Substitution of equation 2.9 into equation 2.8 gives:

$$\Delta S_{298} = \frac{\Delta H_m}{T_m} - \Delta C_p \ln\left(\frac{T_m}{298}\right) \tag{2.10}$$

Substitution of equations 2.10 and 2.7 into equation 1.13 followed by rearrangement gives:

$$\Delta G_{298} = \Delta H_m\left(1 - \frac{298}{T_m}\right) + \Delta C_p\left(298 - T_m - 298\ln\frac{298}{T_m}\right) \tag{2.11}$$

2.2.3 Analysis of DSC data

All analysis of DSC data is carried out making use of the Origin software package, with DSC add-on, from Microcal Inc.

DSC data are routinely recorded by the DA2 software package as differential energy input as a function of time and it is necessary to normalize the data in order to produce the more normal differential heat capacity (or excess heat capacity, c_{ex}) as a function of temperature (figure 2.3) by dividing by the mean scan rate. It should be noted that the scan rate set is nominal only and the actual mean scan rate achieved is recorded and used for this purpose.

Due to variations in the total heat capacity and heat losses in the machine over the temperature ranges commonly employed the actual scan rate varies slightly. This manifests itself as a curvature of the DSC trace. Also instrumental peculiarities such as mismatch of the cells leads to artefacts on the thermogram. These are simply removed by subtracting a buffer base line obtained under precisely the same conditions (figure 2.3) by filling both sample and reference cell with the buffer being used. For greatest accuracy it is necessary to obtain a base line for each new buffer system used no matter how small the differences may be, indeed it is best to use buffer which has been equilibrated (usually by dialysis) with the buffer in the sample solution for all buffer base lines.

It is usual to normalize the data with respect to the total amount of sample in the cell in order to obtain differential excess heat capacity per mole (figure 2.4). This can be done by the Origin software, which has the actual cell volumes set as a default and gives a prompt to the operator for the sample concentration.

The Origin software DSC add-on contains a number of models which may be selected, where appropriate, for fitting to the experimental data. The package is also able to perform deconvolution of multiple, overlapping transitions. The use of these models demands that a baseline be provided for integration. This is done by transforming the essentially flat pre-transition portion of the trace into the post-transition portion according to some simple mathematical model such as assuming a cubic relationship (figure 2.5).

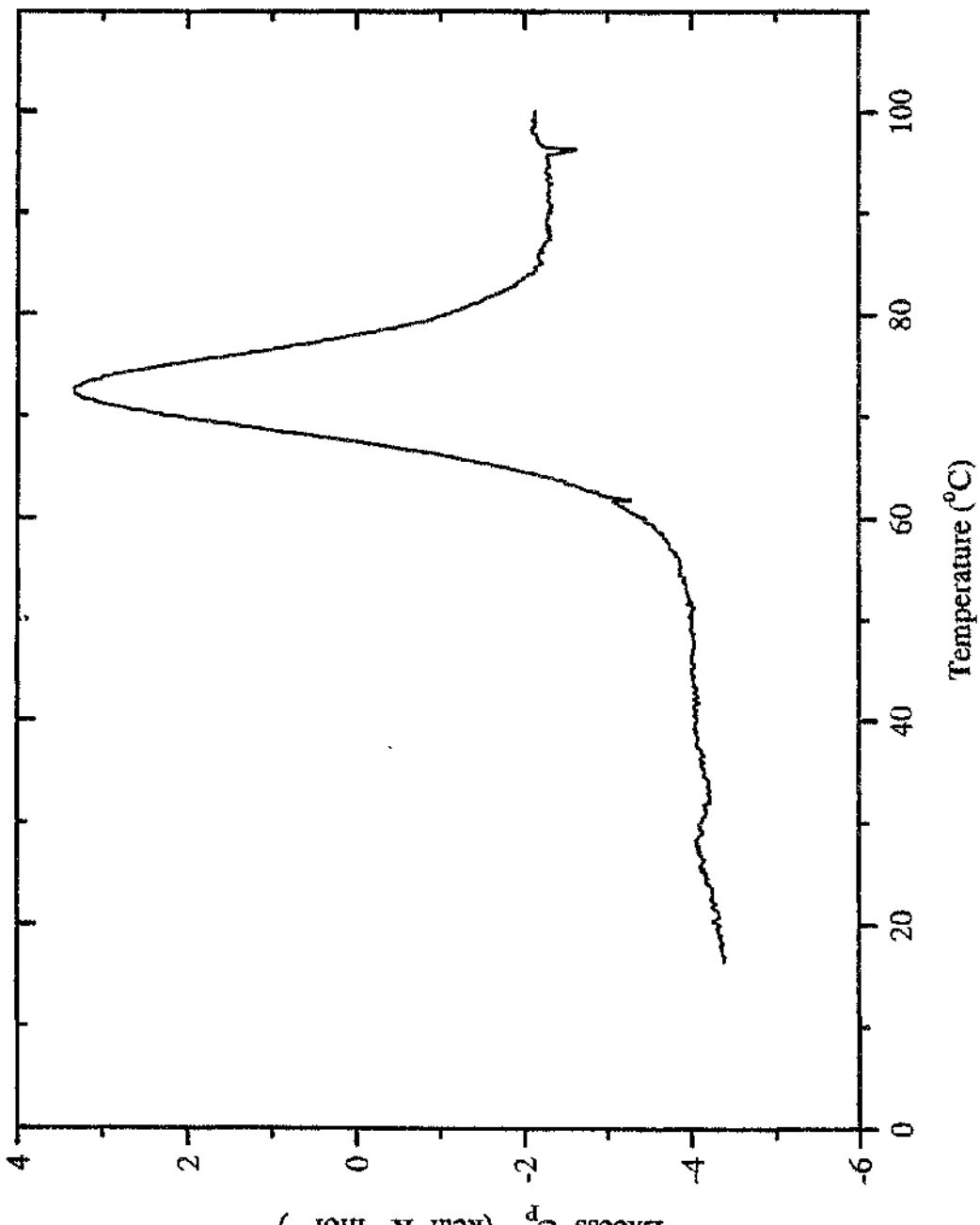

Figure 2.4 Normalised DSC trace with buffer base line subtracted (data obtained as for Figure 2.3).

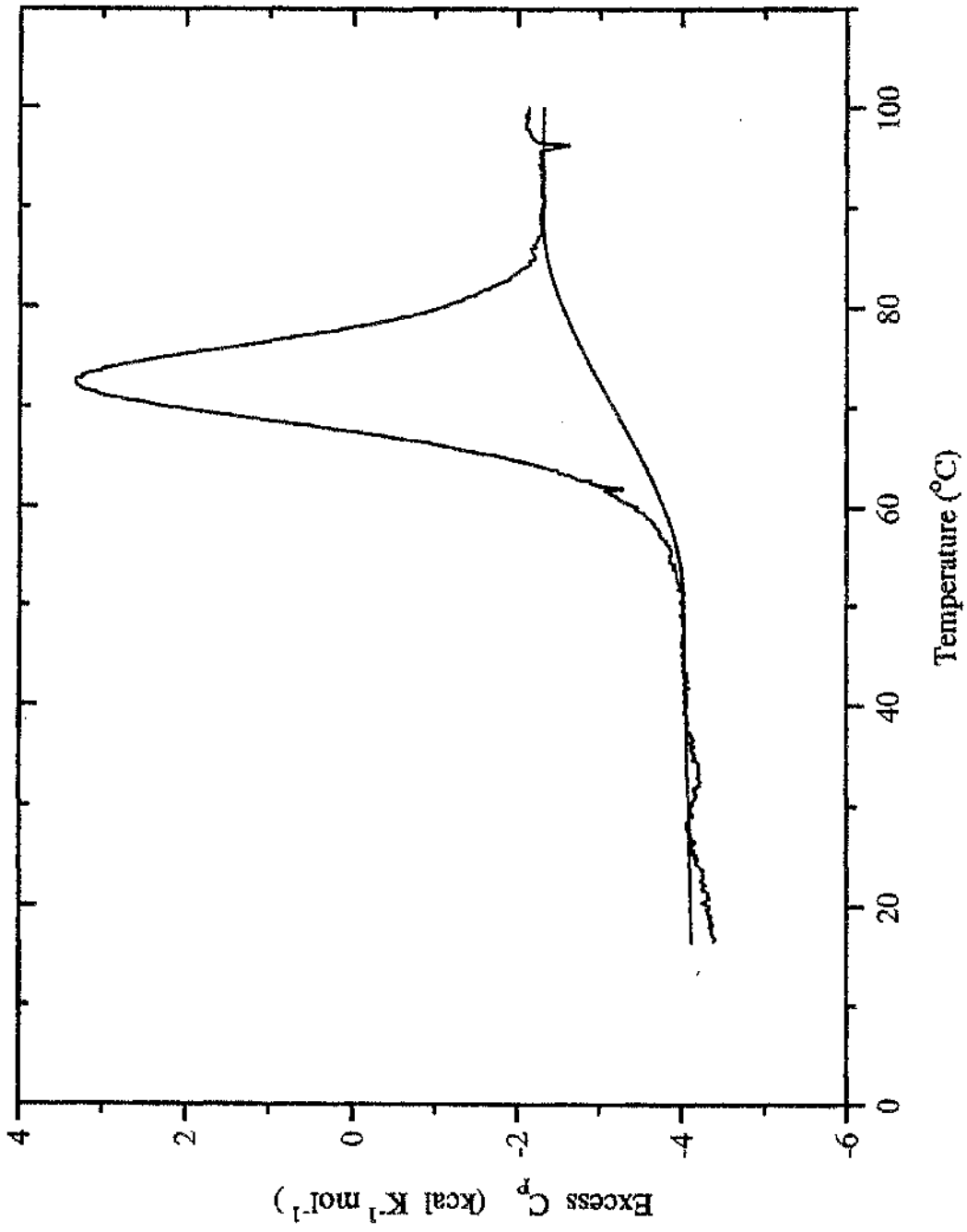

Figure 2.5 DSC trace as before with a cubic baseline drawn for integration.

2.2.4 Models used in DSC analysis

Analysis of the DSC data is made by fitting an appropriate model to the experimentally derived thermogram. For this the software package Origin, from Microcal Inc., is used. All models used to fit DSC data are based upon the van't Hoff equation (cf equation 1.24):

$$\left(\frac{\partial \ln K}{\partial T}\right)_p = \frac{\Delta H_{vH}}{RT^2} \tag{2.12}$$

In the general case it is assumed that the protein consits of one or more structural domains (A, B, C, etc), each of which can undergo a transition between folded and unfolded forms independently of each other (Unfolded forms designated with a prime e.g A' designates the unfolded form of the A domain). Equilibrium constants are expressed in terms of fractions e.g.

$$K_A = \frac{f_{A'}}{f_A} \tag{2.13}$$

where K_A is the equilibrium constant for the unfolding of the A domain, $f_{A'}$ and f_A are the fractions of unfolded and folded A respectively. Remembering that $f_{A'}$ and f_A are simply related as follows:

$$f_A = 1 - f_{A'} \tag{2.14}$$

The total molar enthalpy for the system is given by:

$$H = H_N + f_{A'}\Delta H_A + f_{B'}\Delta H_B + \ldots \tag{2.15}$$

where H_N is the total enthalpy of the native state, to which all measurements of ΔH are relative. The total molar heat capacity is the temperature derivative of equation (2.15):

$$C_p = C_{pN} + \left(f_{A'}\Delta C_{pA} + \Delta H_A \frac{\partial f_{A'}}{\partial T} \right) \tag{2.16}$$

where C_{pN} is the molar heat capacity of the folded state. Substituting equilibrium constants in place of fractional concentrations in 2.16 and then carrying through the differentiation gives:

$$\frac{\partial f_{A'}}{\partial T} = \frac{K_A}{(1+K_A)^2}\frac{\partial K_A}{\partial T} \tag{2.17}$$

Notice that the final term is familiar from the van't Hoff equation 2.12. Substituting equations 2.12 and 2.17 into equation 2.16 gives the following:

$$C_p = C_{pN} + \left(\frac{K_A \Delta C_{pA}}{1+K_A} + \frac{K_A \Delta H_{vH,A} \Delta H_A}{(1+K_A)^2 RT^2} \right) + \ldots \tag{2.18}$$

This equation is general and forms the basis for all subsequent models, which incorporate various assumptions. If we assume that the transition occurs with a permanent change in heat capacity (ΔC_p) and that each transition is a two state, all or none type transition then van't Hoff and calorimetric enthalpies are equal. If we also assume that C_{pN} is a linear function of temperature then equation 2.18 gives:

$$C_p(T) = B_0 + B_1 T + \left(\frac{K_A(T)\Delta C_{pA}}{1+K_A(T)} + \frac{K_A(T)\Delta H_A(T)^2}{\left(1+K_A(T)\right)^2 RT^2} \right) + \ldots \quad (2.19)$$

where $C_p(T)$, $K_A(T)$ and $\Delta H_A(T)$ are the heat capacity, equilibrium constant and enthalpy change at temperature T respectively. This can be expressed in terms of the T_m of the transition as follows:

$$\Delta H_A(T) = \Delta H_{m,A} + \Delta C_{pA}\left(T - T_{m,A}\right) \quad (2.20)$$

Integrating the van't Hoff equation (2.12) from T_{mA} to T gives:

$$K_A(T) = \exp\left\{ \frac{-\Delta H_{m,A}}{RT}\left(1 - \frac{T}{T_{m,A}}\right) - \frac{\Delta C_{pA}}{RT}\left(T - T_{m,A} - T\ln\frac{T}{T_{m,A}}\right) \right\} \quad (2.21)$$

Substituting equations 2.20 and 2.21 into equation 2.19 gives the model to which DSC data can be fitted. The operator gives initial guesses and the model is fit by non-linear-least-squares. If it is assumed that all changes in heat capacity are zero then these equations reduce to:

$$C_p(T) = \frac{K_A(T)\Delta H_A(T)^2}{\left(1+K_A(T)\right)^2 RT^2} + \ldots \quad (2.22)$$

$$K_A(T) = \exp\left\{ \frac{-\Delta H_{m,A}}{RT}\left(1 - \frac{T}{T_{m,A}}\right) \right\} \quad (2.23)$$

again curve fitting is by non-linear-least-squares fitting. Another level of complication is introduced when subunit dissociation occures simultaneously with unfolding:

$$A_n \rightleftharpoons nA \tag{2.24}$$

where n is the number of subunits. The equilibrium constant for this reaction is given by:

$$K = \frac{[A]^n}{[A_n]} \tag{2.25}$$

If f is the fraction of molecules in the A state and 1-f is the fraction in the A_n state then:

$$f = \frac{[A]}{nC_t} \tag{2.26}$$

$$1 - f = \frac{[A_n]}{C_t} \tag{2.27}$$

where C_t is the total bulk concdentration of n-mer, combining eqiuations 2.25, 2.26 and 2.27 gives the following term for the equilibrium constant:

$$K = \frac{f^n n^n C_t^{n-1}}{1 - f} \tag{2.28}$$

Using equation 2.28 along with the following:

$$\frac{\partial \ln K(T)}{\partial\left(\frac{1}{T}\right)} = \frac{-\Delta H_{vH}(T)}{R} \tag{2.29}$$

$$\Delta H(T) = \Delta H_{m,A} + \Delta C_{pA}\left(T - T_{m,A}\right) \tag{2.30}$$

and integrating from T_{mA} (where f = 1 - f = 0.5) to T gives:

$$K(T) = 5^{n-1} n^n C_t^{n-1} \exp\left\{\frac{-\Delta H_{m,A}}{RT}\left(1 - \frac{T}{T_{m,A}}\right) - \frac{\Delta C_{pA}}{RT}\left(T - T_{m,A} - T\ln\frac{T}{T_{m,A}}\right)\right\} \tag{2.31}$$

using K(T) from equation 2.31 together with equation 2.28 allows calculation of f(T). The excess heat capacity can then be calculated from the following:

$$C_p(T) = B_0 + B_1 T + f(T)\Delta C_p + \frac{\Delta H_A(T)}{RT^2}\left(\frac{1 - f(T)}{1 - n + \frac{n}{f(T)}}\right) \tag{2.32}$$

The above eqautions are also valid for a system which undergoes an increase in the degree of oligomerisation concomitant with unfolding as long as the total bulk concentration is expressed in terms of monomer concentration and n above is replaced with 1/j. Where j is as indicated below:

$$A \rightleftharpoons \frac{1}{j} A_j \tag{2.33}$$

2.3 Isothermal titration calorimetry

In isothermal titration calorimetry (ITC) processes which occur at essentially constant temperature are investigated. In principle any kind of reaction can be investigated, although in terms of biological macromolecules ligand binding is arguably the most interesting. The small heat effects associated with binding phenomena are measured and analysis of the data, as described in some detail later yields the heat of binding and equilibrium constants for the binding reaction. The instrument of choice for biological samples is the Omega titration calorimeter from Microcal Inc. One major advantage which the Omega machine has over other reaction calorimeters (such as the LKB batch flow calorimeters) is the very fast response and equilibration times, meaning whole experiments can be carried out rapidly, typically in less than one hour.

2.3.1 The ITC instrument

The Microcal Inc. Omega ITC instrument (figure 2.6) was used for all titration calorimetry carried out in this thesis. The machine is constructed on the same principle as the MC-2D DSC instrument (section 2.2.1). The calorimeter is made up of two cells, made of Hastelloy C, which are contained within an adiabatic jacket. The jacket must be evacuated at low temperature in order to prevent condensation, but can be left at atmospheric pressure otherwise. The cells are lollipop shaped, with a volume of approximately 1.4 ml and samples are loaded via narrow access tubes. One cell (the reference cell) is filled with purified water (containing sodium azide to prevent growth of microorganisms) which need only be replaced occasionally.

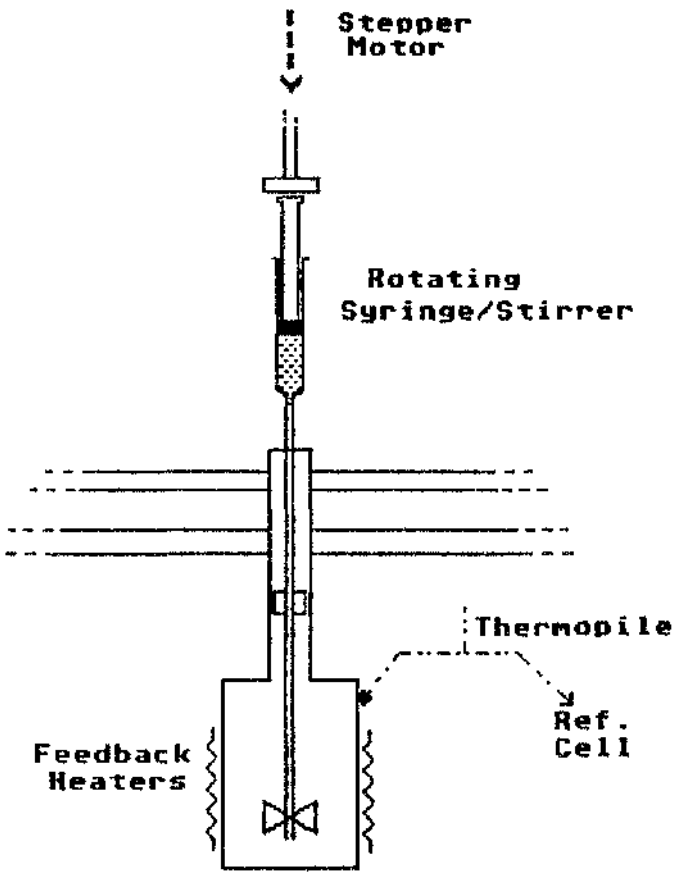

Figure 2.6 The Microcal Omega isothermal titration calorimeter.

The binding reaction takes place in the other cell, the reaction or sample cell. This cell is fitted with a rotating syringe which is used both to deliver one of the reactants in carefully metered aliquots, and to stir the mixture. The plunger of the syringe is actuated by a linear stepper motor which, under the control of a PC, is able to depress the plunger by very precise amounts, ensuring accurate delivery of the reactant.

A feedback system, similar to the one in place in the DSC instrument, monitors the differences in temperature between the reference and reaction cells and between the cells and the jacket and supplies heat where appropriate in order to maintain a fine balance in temperatures. As for the DSC it is the difference in heat supplied to the two cells which is recorded as a function of time.

A PC, running suitable software available from Microcal Inc., controls the action of the syringe stepper motor, and the operator is able to set parameters such as duration of injection, distance by which the plunger is depressed (and therefore the volume of injection) and time between subsequent injections. The PC also records the data ready for analysis using the Microcal Origin package with an ITC add-on installed. The speed of the rotating, stirring syringe is not controlled by the PC and care must be taken to ensure that the same speed is used for the titration and all controls as this influences the heat of mixing. The syringe speed was generally set at 400 RPM.

2.3.2 The information available through ITC

The amount of heat transferred into or out of the system following an injection can be obtained by integration of the heat pulse recorded (figure 2.7). If titration is carried on until saturation the summation of the heat effect associated with each injection gives the total reaction heat, Q_r (after correction with appropriate control experiments, see section 2.3.3). It is possible to obtain the binding enthalpy directly from Q_r by dividing by the protein concentration, [M]:

$$\Delta H^o = \frac{Q_r}{[M]} \tag{2.34}$$

Binding constants can be obtained by using heat production or uptake as a probe for the extent of the reaction:

$$Q_r = \frac{[ML]\Delta H^o}{[M]_t} \tag{2.35}$$

where [ML] is the concentration of the complex and $[M]_t$ is the total protein concentration. For simple one to one binding this can be expressed:

$$Q_r = \frac{K_b[L]\Delta H^o}{1+K_b[L]} \tag{2.36}$$

where K_b is the binding constant and [L] is the free ligand concentration. Rearrangement of equation 2.36 gives:

$$\frac{1}{Q_r} = \frac{1}{\Delta H^o} + \frac{1}{K_b[L]\Delta H^o} \tag{2.37}$$

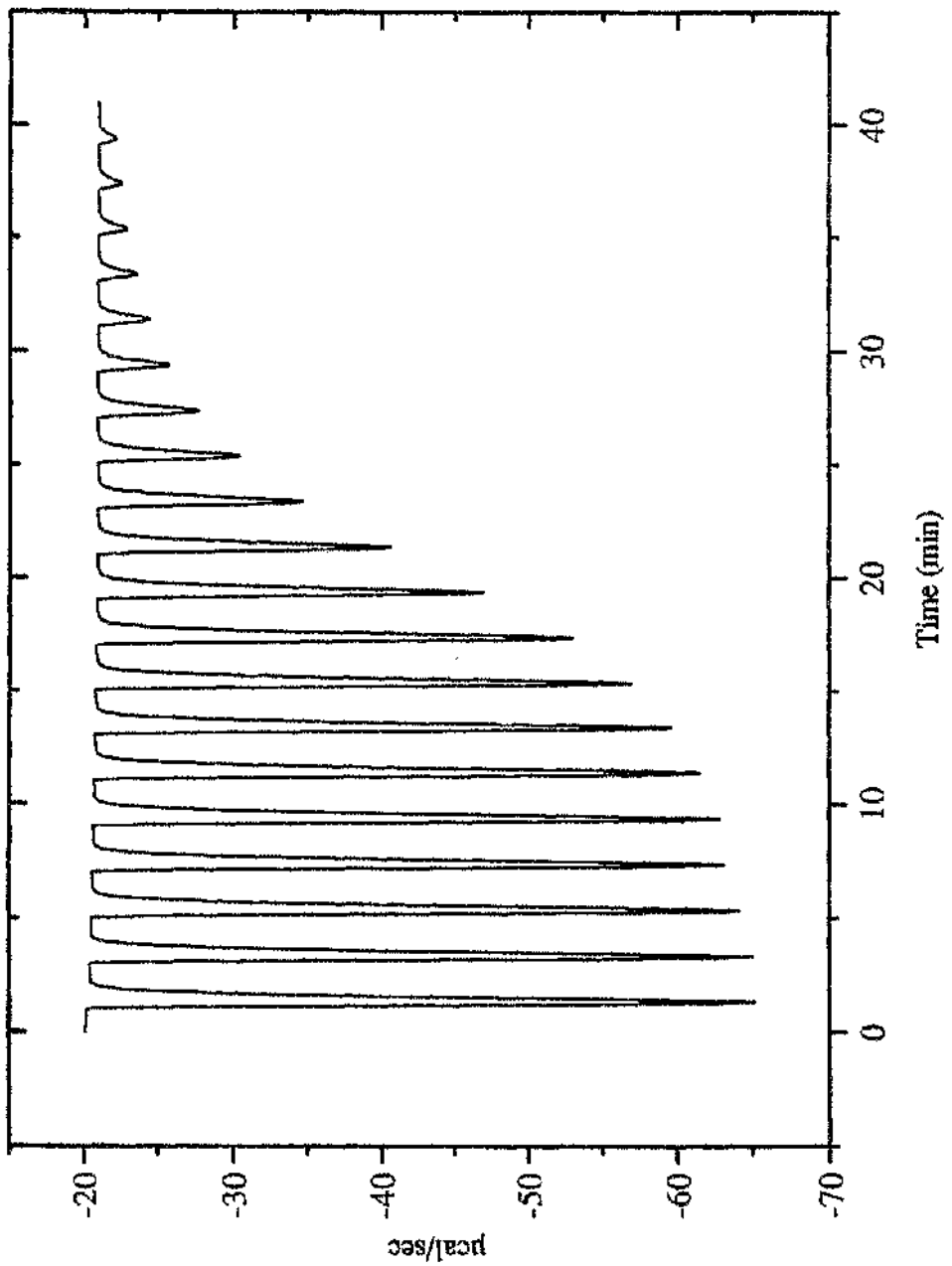

Figure 2.7 Typical heat pulses obtained from a series of ITC injections (simulated data).

Under conditions of weak binding where the approximation that free ligand concentration equals total ligand concentration can be made a double reciprocal plot of $1/Q_r$ versus $1/[L]$ yields a straight line with intercept of inverse ΔH^o and slope equal to $1/\Delta H^o K_b$. However binding is generally too strong for this approximation to hold and free ligand concentration at equilibrium is not comparable to the total ligand concentration. It can, however, be estimated from known quantities in the following way:

$$[L] = [L]_t + \frac{Q_r [M]_t}{\Delta H^\circ} \tag{2.38}$$

where $[L]_t$ is the total ligand concentration. Substituting equation 2.38 into equation 2.37 gives the hyperbolic binding equation for one to one binding. Under conditions where the stoichiometry is not known it is necessary to carry out full least squares fitting of the complete hyperbolic binding equation, which explicitly includes the number of equivalent binding sites, N, to the experimental data in order to obtain accurate values for ΔH^o and K_b:

$$Q_r = -\Delta H^\circ \left(N[M]_t + [L]_t + \frac{1}{K_b} \right) \frac{\left(1 - \left(1 - \frac{4N[M]_t[L]_t}{\left(N[M]_t + [L]_t + \frac{1}{K_b} \right)^2} \right)^{1/2} \right)}{2[M]_t} \tag{2.39}$$

For a system which contains multiple non-interacting binding sites the heat effect will be:

$$Q_r = \sum_{i=1}^{j} \frac{N_i \Delta H_i^{\circ} [L]_t K_i}{1 + K_i [L]_t} \qquad (2.40)$$

where j is the number of distinct binding sites, K_i, ΔH^{O}_i and N_i are the apparent binding constant, enthalpy change and number for each set of sites respectively.

Given K_b and ΔH^{O} it is a simple matter for the standard free energy, ΔG^{O}, and entropy changes, ΔS^{O}, to be calculated from the equations 1.13 and 1.22.

A series of experiments at different temperatures, but under otherwise identical conditions allows calculation of ΔC_p from the temperature dependence of the calorimetrically determined ΔH^{O}. The van't Hoff enthalpy can also be calculated from the temperature dependence of the equilibrium binding constant, K_b. The number of identical binding sites on the protein can, in principle, be obtained from a comparison of the van't Hoff and calorimetric enthalpies in much the same way as it is possible to calculate of the size of the cooperative unit from DSC data (section 2.2.2), although this is rarely done since N is routinely determined by fitting the titration curve.

2.3.3 Control experiments

As for DSC the calorimeter measures all heat transfers, no matter what their source, so the observed heat effect, Q_{obs}, includes the required heat of reaction, Q_r, plus the heat of dilution of both ligand and protein, Q_{dil} and the heat of mixing, Q_{mix}:

$$Q_{obs} = Q_r + Q_{dil} + Q_{mix} \qquad (2.41)$$

Therefore in order to obtain the heat resulting only from the binding reaction it is necessary to carry out a number of control experiments

Expt.	Contents of cell	Contents of cell	Heat effects
1	Protein	Ligand	A+B+C+D
2	Protein	Buffer	C+D
3	Buffer	Ligand	B+D
4	Buffer	Buffer	D

Table 2.1 control experiments necessary for correction of observed heat effects in titration experiments.

where the heat effects are as follows:

A = Heat of Protein-ligand binding reaction, Q_r.

B = Heat of dilution of the ligand (this is usually the largest control since the ligand in the syringe is reasonably concentrated, and is being diluted into a much larger volume upon injection into cell).

C = Heat of dilution of protein ($B + C = Q_{dil}$)

D = Heat of mixing (usually very small)

The required heat of binding is given by:

$$Q_r = A = 1 - 2 - 3 + 4 \qquad (2.42)$$

2.3.4 Analysis of ITC data

All analysis of experimental data was carried out using the Origin software package (including the ITC add-on) from Microcal Inc.. The raw data obtained from the titration experiment is a record of the energy required to maintain the temperatures of the cells and jacket equal as a function of time. For each set of injections (including all of the controls) a baseline is drawn, the peaks are integrated and the controls are subtracted to give corrected heats, which are then plotted as a function of injection number (figure 2.8).

The Origin software package contains a number of mathematical models which can be selected for fitting to the experimental data by the method of least squares. The model then gives values for N, K_b and ΔH^o. For weak binding situations the stoichiometry is generally poorly defined and under these circumstances it is necessary to fix the number of binding sites in order to give reasonable values for the binding constant and the enthalpy change.

When fitting a model for two sets of sites to the data it is not always possible to obtain a unique fit since there are six floating parameters. Again fixing the stoichiometry may help, but care needs to be exercised in choosing what values to set. The situation is complicated further still by the possibility that binding sites may be interactive. The binding of the first ligand may enhance binding of the second, i.e. cooperativity (Cooper et al, 1994), or it may decrease affinity in the second. This situation will not be encountered in this thesis.

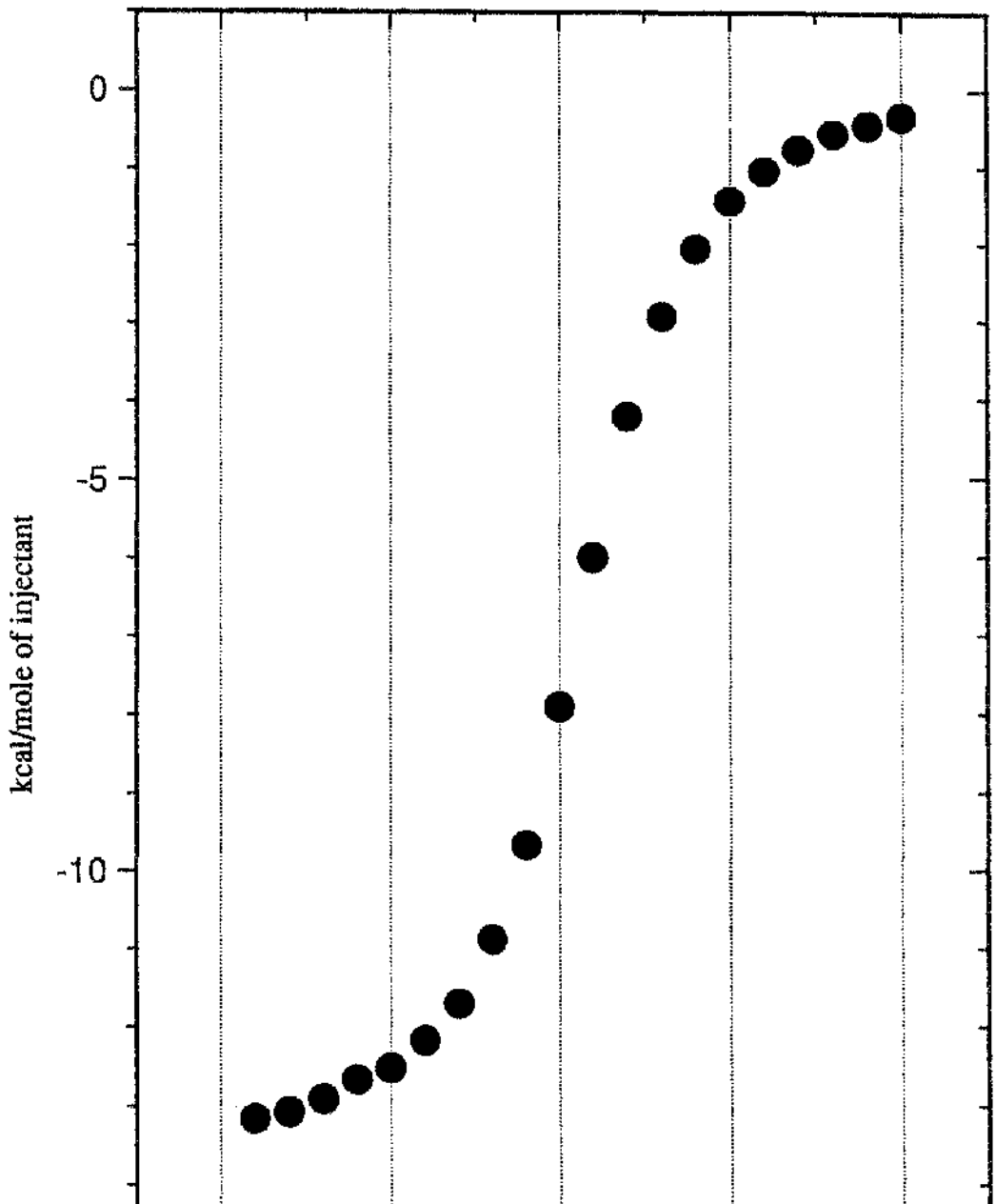

Figure 2.8 Integrated heats from pulses in Figure 2.7.

2.3.5 Models used in ITC analysis

The mathematical models used by the Origin software package for fitting to experimental data are as follows.

When fitting data obtained for a one-to-one binding reaction the binding constant is given by the equation:

$$K_b = \frac{\theta}{(1-\theta)[L]} \tag{2.43}$$

where θ is the fraction of sites occupied by ligand. The free ligand concentration at equilibrium can be related to the total ligand concentration by the equation:

$$[L] = [L]_t - N\theta[M]_t \tag{2.44}$$

Combining equations 2.43 and 2.44 gives the equation:

$$\theta^2 - \theta\left(1 + \frac{[L]_t}{N[M]_t} + \frac{1}{NK_b[M]_t}\right) + \frac{[L]_t}{N[M]_t} = 0 \tag{2.45}$$

The total heat content of the solution, Q, is equal to:

$$Q = N\theta[M]_t \Delta HV_o \tag{2.46}$$

where V_0 is the active cell volume. Solving equation 2.45 for θ and substituting into equation 2.35 gives:

$$Q = \frac{N[M]_t \Delta H V_o}{2}\left(1+\frac{[L]_t}{N[M]_t}+\frac{1}{NK[M]_t}-\sqrt{\left(1+\frac{[L]_t}{N[M]_t}+\frac{1}{NK[M]_t}\right)^2-\frac{4[L]_t}{N[M]_t}}\right) \tag{2.47}$$

This equation can be solved for Q_i given values for N, K_b and ΔH at the end of the ith injection. The calorimeter actually measures the difference in heat between the ith and (i-1)th injection:

$$\Delta Q_i = Q_i - Q_{i-1} \tag{2.48}$$

where ΔQ_i is the difference in heat, Q_i and Q_{i-1} are the heats of the ith and (i-1)th injections respectively. Initial guesses of the values of the fitting parameters (N, K_b and ΔH) are made and these are substituted into equation 2.47 to find values for ΔQ_i for each injection. These calculated ΔQ_i values are compared with the experimental values and statistical methods are employed to find new values of the fitting parameters. A number of iterations are carried out until the values converge.

When fitting data where it is assumed that there are two independent sets of binding sites values of N, K_b and ΔH are obtained for both sets. The following equations are comparable to those used for one set of sites (equations 2.43 to 2.48).

The fractional saturation for type 1 and type 2 sites (θ_1 and θ_2) are given by:

$$K_1 = \frac{\theta_1}{(1-\theta_1)[L]} \tag{2.49}$$

$$K_2 = \frac{\theta_2}{(1-\theta_2)[L]} \tag{2.50}$$

and the free ligand concentration is given by:

$$[L] = [L]_t - [M]_t(N_1\theta_1 + N_2\theta_2) \tag{2.51}$$

where N_1 is the number of type 1 binding sites and N_2 is the number of type 2 binding sites. Combining equations 2.49, 2.50 and 2.51, followed by rearranging gives the cubic equation:

$$[L]^3 + \left(\frac{1}{K_1} + \frac{1}{K_2} + (N_1 + N_2)[M]_t - [L]_t\right)[L]^2 + \left(\left(\frac{N_1}{K_2} + \frac{N_2}{K_1}\right)[M]_t - \left(\frac{1}{K_1} + \frac{1}{K_2}\right)[L]_t + \frac{1}{K_1K_2}\right)[L] - \frac{[L]_t}{K_1K_2} = 0 \tag{2.52}$$

Given values for $[L]_t$, $[M]_t$, N_1, N_2, K_1 and K_2 equation 2.52 can be solved for [L]. Substitution of [L] into equations 2.49 and 2.50 gives values for θ_1 and θ_2.

The heat content of the solution after the ith injection is given by:

$$Q_i = [M]_t V_o(N_1\theta_1\Delta H_1 + N_2\theta_2\Delta H_2) \tag{2.53}$$

As with the model for one set of sites intial guesses are made of the fitting parameters, which are optimized by statistical techniques until convergence.

"too much a chemist and a chemist too long to consider myself a real man of letters; yet too distracted by the vari-coloured, tragic or strange landscape to feel a chemist in every fibre."

Primo Levi

Chapter three

Materials and methods

3.1 Proteins

WT-α-Lactalbumin (Type I) and calcium depleted WT-α-lactalbumin (Type III) were purchased from Sigma Chemical Co., 3SS-α-lactalbumin (section 3.1.1) was obtained from Dr. T. E. Creighton, EMBL, Heidelberg, Germany.

Pf1-gV protein and complex (section 3.1.2) were obtained from Prof G.G. Kneale and Dr K.G. Davis, Portsmouth University.

MS2 viral coat protein and AhrC were prepared and purified (in the laboratory of Prof. P.G. Stockley, Department of Genetics, University of Leeds) as described in sections 3.1.3 and 3.1.4 respectively.

3.1.1 Preparation of 3SS-α-lactalbumin.

3SS-BLA was prepared on our behalf by members of the laboratory of Thomas Creighton (Heidelberg) according to the following protocol (Ewbank & Creighton, 1993). BLA was taken up in buffer (0.1M Tris, 0.2M KCl, 10mM $CaCl_2$, pH 7.0) to a concentration of 20 mg/ml. To this one tenth volume of 60mM dithiothreitol (DTT) in the same buffer was added and the mixture was incubated for two minutes at 25 °C. Following this period free thiols were trapped by addition of one fifth volume of 0.6M iodoacetamide in 0.5M Tris pH 8.0 and incubation for a further 2 minutes at 25 °C. The protein was separated from excess blocking agent by rapid gel filtration into 20mM ammonium acetate (Pharmacia fast desalting column) and further purified by reversed-phase HPLC (Dynamax-300A C_4 column 4.6x250mm, with a linear gradient of 25% to 50% acetonitrile in 0.1% TFA). The protein was then lyophilized and delivered to Glasgow for DSC.

3.1.2 Preparation of Pf1-gV protein and complex.

Pf1-gV protein and complex was prepared on our behalf by members of the laboratory of G. Geoff Kneale (Portsmouth) according to the following protocol (Plyte & Kneale, 1993). *Pseudomonas aeruginosa* (strain K) was grown to midlog phase, infected with phage Pf1 and grown for a further 4 hours. The cells were pelleted by centrifugation at 4000 RPM for 30 minutes at 4 °C, and then resuspended in 200 mL of buffer (10 mM Tris, 100 mM NaCl, pH 7.5) and centrifuged a further twice to remove all trace of the growth medium.

The resuspended pellet was applied to a Sepharose CL-2B column and equilibrated with 10 mM Tris, 100 mM NaCl, pH 7.5 buffer (buffer A). Fractions discovered to contain complex (using SDS-PAGE) were pooled and concentrated by centrifugation at 50,000 RPM at 4 °C for 3 hours. The pellet thus formed was resuspended in buffer B (as buffer A but with 1 mM EDTA added) before application onto a phenyl-sepharose fast flow (high sub) hydrophobic interaction column equilibrated with buffer A. Fractions found to contain complex were concentrated as before. The pellet was resuspended in the buffer of choice and clarified by centrifugation at 15,000 RPM for at 4 °C for 30 minutes (Davis et al, 1995).

3.1.3 MS2 preparation and purification.

The following scheme was obtained from Dr Robert Mastico, and was used to prepare and purify MS2 viral coat protein.

1 Cell growth.

E. coli cells containing a plasmid for expression of the MS2 coat protein were grown at 37 °C in 4L of 2TY broth (16g/l tryptone, 10g/l yeast extract, 5g/l sodium chloride titrated to pH 7 with sodium hydroxide), with 100 μg/ml ampicillin, until the cells entered the log phase (i.e. when optical density reached 0.40-0.60 absorbance units). Then the inducer Isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 1mM, and allowed to incubate overnight at 37 °C.

2 Initial isolation.

The cells were pelleted by centifugation at 3,000 RPM for 30 min at 4 °C and then resuspended in cold buffer (50mM Hepes, 100mM NaCl, 10 mM DTT, 5mM EDTA and 1mM phenylmethylsulfonyl fluoride (PMSF) titrated to pH 7.4 with NaOH). Cells were then sonicated to break cell walls followed by centrifugation at 15,000 RPM for 20 min at 4 °C. Magnesium acetate and DNase I to final concentrations of 6mM and 10 μg/ml respectively were added to the supernatant, and incubated for 30 min at room temperature, with constant stirring, followed by further centrifugation at 5,000 RPM for 30 min at 4 °C, retaining the supernatant.

3 Ammonium sulphate precipitation.

Ammonium sulphate was added to the supernatant to 30% saturation (16.4g/100ml) and allowed to stand for 2 hours at 0°C. The partially purified protein was then seperated from the impurities thus precipitated by centrifugation at 5,000 RPM for 30 min, 4°C. SDS-PAGE was employed to check that the protein was located in the supernatant.

4 Purification.

MS2 viral coat protein empty capsids were finally purified by differential sedimentation through a linear sucrose gradient with ultra-centrifugation at 28,000 RPM for 4-5 hours at 4 °C. Layers of 15, 30 and 45% w/w sucrose in 10mM Hepes, 100mM NaCl, 1mM EDTA pH 7.4 were placed in a centrifuge

tube with most dense layer at the bottom, and this was allowed to linearise overnight at 0°C. After centrifugation the contents of the tubes were removed from the bottom, and fractionated, monitoring UV absorbance via an inline flow cell. Fractions were assayed for the presence of MS2 Viral Coat protein by SDS-PAGE.

5 Precipitation.

Protein was precipitated, overnight, with 50% saturation ammonium sulphate, and then pelleted by centrifugation at 3,000 RPM for 2 hours at 4°C. The pellet was resuspended as required, followed by dialysis to remove ammonium sulphate.

3.1.4 AhrC preparation and purification.

AhrC was prepared according to the following scheme provided by Coleen Miller.

1 Cell growth.

A rich growth medium was required, containing 100 g/L yeast extract, 56 g/L KH_2PO_4, 378 g/L $K_2HPO_4.3H_2O$, 5 g/L $MgSO_4.7H_2O$, 10 g/L glucose, 10 mg/L thiamine and 200 mg/L ampicillin. *E. coli* Cells, containing a plasmid for expression of the protein, were grown in 10 x 750 ml protions of this medium at 37°C until the optical density at 600 nm was 1.2 absorbance units. IPTG was

then added to a final concentration of 1.2 g/L, followed by a further 3 hour incubation period. Cells were then harvested by centrifuging at 3,000 RPM for 15 min.

2 Isolation.

A stock of ARG buffer (20 mM Tris/HCl, 10 mM $MgCl_2$, 10 mM 2-mercapto-ethanol, 1 mM PMSF, 25 μM N-tosyl-L-phenylalanine chloromethyl ketone (TPCK) at pH 7.5) was prepared (PMSF and TPCK are included as protease inhibitors). Cells were resuspended in 150 ml ice cold ARG buffer, and cell walls broken by sonicating in 15 ml aliquots. The aliquots were pooled and centrifuged at 12,000 RPM for 30 min. The pellet was resuspended in 120 ml ARG buffer and treated with 1 μg/ml DNase I at 37°C for 15 min, NaCl to a final concentration of 0.5 M was added and incubated for a further 15 min at 37 °C, following which the solution was centrifuged at 12,000 RPM for 30 min. The supernatant was dialyzed against two portions of ARG buffer with 75 mM NaCl for a total time of around three hours, causing the AhrC to precipitate selectively from solution. The precipitate was harvested by centrifugation at 12,000 RPM for 30 min. The pellet was resuspended in 75 ml ARG buffer with 0.5 M NaCl, then dialyzed against ARG bufer with 0.3 M NaCl overnight.

3 Purification.

An S-sepahrose column was equilibrated with ARG buffer with 0.3 M NaCl, and buffer was run through column until buffer running onto and off from column had the same conductivity. Dialyzed protein was then loaded onto the column

and washed with 0.3 M NaCl until the flow through showed no absorbance at 280 nm. The column was then developed with a 0.3 to 0.6 M NaCl gradient in ARG buffer at a flow rate of approximately 1 ml/min. Fractions with an absorbance at 280 nm were assayed for AhrC by SDS-PAGE, and fractions containing AhrC were pooled, and protein precipitated with 25% PEG-6000.

3.2 Chemicals.

3.2.1 General chemicals.

Unless otherwise stated all chemicals were of Analar grade, and were used without further purification, with the exception of succinimide used in fluorescence quenching studies. Succinimide was purchased from Sigma Chemical Co. and was re-crystallised three times from hot ethanol prior to use until no fluorescence from trace impurities was detectable.

3.2.2 Chemicals used in protein preparations.

Isopropyl-Beta-D-Thioglactoside (IPTG) was purchased from NBLT Enzymes Ltd.

Phenylmethylsulfonylfluoride (PMSF), N-[2-hydroxyethyl]piperazine-N'-[2-ethane sulfonic acid] (Hepes), tris(hydroxymethyl)methylamine (Tris), piperazine,

thiamine, 2-mercaptoethanol, ethylenediaminetetracetic acid (EDTA), dithiothreitol (DTT), ammonium sulphate, N-tosyl-L-phenylalaninechloro-methyl ketone (TPCK), acrylamide, N,N'-methylene-bis-acrylamide and 2-mercaptoethanol were all purchased from Sigma Chemical Co.

Ammonium persulphate (AMPS), N,N,N',N'-tetramethyl-diamine (TEMED) were purchased from Shandon Ltd.

3.3 Differential Scanning Calorimetry.

The Microcalorimeters used are described in some detail in chapter 2, however the general experimental set up is described here and in section 3.4.

3.3.1 Procedure for obtaining a thermogram.

All DSC was carried out using a Microcal MC-2D ultra sensitive calorimeter, fitted with an EM Electronics N2a nanovolt pre-amplifier. The DSC was operated under the control of a PC which was also used for data collection. A nominal scan rate of 60 degrees/hour and a data acquisition and averaging time, or "filter", of 15s, was employed, unless otherwise stated in the text.

Where possible, samples for DSC were equilibrated with appropriate buffer. For the samples prepared in Glasgow this was done either by dialysis or by simply dissolving, in buffer, protein which had been extensively dialysed against pure water and

then lyophilised. Samples of Pf1 gene five complex provided by the Portsmouth group were presented along with appropriately equilibrated buffer. This buffer was used both as a reference, and for buffer baseline scans. Both sample and buffer were degassed by placing the solution under vacuum in a small desiccator fitted with a water aspirator, and gently stirring using a magnetic stirrer.

Both cells of the calorimeter were rinsed at least five times with buffer prior to filling. Filling the cells was carried out using a syringe fitted with a length of narrow gauge Teflon tubing. Care was taken to ensure that the syringe, and tubing, were free from air bubbles. It is critically important that there are no bubbles either in the cells or in the inlet tube. First one or more buffer base line scans were obtained by filling both sample and reference cells with the buffer and scanning in the normal manner. Multiple scans, which could be done overnight, were generally required when the machine had been unused for a period of time.

The sample solution was then loaded into the sample cell. The small amount of buffer left in the cell prior to filling leads to a slight dilution of the sample, however this is avoided by rinsing the cell with a little of the sample and then mixing the rinsing with the rest of the sample or discarding. A little of the sample was retained for subsequent protein concentration estimation. During DSC a pressure head of 1-2 atm of nitrogen was maintained in order to prevent boiling of the sample and to suppress bubble formation.

The procedure for obtaining a thermogram, both for the sample and the buffer base line, was as follows. The scan parameters were entered into the computer. These consist of: number of scans, initial temperature (20°C unless otherwise stated), final

temperature (100°C unless otherwise stated, although it is possible either to stop the scan before this temperature or allow the scan to continue, up to a maximum of 110°C if necessary, at any point during the scan.), scan rate, filter (as indicated above, these were usually set to 60 deg/hour and 15s respectively) and file name(s). Prior to starting, the cell temperatures were allowed to equilibrate, with the water circulator on, for 1 hour, or until the difference in temperature between the cells and the adiabatic jacket was less than 1°C, this generally takes 20-30 minutes. A further, 5 minute, period of equilibration, during which the feedback system (see section 2.2.1) is used to finely balance the temperatures, preceded the scan.

3.3.2 Analysis of DSC data.

All DSC data were analysed making use of the Origin software package available from Microcal, with the DSC add-on installed.

First it was necessary to normalise the thermograms with respect to the actual scan rate. Then the buffer base line was subtracted from the sample scan followed by normalisation with respect to protein concentration, and cell volume. A non-two-state model was, generally, fitted to this normalized experimental data using the facility built into the software package (Where this is not the case details will be given in the text). The mathematical basis for the models used by Origin is described in section 2.2.4.

3.3.3 Temperature axis calibration.

Occasionally it was expedient to check the temperature calibration of the DSC. Standard samples provided by Microcal were used for this task. The samples consist of sealed steel capillaries containing small amounts of hydrocarbons with known melting points, which are placed into water filled cells, and a scan started from 20°C to 80°C at a scan rate of 45 deg/hour, with a filter of 8s. One of the capillaries contains a substance with a melting point of 28.2°C the other contains a substance with a transition at 73.8°C and a melting point of 75.9°C. Since all transitions are endothermic the standard in the sample cell gives rise to a positive peak(s) while the one in the reference cell gives rise to a negative peak(s).

3.3.4 Excess heat capacity axis calibration.

It was also expedient to make occasional checks on the calibration of the excess heat capacity measurements. For this purpose electrical calibration is sufficient. With both cells filled with water, a small current, up to a maximum equivalent of 12 mcal/min, is passed through the calibration heater, for 4 minutes, and the amount of heat supplied by the feedback system to maintain cell temperature equilibrium is measured in the usual manner. The height and area of such calibration pulses should match the values entered, after subtraction of a water/water baseline.

3.4 Isothermal Titration Calorimetry.

3.4.1 Procedure for obtaining a binding isotherm.

All titration calorimetry was carried out using an Omega ITC instrument from Microcal, operating under PC control. The reference cell was filled with distilled, degassed water containing 0.05% sodium azide (in order to prevent growth of microorganisms). This reference is only occasionally re-filled. Samples were prepared for ITC as for DSC. The sample cell was filled with protein in the same manner as the DSC cells, taking care to exclude possible bubbles. After filling the system was allowed to reach thermal equilibrium, at the desired temperature (usually 25°C). This could take some considerable time (especially with the cold room temperatures occasionally experienced in Glasgow!); to speed up the process the filling syringe could be gently warmed, using the heat of the hand or, with great care, a hair drier set to blow warm air. The stirring injection syringe was then filled with substrate (synthetic RNA 19mer for MS2 experiments and Arginine for AhrC experiments). Again care was taken in order to ensure that no bubbles were trapped in the syringe. The syringe was mounted in the calorimeter and used to stir at 400 RPM. A further period of final equilibration was allowed (approximately 5 min), after which data acquisition with no injections was initiated; when this showed a stable, noise-free baseline injections could begin. Details of the injection schedule to be followed (indicated where appropriate in the relevant chapter) were entered into the PC, which then controlled the stepper motor (attached to the syringe plunger) and data acquisition. The data were stored and analysed later using the Microcal Origin software package, as described later.

Appropriate control experiments were set up in exactly the same fashion. These controls were as described in section 2.3.3.

3.4.2 Analysis of ITC data.

As with DSC, the Origin software package from Microcal has an add-on available for analysis of ITC data, this was used for all analysis. For each injection a baseline is drawn and the peaks are integrated, to give heats which can then be analysed according to simple models to give stoichiometry, binding constant and ΔH. A fuller description is to be found in section 2.3.

3.4.3 Calibration.

Electrical calibration, similar to that described for DSC is provided on the Omega machine. This was used for routine calibration. Further calibration using standard reactions were performed periodically by Mrs. Margaret Nuttley and Dr. Alan Cooper.

3.5 Fluorescence spectroscopy.

3.5.1 Intrinsic fluorescence.

Fluorescence spectroscopy was carried out at 25°C in a Spex inc. FluoroMax spectrofluorometer, fitted with a circulating, thermo-statted water bath. Control of the machine, and data acquisition were accomplished using a PC running the dM3000 software package, also from Spex inc.

In order to prevent complication by internal filtering affect and to prevent saturation of the detector samples for fluorescence were diluted with appropriate buffer to give an absorption maximum of approximately 0.1 absorbance units (1cm path length).

The protein emission spectra were obtained by exciting the samples at 295nm with both emission and excitation slits set to 4.25nm (equivalent to slit widths of 1 mm using the Spex machine which has a dispersion of 4.25 nm/mm. bandpass = slit width (mm) x dispersion (nm/mm)). Data were collected every 1nm and averaged over 1s. In all cases an emission spectrum was obtained for the buffer, and this was subtracted from protein fluorescence measurements. This was crucial since the emissions measured overlapped with the water Raman band.

3.5.2 Fluorescence quenching.

Fluorescence quenching experiments were carried out exactly as described above for protein intrinsic fluorescence measurements, with the exception that the initial volume of each sample in the cuvette was 2.0ml and this was titrated with accurately prepared solutions of approximately 2M succinimide containing the protein, in 8 steps to give a final succinimide concentration of approximately 600mM. The maximum fluorescence intensity, after correction for water Raman band, was measured and a Stern-Volmer plot of I_0/I versus succinimide concentration was constructed, where I_0 is the intensity with no added quencher and I is the measured intensity. Quenching data were analysed in each of two ways. Firstly the following model was constructed (see chapter 4 section 4.4 for the derivation of this equation) in which it was assumed that there existed two populations of tryptophans with different, non-zero, quenching rates,

$$\frac{I_0}{I} = \frac{(1+K_A[Q])(1+K_B[Q])}{1+(K_B + f_B(K_A - K_B))[Q]} \qquad (3.1)$$

Where I_0 is the initial fluorescence intensity, I is the intensity measured, K_A and K_B are the quenching rate constants for population a and b respectively, f_B is the fraction of fluorescence intensity deriving from population b and [Q] is the quencher concentration

This gave a curve which fitted the experimental data with some accuracy. However, the approximation that there were two populations is possibly inadequate. BLA in fact contains four tryptophan residues each of which is likely to undergo quenching at different rates. No model was constructed for this case since any such model would necessarily have too many variables to be of any practical use.

Eftink and Ghiron (1976a) suggest that quenching which displays non-linear Stern-Volmer plots, such as found here, can be compared by limiting the analysis to the initial portion of the plot, which reflects the quenching of the more accessible trp residues. The Stern-Volmer plots here were found to be closely approximated by two straight lines, reflecting the quenching of more accessible residues in the early portion and more buried residues as quencher concentration rises. Quenching rates are quoted on this basis for the more deeply buried residues.

3.6 Protein concentration estimation.

3.6.1 UV absorption of proteins.

All proteins containing the aromatic residues tyrosine and tryptophan have a significant absorbance maximum at 280nm. Where the extinction coefficient (ε) is known it is possible to calculate the concentration of a protein sample according to the Beer-Lambert law. Where available, literature values for the extinction coefficient were used. Where these were not available estimations of the extinction coefficient were made using the method of Gill and von Hippel (1989), described below. The following extinction coefficients and molecular weights were used.

Protein	$\varepsilon^{280}_{1mg/ml}$	molecular weight
α-Lactalbumin	2.01°	14,186°
3SS-α-Lactalbumin	2.01+	14,186+
lysozyme	2.65°	14,500°
AhrC	0.586*	16,673^

TABLE 3.1 Summary of extinction coefficients used. + assumed to be the same as for WT-α-lactalbumin. * calculated by the method of Gill and von Hippel, see below. ° From McKenzie & White, 1991. ^ From Czaplewski et al, 1992.

MS2 viral coat protein samples had a UV absorbance maximum at 260 nm indicative of substantial contamination by nucleic acid. It is assumed that the contamination takes the form of small fragments of nucleic acid on the basis that the DNase treatment step of the preparation is expected to break all the long chains (section 3.1.2). The coat protein is known to be capable of encapsidating non-phage nucleic acid (Pickett & Peabody, 1993). This contamination made it impossible to calculate concentrations from UV absorbance since the nucleic acid absorbance overlaps that of the protein. It was therefore necessary to employ a method which did not rely on measurement of A_{280}, and so the coomassie brilliant blue assay and the method of Warburg and Christian were used, as described below.

3.6.2 Estimation of protein extinction coefficients.

As stated above, where a literature value for the extinction coefficient was not available the method of Gill and von Hippel (1989) was employed to calculate the extinction coefficient. This method is, in theory, applicable to any protein for which the number of tryptophan, tyrosine and cysteine residues is known, however it is an empirical rule and is strictly accurate only for denatured proteins. It has been found to be accurate (± 5%) for most proteins checked thus far although no attempt was made to test its accuracy for the protein AhrC. The method is based on Equation 3.2, which gives the molar extinction coefficient for the native protein:

$$\varepsilon_{protein} = a\varepsilon_{Tyr} + b\varepsilon_{Trp} + c\varepsilon_{Cys} \qquad (3.2)$$

in which a, b and c are the number of each type of residue in the protein and ε_{Tyr}, ε_{Trp} and ε_{Cys} are the molar extinction coefficients of tyrosine, tryptophan, and cysteine respectively, based on model compound studies. The values used for these parameters are those in table 3.2

Residue	ε^{280}
Trp	5690
Tyr	1280
Cys	120

TABLE 3.2 Summary of the extinction coefficients (at 280nm) of the amino acids which absorb light in the UV region.

AhrC is found to contain 1 Trp, 3 Tyr and 2 Cys residues and so the method of Gill and von Hipple gives it an extinction coefficient of $\varepsilon^{280}_{1mg/ml} = 0.585$.

3.6.3 Coomassie brilliant blue assay.

In free solution the dye coomassie brilliant blue (coomassie G250) has a brown colour, with an absorbance maximum at 465nm, however when the dye is bound to protein it has a brilliant blue colour with absorbance maximum at 620nm. The absorbance ratio A620/A465 is measured and plotted against standard protein solutions of known concentration to form a calibration graph. The proteins used as standards were α-lactalbumin, lysozyme and bovine serum albumin.

The dye solution used for the assay was a 0.06% solution of coomassie G250 in 3% perchloric acid. The solution was stirred overnight and filtered to remove any residual solid material from the dye. Measurements of the A620/A465 ratio were taken against a water reference, and a blank measurement, in which 1ml coomassie stain was mixed with 1ml water was taken and this was subtracted from all subsequent measurements. For each protein used as standard, samples were prepared with accurate concentrations of approximately 10, 20, 30, 40 and 50μg/ml. The calibration curve was constructed by taking 1 ml of protein solution, adding 1 ml of coomassie stain and allowing the colour to develop (around 5-10 minutes), and then measuring the A620/A465 ration. Measurement of the concentration of the protein in question was carried out by diluting the sample to give an expected concentration somewhere around the middle of the calibration range and taking A620/A465 measurement and then calculating the concentration from the value obtained from the calibration curve and the

dilution factor used. In practice two or three different dilutions of the sample of interest were made and an average value calculated.

3.6.4 Warburg and Christian method

The method of Warburg and Christian (Warburg & Christian, 1942) involves measuring the ratio of absorbance at 280 nm and 260 nm, and using an empirical factor, given in the table below, to calculate the concentration of protein in samples contaminated with nucleic acid. The A_{280}/A_{260} ratio of a suitably diluted protein solution is measured and the protein concentration calculated from the following equation:

$$C = \frac{A_{280}F}{d} \tag{3.3}$$

where C is the concentration of the protein in mg/ml F is the factor from taken from table 3.3 and d is the path length in cm.

A_{280}/A_{260}	% Nucleic acid	Factor	A_{280}/A_{260}	% Nucleic acid	Factor
1.75	0	1.118	0.86	5.20	0.671
1.60	0.30	1.078	0.84	5.60	0.650
1.50	0.56	1.047	0.82	6.10	0.628
1.40	0.87	1.011	0.80	6.60	0.605
1.30	1.26	0.969	0.78	7.10	0.581
1.25	1.49	0.946	0.76	7.80	0.555
1.20	1.75	0.921	0.74	8.50	0.528
1.15	2.05	0.893	0.72	9.30	0.500
1.10	2.40	0.863	0.70	10.3	0.470
1.05	2.80	0.831	0.68	11.4	0.438
1.00	3.30	0.794	0.66	12.8	0.404
0.96	3.70	0.763	0.64	14.5	0.368
0.92	4.30	0.728	0.62	16.6	0.330
0.90	4.60	0.710	0.60	19.2	0.289
0.88	4.90	0.691			

TABLE 3.3 Correction factors for use in Warburg and Christian method. Nucleic acid is expressed as a percentage of the total (protein + nucleic acid).

"The great tragedy of science - the slaying of a beautiful theory by an ugly fact."

T.H.Huxley.

Chapter Four

α-Lactalbumin

4.1 Introduction

The thermodynamic basis for the stability of the three dimensional structure of proteins is poorly understood, partly due to the large number of interactions involved. One of the interactions responsible for protein stability which is most readily accessible to the experimenter is the disulfide bridge (section 1.2.6). Disulfide bridges are commonly found in extracellular proteins, where the generally oxidizing conditions favour their formation. It is often possible to reduce one or more of the disulfide bridges within a protein, and prevent its reformation using some suitable blocking group. This has the advantage over other possible perturbations that there is, generally, no appreciable effect upon other interactions; unlike, for example, changing from hydrogenated to deuterated solvents, which has a broad influence (chapter 5). Following the recent calorimetric experiments on the consequences of the disulfide bridge between residues 6 and 127 of HEWL for the protein's stability (Cooper et al, 1992), it was decided to study the structurally homologous protein bovine α-lactalbumin (BLA).

BLA is a member of a family of α-lactalbumins from various sources. BLA (figure 4.1) contains 123 amino acid residues, with a molecular weight of 14,186. Like HEWL, BLA has 8 Cys residues (Schmidt & Ebner, 1971) making up 4 disulphide bridges between the Cys residues at 6-120, 28-111, 61-77 and 73-91 (Brew et al, 1970). As noted above BLA has a high degree of sequence and structural homology with HEWL (McKenzie & White, 1991), compare figures 4.1 and 5.1. Despite these similarities, their functions are quite different.

BLA, in common with the rest of the α-lactalbumin family is a calcium binding protein (Hiraoka et al, 1980). Estimates of the binding constant have been given by various authors as 10^6-10^9 M^{-1} (Hiraoka et al, 1980; Permyakov et al, 1981; Murakami et al, 1982; Segawa & Sugai, 1983; Hiraoka & Sugai, 1984; Bryant & Andrews, 1984; Hiraoka & Sugai, 1985; Desmet & Cauwelaert, 1988). BLA from which the bound calcium ions have been removed is referred to as apo-α-lactalbumin (apo-BLA) and where necessary for clarity protein with the calcium present may be referred to as holo-α-lactalbumin (holo-BLA).

BLA is one of the major milk whey proteins (the other is beta-lactoglobulin). Along with the protein galactosyl transferase it is responsible for the synthesis of lactose (Ebner et al, 1966). The role of BLA is to alter the specificity of the galactosyl transferase protein (Brew et al, 1968) it is therefore known as a specifier protein. (See also McKenzie & White, 1991; Schmidt & Ebner, 1971)

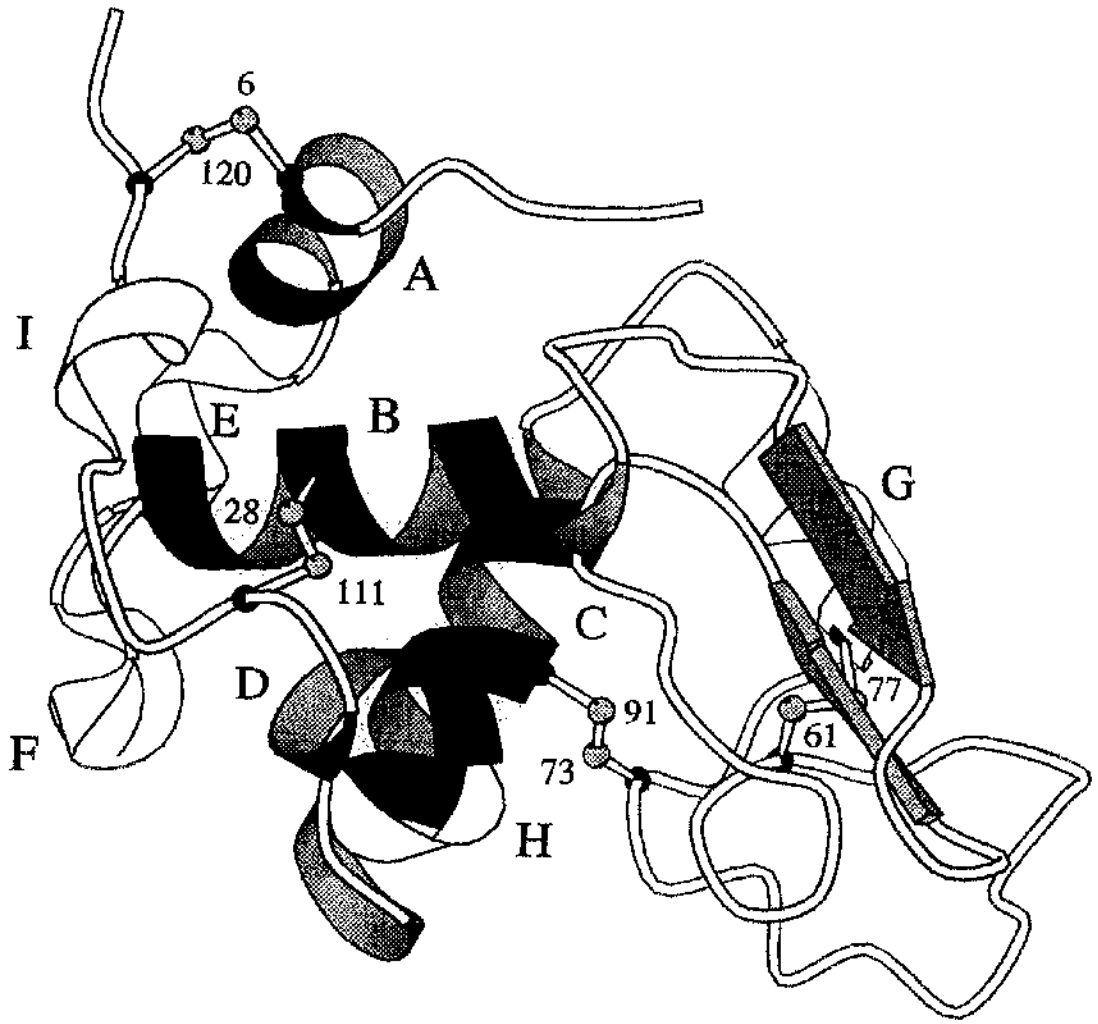

Figure 4.1 The three dimensional structure of bovine α-lactalbumin.

BLA has been the subject of much discussion recently, especially concerning the role of the partially folded intermediate state, known as the molten globule (MG) state (Kuwajima, 1977 & 1989; Ptitsyn, 1992). The MG state is believed to be an important kinetic intermediate in the protein folding process (Kuwajima et al, 1985; Ikeguchi et al, 1986), and so has stimulated much interest. The MG state can be stabilised at room temperature by lowering of the pH (Kuwajima, 1976) or by removal of the bound calcium (Hiraoka & Sugai, 1984; Ikeguchi et al, 1986; Yutani et al, 1992). Addition of sodium to apo-BLA is known to stabilise the native (N) state, presumably by binding to the, now free, calcium binding site (Desmet et al, 1987; Hiraoka & Sugai, 1984).

Despite much work the structure of the MG remains poorly defined. It is by no means certain that the acid MG has identical structure to the apo MG (Shimizu et al, 1993; Keiderling et al, 1994; Ford et al, 1995; Wilson et al, 1995) and the term is often used very loosely to describe what may in fact be a range of states (Finkelstein & Shakhnovich, 1989; Shakhnovich & Finkelstein, 1989; Ptitsyn, 1992; Kiefhaber et al, 1995), however all of the states referred to as MG have a number of factors in common. They all share a loss of the near-ultra violet circular dichroism (UV CD) band, indicating a loss of tertiary structure, while the far-UV CD indicates that the MG contains an amount of secondary structure similar to that found in the N state (Kuwajima et al, 1977 & 1985; Dolgikh et al, 1981 & 1985). It has been suggested that the tertiary structure of the MG fluctuates with a rate constant of 10^7 s^{-1} (Dolgikh et al, 1981).

^{1}H NMR results show that, while the MG has lost most of the N state's long range tertiary structure some tertiary structure none the less remains (Shimizu et al, 1993; Alexandrescu et al, 1993). Further, the secondary structure which exists in the MG state is not necessarily that found in the N state (Alexandrescu et al, 1993). This has also been suggested by Raman optical activity (ROA) and vibrational circular dichroism (VCD) studies (Keiderling et al 1994; Wilson et al, 1995).

Further evidence for the lack of tertiary structure in the MG state comes from studies in which one of the proteins 4 disulfide bridges has been reduced. In this modified form (3SS-BLA) the MG can undergo spontaneous disulfide rearrangement to give a range of non-native pairings (Ewbank & Creighton, 1991), which are inconsistent with a highly ordered tertiary structure.

Given the lack of specific tertiary interactions it is perhaps surprising that the MG has a compact structure, comparable to that of the N state, however, intrinsic viscosity measurements (Dolgikh et al, 1986) quasielastic light scattering (Gast et al, 1986) and high angle diffuse X-ray scattering (Damaschun et al, 1986) all point towards the MG having similar hydrodynamic properties to the N state.

There has been considerable discussion of the enthalpy change associated with the transition from the MG to the denatured (D) states (Xie et al, 1991 & 1993; Yutani et al, 1992). Xie et al found that there was an enthalpy change for the MG to D transition (associated with the loss of the secondary structure) of 100.8 kJmol^{-1} whereas the enthalpy change for the transition from N to MG is only 32.2 kJmol^{-1} (Xie et al, 1991). This is in

direct opposition to the results of Yutani et al who found no enthalpy change upon unfolding of the residual structure in the MG (Yutani et al, 1992). Yutani et al are in agreement with the interpretations of Shimizu et al who state that the transition from MG to D is not a cooperative two state process based on the their NMR results (Shimizu et al, 1993) which show that the unfolding of different regions of the structure are not coincident.

It has already been mentioned that BLA contains four disulphide bridges. It is possible to reduce specifically one of these, making a three disulfide containing modified form of the protein, referred to as 3SS-BLA. Shechter et al (1973) found that the reduction of BLA with dithiothreitol (DTT) proceeded in two steps, one fast step followed by one much slower step. If at the end of the fast step the partially reduced protein was carboxymethylated using 2-iodoacetic acid it was found that the disulfide between residues 6 and 120 had been specifically reduced and reformation of the disulfide prevented by the formation of two carboxymethylcysteine residues. Iyer & Klee (Iyer & Klee, 1973) further found that if the reduction with DTT was carried out at 0°C only the C6-C120 disulfide bridge was reduced.

This hyper-reactivity has been quantified, and it was found that the C6-C120 disulfide bridge was 140 times more reactive than average. This has been explained in terms of molecular strain imposed on the crosslink by the geometry into which it is forced upon folding to the native state (Kuwajima et al, 1990).

3SS-BLA and wild type (WT) 4SS-BLA bind equally with anti-BLA antibodies (Shechter et al, 1973), indicating that there are no gross changes in the structure of the

protein upon modification. This is backed up by the CD (Kuwajima et al 1990) and NMR (Ikeguchi et al, 1992) data. However, there are minor differences as evidenced by the fact that 3SS-BLA is more readily digested by trypsin (Shechter et al, 1973). This may be due to minor conformational differences or to differences in the dynamic properties of the two.

4.1.1 Previous work on HEWL

Like BLA it is possible to remove one disulfide bridge from the homologous protein HEWL forming the modified protein 3SS-HEWL. The effect of this modification has been studied using DSC and NMR (Radford et al, 1991; Cooper et al, 1992).

It was found that the removal of the crosslink and the introduction of the carboxymethyl blocking group did not appear to alter in any way the secondary or tertiary structure of the protein. However it did, as expected, reduce the thermal stability of the protein by a significant amount. The study showed that 3SS-HEWL was between 25 and 30°C less stable that the wild type protein. Analysis of the thermodynamic parameters of unfolding showed that this reduction in stability could be attributed entirely to the increased stability of the unfolded state of the modified protein resulting from the much increased conformational flexibility and hence absolute entropy in the absence of the constraining crosslink. The difference in the unfolding entropy of the two forms of the protein was calculated to be $\Delta\Delta S298 = 90\ JK^{-1}mol^{-1}$, while there was no significant difference in their unfolding enthalpy. This was in good agreement with the theoretically predicted values of between 70 and 95 $JK^{-1}mol^{-1}$ (Section 1.2.6; Schellman, 1955; Poland & Scheraga, 1965;

Pace et al, 1988). Cooper et al also found that there was no difference in ΔC_p between the two forms of the protein.

In other experiments (Kuroki et al, 1992a) the effect of the 77-95 disulfide bridge in human lysozyme was examined by comparison of the wild type protein to a mutant in which the cysteine residues at positions 77 and 95 were each replaced by alanine. This was found to reduce the thermal stability of the protein by over 14°C. Contrary to the results of the previously described work it was found that there was a difference in the enthalpy of unfolding for the two proteins and the difference in free energy of $\Delta\Delta G = 8.4$ kJmol^{-1} was attributed to this (Kuroki et al, 1992a). These findings are in line with the theory presented by Doig and Williams (Section 1.2.6; Doig & Williams, 1991). However, the Doig and Williams theory predicts that the removal of the crosslink should result in an increased ΔC_p. No such increase in ΔC_p was observed. The effect of the mutations on the dynamics of the protein was probed by normal mode refinement of a 1.5 Å resolution set of X-ray data (Kidera et al, 1994). Simply comparing the isotropic temperature factors for non-hydrogen atoms (B-factors) it appears that the mutant has a significantly more flexible structure than the wild type protein, however B-factors include both internal terms, which give an insight into the dynamic structure of the protein, and external terms, which depend on the quality of the crystals.

Normal mode refinement allows these two to be considered separately (Kidera & Go, 1990, 1992; Kidera et al, 1992). It was found that the large differences in B-factors resulted mostly from external factors and that only residues in the vicinity of the mutations

were significantly more flexible. The conclusion was that removal of the 77-95 dislufide bridge did not have a significant effect on the dynamic structure of the protein.

4.1.2 Previous work on BLA

3SS-BLA is known to adopt a native like conformation (Shechter et al, 1973; Kuwajima et al, 1990; Ikeguchi et al, 1992) under the conditions to be used. This is crucial if one wishes to carry out the kind of comparative studies discussed below and relate changes observed to a particular phenomenon.

The contribution of the 6-120 disulfide bridge to the stability of the native conformation of BLA has been studied previously (Ikeguchi et al, 1992) by non-calorimetric means. In that study circular dichroism (CD) measurements of disulfide intact and modified BLA were made and the free energy of unfolding, in the presence of chemical denaturants, was calculated for both. This gave a value for the difference in free energy between the intact protein and the protein lacking the disulfide bridge between cysteine 6 and cysteine 120 of $\Delta\Delta G^{\circ} = 13.0$ kJmol^{-1} (in the presence of 1mM Ca^{2+}, at 4.5 °C). The authors attributed the decreased stability, resulting from removal of the disulfide bridge entirely to the decrease in conformational entropy of the unfolded state. This was done on the basis that the difference in protein-solvent interactions between the native and unfolded conformations of WT-BLA is similar to that between the native and unfolded conformations of 3SS-BLA and, therefore, any solvent ordering contribution (as proposed by Doig and Williams, 1991) to the overall thermodynamics must be small. No account was taken of

possible changes to the entropy of the native state resulting from the removal of the disulfide bridge.

4.2 Thermodynamics of the unfolding of WT- and 3SS-BLA

The DSC thermograms for both wild type and 3SS-BLA (figure 4.2) are typical for small globular proteins, showing a single cooperative transition to the unfolded state. DSC measurements were made on both disulfide intact and modified BLAs under various conditions of pH such that the transition midpoint temperature (T_m) of the intact protein varied from 50-73 °C, and that of the three disulfide protein varied in the range 46-59 °C, giving an overlap of 9 °C. Typical DSC traces for both 3SS-BLA and disulfide intact BLA, under precisely the same buffer conditions, are presented in figure 4.2. The difference in thermal stability is obvious and, as expected, 3SS-BLA is considerably less stable than its four disulfide counterpart. Indeed the difference in T_m (ΔT_m) under these conditions is 13 °C. Under the range of conditions used ΔT_m varies between 10-15 °C.

The T_m of both forms of the protein increases with increasing pH up to pH 5, reaching a plateau in the range pH 5-7.5 before decreasing again as the conditions become more basic (see tables 4.1 and figure 4.3). Above pH 8, under the conditions of these experiments, thermal unfolding curves are strongly affected by exothermic aggregation concomitant with the unfolding transition of the protein, making the calculation of ΔH_m from traces obtained above this pH impossible. Aggregation in DSC thermograms also leads

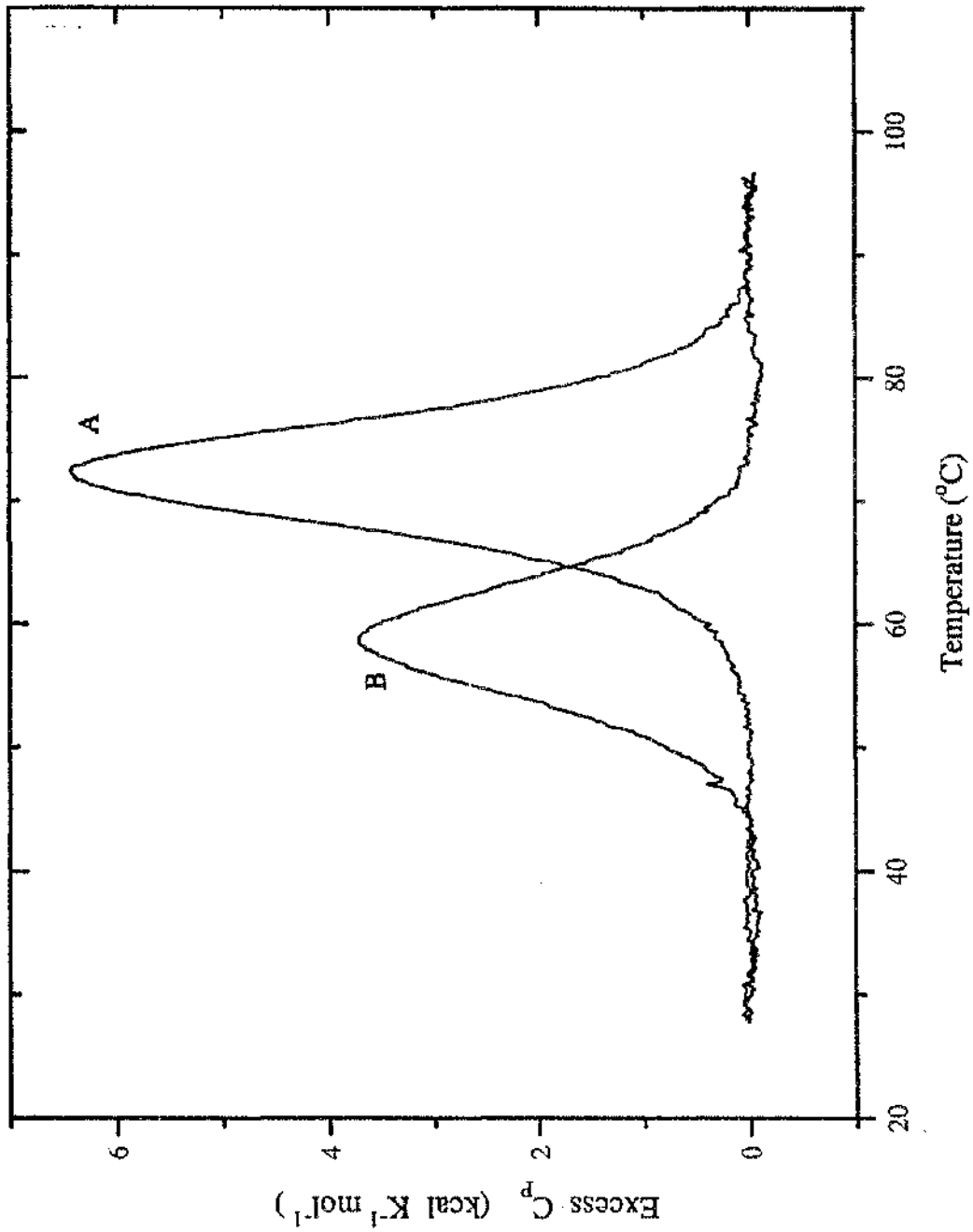

Figure 4.2 Normalised DSC traces for WT (A) and 3SS (B) α-lactalbumin obtained under the identical conditions of Tris buffer pH 7.4, 100mM NaCl, 100mM $CaCl_2$.

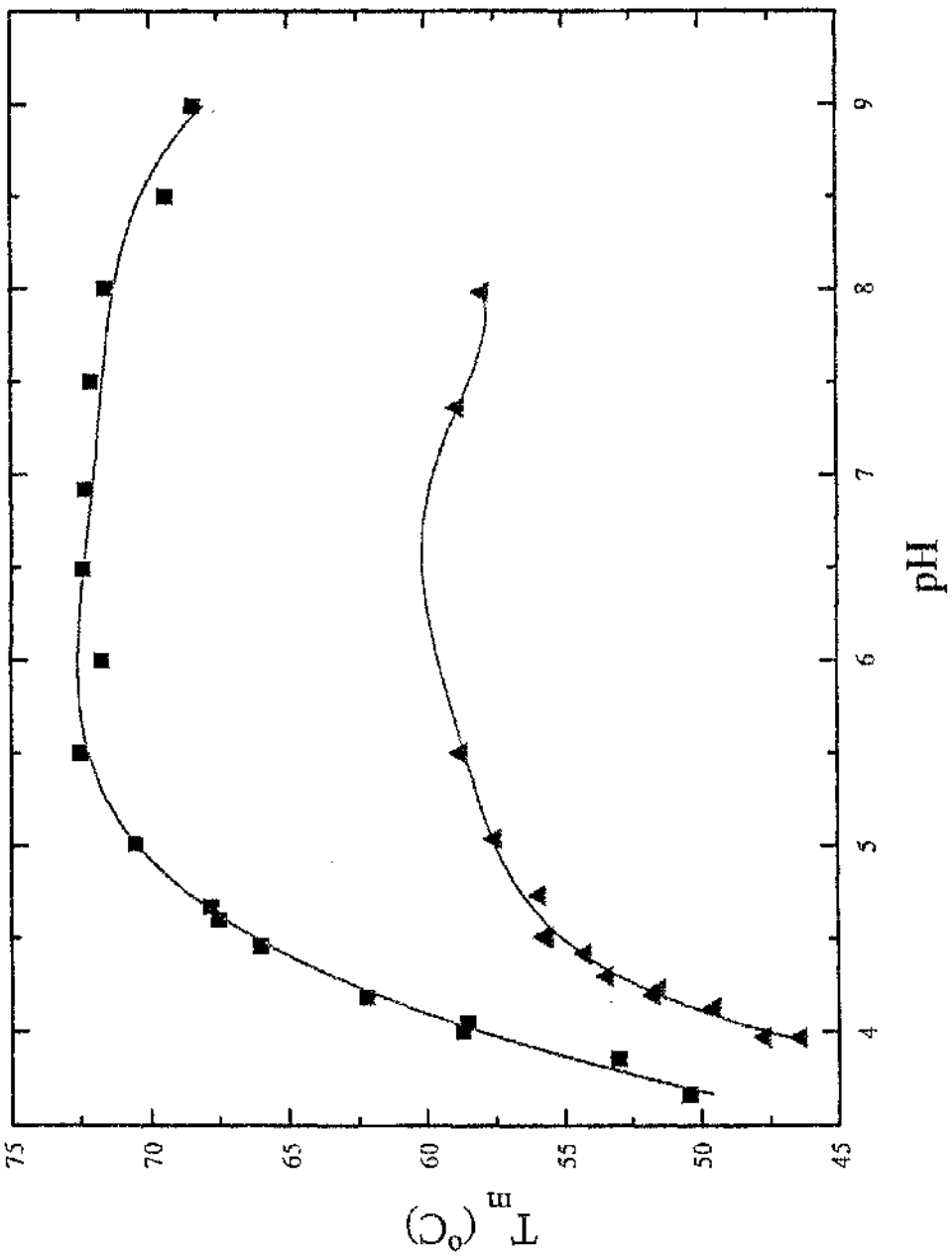

Figure 4.3 Plot of pH versus T_m for WT (■) and 3SS (▲) α-lactalbumin.

to artificial sharpening of the transition, resulting in difficulties when estimating the T_m from these traces. T_m's quoted from these are therefore approximate only.

WT- BLA		3SS- BLA	
pH	Tm (°C)	pH	Tm (°C)
3.67	50.4	3.97	46.4
3.86	53.0	3.97	47.7
4.00	58.7	4.13	49.5
4.05	58.5	4.20	51.7
4.19	62.2	4.23	51.5
4.46	66.0	4.30	53.4
4.60	67.5	4.42	54.2
4.67	67.8	4.51	55.6
5.01	70.5	4.51	55.7
5.50	72.5	4.73	55.9
6.00	71.7	5.04	57.5
6.49	72.4	5.50	58.7
6.92	72.3	7.36	58.8
7.50	72.1	7.98	57.9
8.00	71.6		
8.50	69.5*		
8.99	68.5*		

TABLE 4.1 Summary of results on the pH dependence of the T_m of WT and 3SS-BLA. *These T_m's are approximate.

All of these experiments were carried out in the presence of 0.1M calcium chloride and 0.1M sodium chloride. The buffers used were sodium acetate/acetic acid (pH 3.7 - 5.5), piperazine (pH 5.0 - 6.5) or TRIS/HCl (pH 7.0 - 9.0) and buffer salts were present at 0.1M concentrations in all cases.

This kind of pH dependence is typical of small globular proteins (Yang and Honig, 1993). The isoelectric point (pI, the pH where the protein has no net charge) of BLA is 5.6 which coincides well with the observed maximum in thermal stability as may be expected. Substantially similar results have been obtained for HEWL (Pfeil & Privalov, 1976; Cooper et al, 1992) and for the methionine repressor protein from E. coli, MetJ (Johnson et al, 1992).

With the pH held constant at 8.00 in Tris buffer, with 0.1M NaCl, the DSC thermograms of both WT-BLA and 3SS-BLA were recorded under a range of calcium chloride concentrations. The difference in T_m between WT and 3SS was 14 °C for all Ca^{2+} ion concentrations Table 4.2), in accordance with the differences noted in the pH effect experiments (section 4.3.2).

WT-BLA			3SS-BLA		
$[Ca^{2+}]$ (mM)	$\ln[Ca^{2+}]$	T_m (°C)	$[Ca^{2+}]$ (mM)	$\ln[Ca^{2+}]$	T_m (°C)
0.05	-3.00	63.7	0.1	-2.30	50.3
0.25	-1.39	65.3	0.7	-0.36	52.2
0.50	-0.69	66.2	1.0	0	53.5
2.50	0.92	68.8	7.0	1.95	55.9
5.00	1.61	69.5	10	2.30	56.4
12.5	2.53	70.2	100	4.61	57.5
25.0	3.22	70.7			
37.5	3.62	71.3			
100	4.61	70.7			
200	5.30	70.6			

TABLE 4.2 Summary of results on the calcium chloride concentration ($[Ca^{2+}]$) dependence of the T_m of WT and 3SS-BLA.

If it is assumed that the Ca^{2+} ion binds specifically to the folded form of the protein then upon unfolding the following equilibrium is set up:

$$NCa^{2+} \rightleftharpoons U + Ca^{2+} \qquad (4.1)$$

Where NCa^{2+} refers to BLA in the native state with calcium bound, U refers to the unfolded state and Ca^{2+} refers to free calcium ions. It is clear that increasing the excess concentration of calcium chloride in the buffer would result in shifting the equilibrium to favour the

reactants, and so would result in a stabilisation of the native state. Plotting T_m versus $\ln[Ca^{2+}]$ under these circumstances would result in a straight line. Since with a simple equilibrium such as described by equation 4.1 the equilibrium constant at T_m is simply given by $1/[Ca^{2+}]$ (since at T_m the concentration of protein in the N and U states is equal). Therefore plotting $\ln[Ca^{2+}]$ is equivalent to plotting -lnK, application of the van't Hoff equation predicts that, if the assumptions concerning the equilibrium are correct, a plot of -lnK versus 1/T should give a straight line (section 1.5). However when this plot is constructed (figure 4.4) the stabilisation experienced by the protein appears to reaches a maximum at around 50 mM. This indicates that the equilibrium is more complex than the simple assumption above suggests. It may be that the unfolded state of the protein is able to bind either the calcium or chloride ions, although much less tightly. It may be that such a situation is better described by the following equilibrium:

$$\begin{array}{ccccc} N.Ca^{2+} & \rightleftharpoons & U + Ca^{2+} & \rightleftharpoons & U.Ca^{2+} \\ \Downarrow & & & & \\ N + Ca^{2+} & & & & \end{array} \qquad (4.2)$$

where N and U refer to native and unfolded protein respectively and $N.Ca^{2+}$ and $U.Ca^{2+}$ refer to the calcium complex with those species. The apparent equlibrium constant, K_{app}, is given by:

$$K_{app} = \frac{[U]_{tot}}{[N]_{tot}} \qquad (4.3)$$

$$[U]_{tot} = [U] + [U.Ca^{2+}] \qquad (4.4)$$

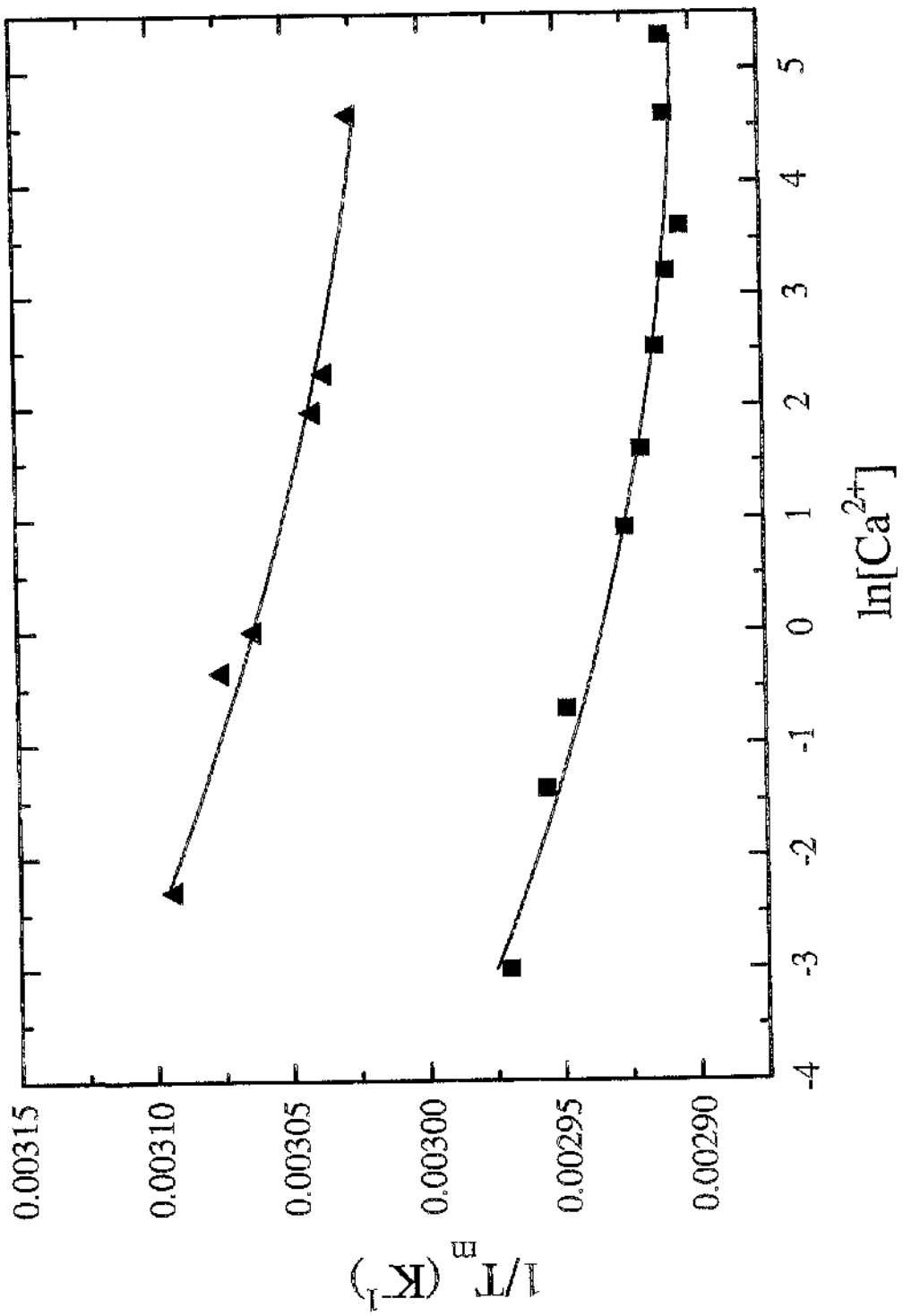

Figure 4.4 Plot of natural log of calcium ion concentration versus $1/T_m$ for WT (■) and 3SS (▲) α-lactalbumin.

$$[N]_{tot} = [N] + [N.Ca^{2+}] \tag{4.5}$$

rearangement of equation 4.3 in terms of the equilbrium constants given in equation 4.2 gives:

$$K_{app} = \frac{[U](1 + K_2[Ca^{2+}])}{[N](1 + K_1[Ca^{2+}])} \tag{4.6}$$

As has been observed before it is difficult to estimate unambiguously the change in heat capacity from a single DSC transition. A much more accurate estimate is obtained when the temperature dependence of the unfolding enthalpy is examined. Recall equation 2.4:

$$\Delta H_1 = \Delta H_0 + \Delta C_p(T_1 - T_0) \tag{2.4}$$

where ΔH_1 and ΔH_0 are the enthalpy changes at temperatures T_1 and T_0 respectively. If T_0 is chosen to be 273K and T_1 chosen to be equal to T_m equation 2.4 conveniently reduces to:

$$\Delta H_m = \Delta H_{273} + \Delta C_p T_m \tag{4.7}$$

where T_m is the transition mid-point temperature in centigrade and ΔH_{273} is the enthalpy change at 0 °C. Therefore given values for ΔH_m and T_m (such as those in tables 4.3 to 4.6) it is possible to calculate, by linear regression, not only ΔC_p but also the enthalpy at whatever reference temperature is required.

Plots of calorimetric enthalpy ($\Delta H_{m,cal}$) and van't Hoff enthalpy ($\Delta H_{m,vH}$) against T_m for WT-BLA give two essentially parallel straight lines (figure 4.5; a similar plot can be made for 3SS-BLA), which are, however, significantly offset to one another. This reflects the way minor errors in baselines tend to increase the calculated value of $\Delta H_{m,vH}$, by causing an artificial sharpening of the transition, and decrease the value of $\Delta H_{m,cal}$ by reducing the area under the transition, and vice versa. For this reason it was decided to employ a weighted average of the two ($\Delta H_{m,wa}$) in all calculations. Haynie and Freire (Haynie & Freire, 1994) recently introduced the following "weighted average enthalpy function"

$$\Delta H_{m,wa} = p\Delta H_{m,vH} + q\Delta H_{m,cal} \qquad (4.8)$$

This weighted average was intended to allow full analysis of so called "marginally stable proteins", where cold denaturation is a significant process observed in the experimental temperature range. However computer simulations in this laboratory (A. Cooper, unpublished results) in which the pre- and post-transition base lines were perturbed, showed that the weighted average given by Haynie and Freire was also able to correct successfully for the induced errors in the simulated data. The values for the weighting factors, p and q, given in that paper, and used here, are 0.65 and 0.35 respectively. Plots of $\Delta H_{m,wa}$ versus T_m give a straight line with much less experimental scatter than either $\Delta H_{m,vH}$ or $\Delta H_{m,cal}$ (figure 4.5). The following tables show the calculated weighted average enthalpy ($\Delta H_{m,wa}$) along with the directly measured calorimetric enthalpy ($\Delta H_{m,cal}$) and the van't Hoff enthalpy ($\Delta H_{m,vH}$) as described elsewhere (section 2.2.2).

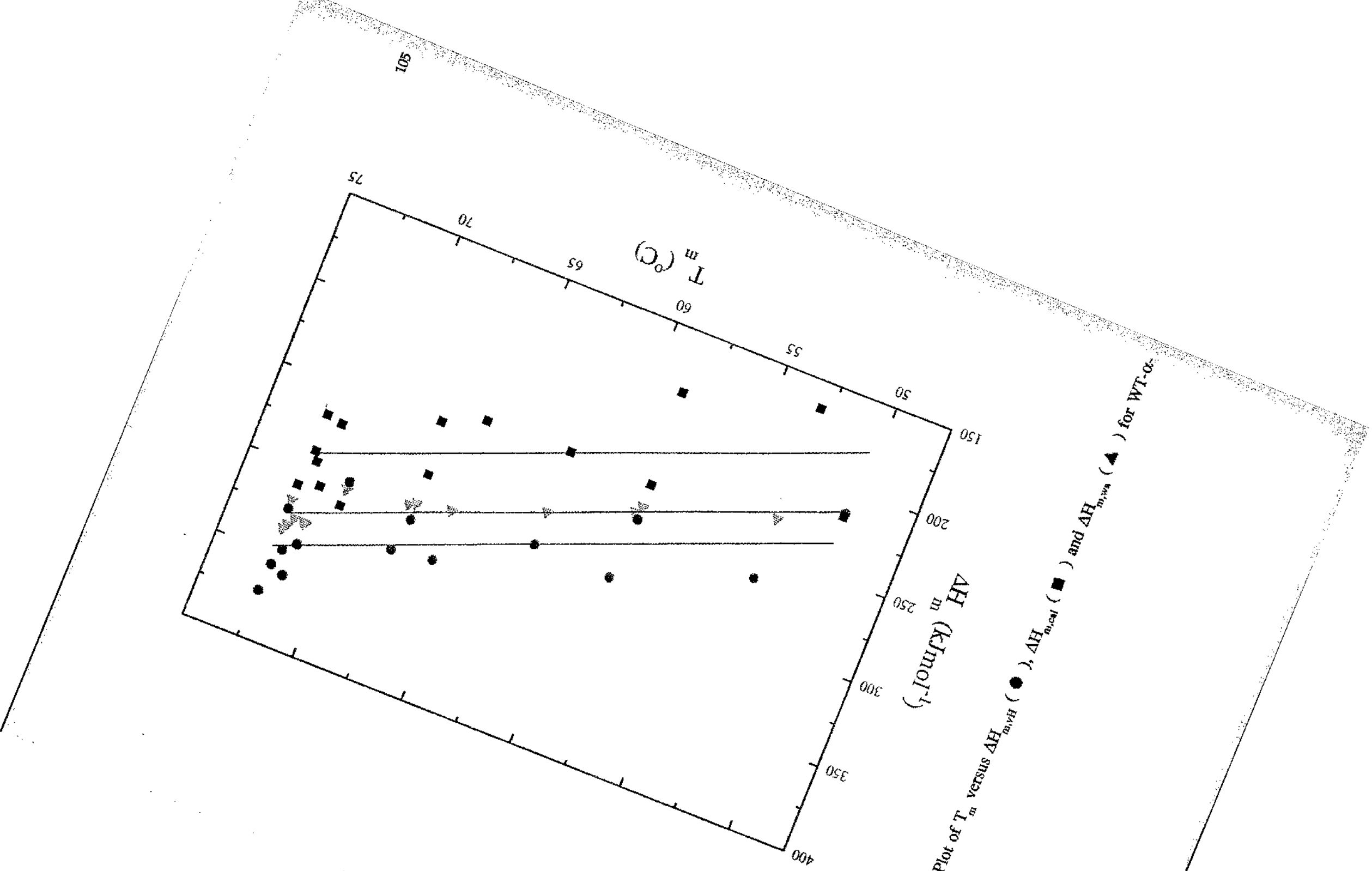

Figure 4.5 Plot of T_m versus $\Delta H_{m,vH}$ (●), $\Delta H_{m,cal}$ (■) and $\Delta H_{m,wa}$ (▲) for WT-α-lactalbumin.

T_m (°C)	$\Delta H_{m.cal}$ (kJmol^{-1})	$\Delta H_{m.vH}$ (kJmol^{-1})	$\Delta H_{m.wa}$ (kJmol^{-1})
63.7	268.5	282.9	277.9
65.3	279.2	317.8	304.3
66.2	299.4	321.1	313.5
68.8	310.2	338.7	328.7
69.5	306.8	332.5	323.5
70.2	309.9	340.1	329.5
70.7	320.3	335.0	329.8
71.3	327.4	333.6	331.4
70.7	289.9	358.5	334.5
70.6	292.7	355.7	333.6

TABLE 4.3 Summary of T_m and the various estimates of ΔH_m for WT-BLA from calcium ion concentration dependence data.

T_m (°C)	$\Delta H_{m.cal}$ (kJmol^{-1})	$\Delta H_{m.vH}$ (kJmol^{-1})	$\Delta H_{m.wa}$ (kJmol^{-1})
50.4	215.9	213.8	214.5
53.0	165.0	265.6	230.4
58.7	184.7	294.8	256.3
58.5	238.7	259.2	252.0
62.2	238.1	292.8	273.7
66.0	238.9	321.5	292.6
67.5	278.4	305.5	296.0
67.8	248.5	324.6	298.0
70.5	312.5	298.4	303.3
72.5	310.4	324.6	319.6
71.7	270.0	360.0	328.5
72.4	268.0	372.5	335.9
72.3	289.3	356.5	333.0
72.1	294.6	346.8	328.6
71.6	306.8	341.0	329.0

TABLE 4.4 Summary of T_m and the various estimates of ΔH_m for WT-BLA from pH dependence data.

T_m (°C)	$\Delta H_{m.cal}$ (kJmol^{-1})	$\Delta H_{m.vH}$ (kJmol^{-1})	$\Delta H_{m.ws}$ (kJmol^{-1})
50.3	180.2	233.5	214.8
52.2	195.2	228.7	217.0
53.5	230.3	251.7	244.2
55.9	204.3	256.8	238.4
56.4	230.3	268.2	255.0
57.5	236.5	275.9	262.1

TABLE 4.5 Summary of T_m and the various estimates of ΔH_m for 3SS-BLA from calcium ion concentration dependence data.

T_m (°C)	$\Delta H_{m.cal}$ ($kJmol^{-1}$)	$\Delta H_{m.vH}$ ($kJmol^{-1}$)	$\Delta H_{m.wa}$ ($kJmol^{-1}$)
46.4	197.5	163.3	175.3
47.7	176.5	183.8	181.3
49.5	132.4	214.2	185.6
51.7	108.1	263.1	208.8
51.5	94.2	269.7	208.3
53.4	135.0	263.5	218.5
54.2	146.6	255.4	216.5
55.6	169.3	258.6	227.3
55.7	170.0	234.2	211.7
55.9	171.5	239.3	215.6
57.5	205.7	233.5	223.8
58.7	209.9	247.6	234.4
58.8	202.6	289.6	259.2
57.9	207.0	284.3	257.3

TABLE 4.6 Summary of T_m and the various estimates of ΔH_m for 3SS-BLA from pH dependence data.

Plotting ΔH_m versus Tm for WT- and 3SS-BLA gives two straight lines which appear to be co-linear within experimental error, indicating a common ΔC_p. Remember it was also found for HEWL that the removal of one disulfide bridge did not affect the observed ΔC_p (Cooper et al, 1992). Linear regression carried out on the plots for WT- and

3SS-BLA from both sets of experiments (figure 4.6) allows calculation of the common ΔC_p for the two forms of the protein. The value obtained in this way is $\Delta C_p = 6.0 \pm 0.2\ kJK^{-1}mol^{-1}$. This value compares reasonably well with the previously reported values of $\Delta C_p = 7.95\ kJK^{-1}mol^{-1}$ (Xie et al, 1991) and $\Delta C_p = 6.28\ kJK^{-1}mol^{-1}$ (Pfeil & Sadowski, 1985).

Given the values of T_m, ΔH_m and ΔC_p it is possible to calculate the standard free energy change at 298K for the unfolding reaction according to equation 2.11 which we encountered earlier:

$$\Delta G_{298} = \Delta H_m\left(1-\frac{298}{T_m}\right)+\Delta C_p\left(298-T_m-298\ln\frac{298}{T_m}\right) \qquad (2.11)$$

This has been done for the data for both forms of the protein and the results are shown in tables 4.7 and 4.8

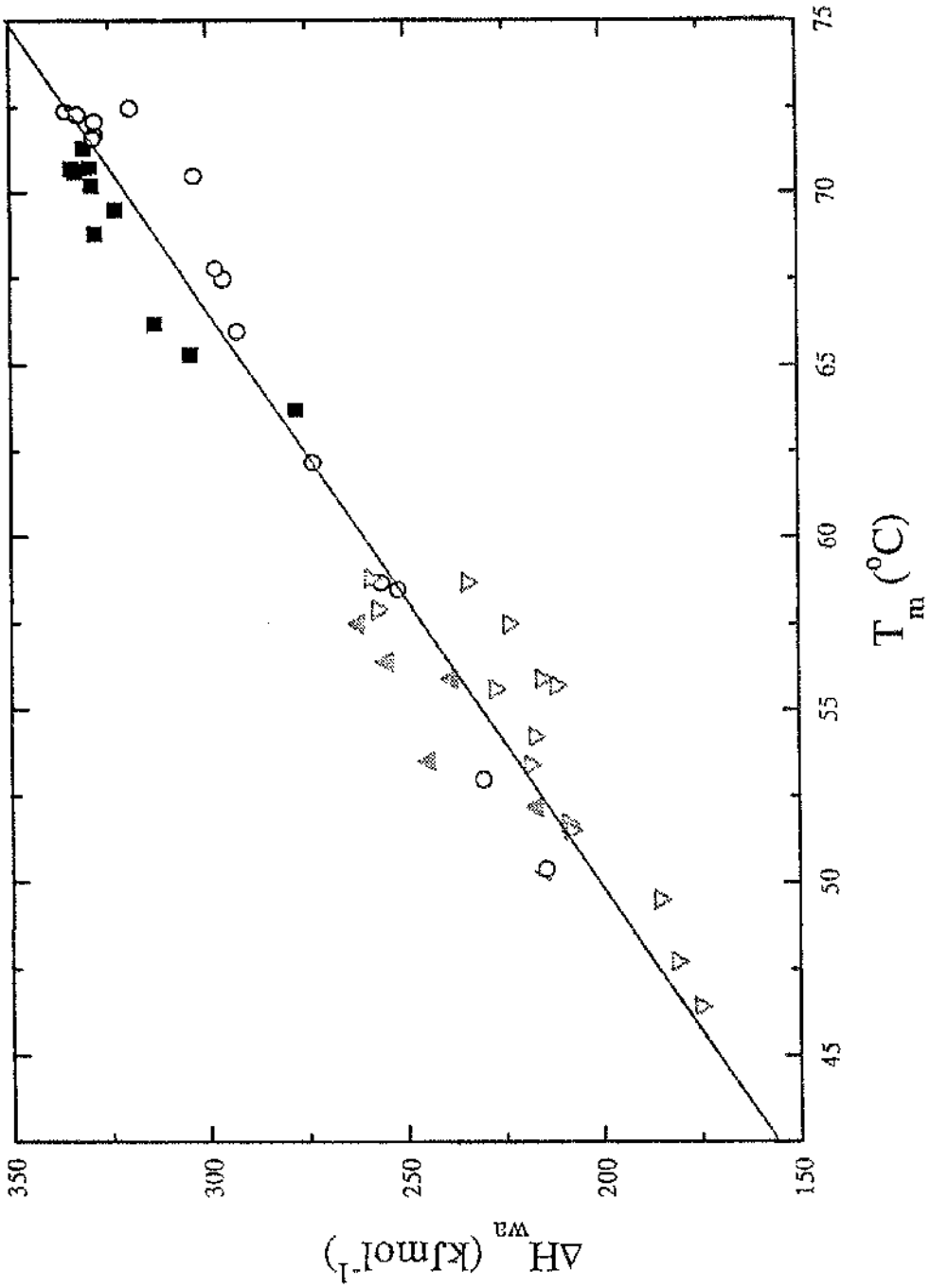

Figure 4.6 Plot of T_m versus weighted average ΔH for WT and 3SS-α-lactalbumin from both calcium (WT (■) & 3SS (▲)) and pH experiments (WT (○) & 3SS (▽)).

WT-BLA			3SS-BLA		
ΔH_m (kJmol^{-1})	T_m (°C)	ΔG_{298} (kJmol^{-1})	ΔH_m (kJmol^{-1})	T_m (°C)	ΔG_{298} (kJmol^{-1})
214.5	50.4	10.70	175.3	46.4	7.34
230.4	53.0	12.36	181.3	47.7	7.89
256.3	58.7	15.40	185.6	49.5	8.37
252.0	58.5	14.95	208.8	51.7	10.40
273.7	62.2	17.50	208.3	51.5	10.33
292.6	66.0	19.88	218.5	53.4	11.38
296.0	67.5	20.33	216.5	54.2	11.26
298.0	67.8	20.58	227.3	55.6	12.34
303.3	70.5	21.24	211.7	55.7	10.89
319.6	72.5	23.38	215.6	55.9	11.26
328.5	71.7	24.60	223.8	57.5	12.09
335.9	72.4	25.62	234.4	58.7	13.18
333.0	72.3	25.22	259.2	58.8	15.70
328.6	72.1	24.61	257.3	57.9	15.42
329.0	71.6	24.68			

TABLE 4.7 Summary of standard free energies of unfolding for WT- and 3SS-BLA derived from the pH dependence data.

WT-BLA			3SS-BLA		
ΔH_{m} (kJmol^{-1})	T_m (°C)	ΔG_{298} (kJmol^{-1})	ΔH_{m} (kJmol^{-1})	T_m (°C)	ΔG_{298} (kJmol^{-1})
277.9	63.7	18.05	214.8	50.3	10.71
304.3	65.3	21.24	217.0	52.2	11.13
313.5	66.2	22.42	244.2	53.5	13.63
328.7	68.8	24.52	238.4	55.9	13.40
323.5	69.5	23.89	255.0	56.4	15.03
329.5	70.2	24.70	262.1	57.5	15.86
329.8	70.7	24.76			
331.4	71.3	24.99			
334.5	70.7	25.38			
333.6	70.6	25.26			

TABLE 4.8 Summary of standard free energies of unfolding for WT- and 3SS-BLA derived from the calcium ion concentration dependence data.

A plot showing the variation of the standard free energy (ΔG^{o}_{298}) with pH for both forms of the protein is given in figure 4.7. From this it is clear that the difference in the standard free energy change is $\Delta\Delta G^{o} = 9.6 \pm 1.2$ kJmol^{-1}. A similar plot is shown in figure 4.8 in which standard free energy change is plotted as a function of the natural log of the calcium ion concentration. This shows the comparable value found from the calcium ion concentration experiments of $\Delta\Delta G^{o} = 9.9 \pm 0.3$ kJmol^{-1}.

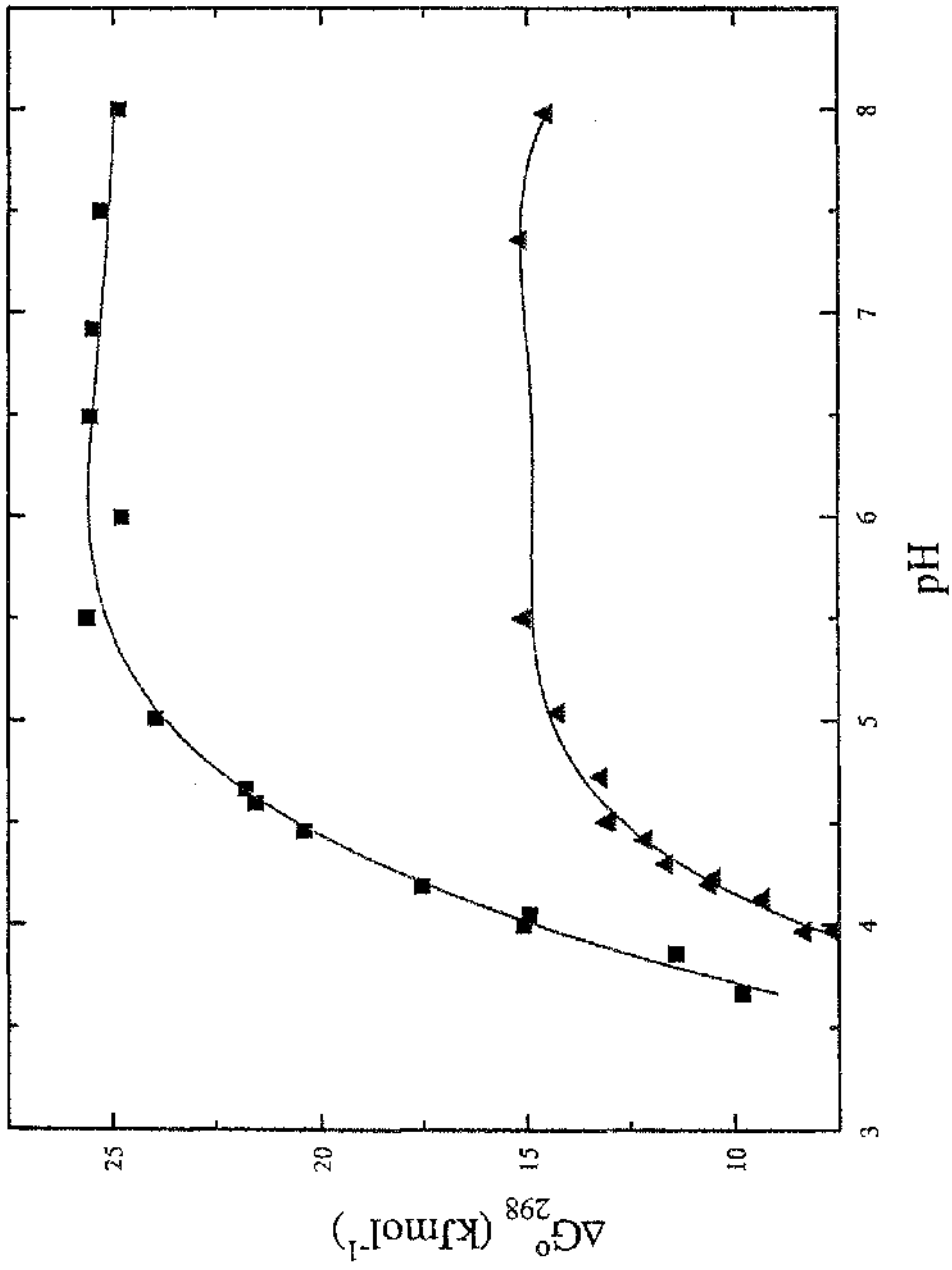

Figure 4.7 Plot of pH versus free energy for WT (■) and 3SS (▲) α-lactalbumin.

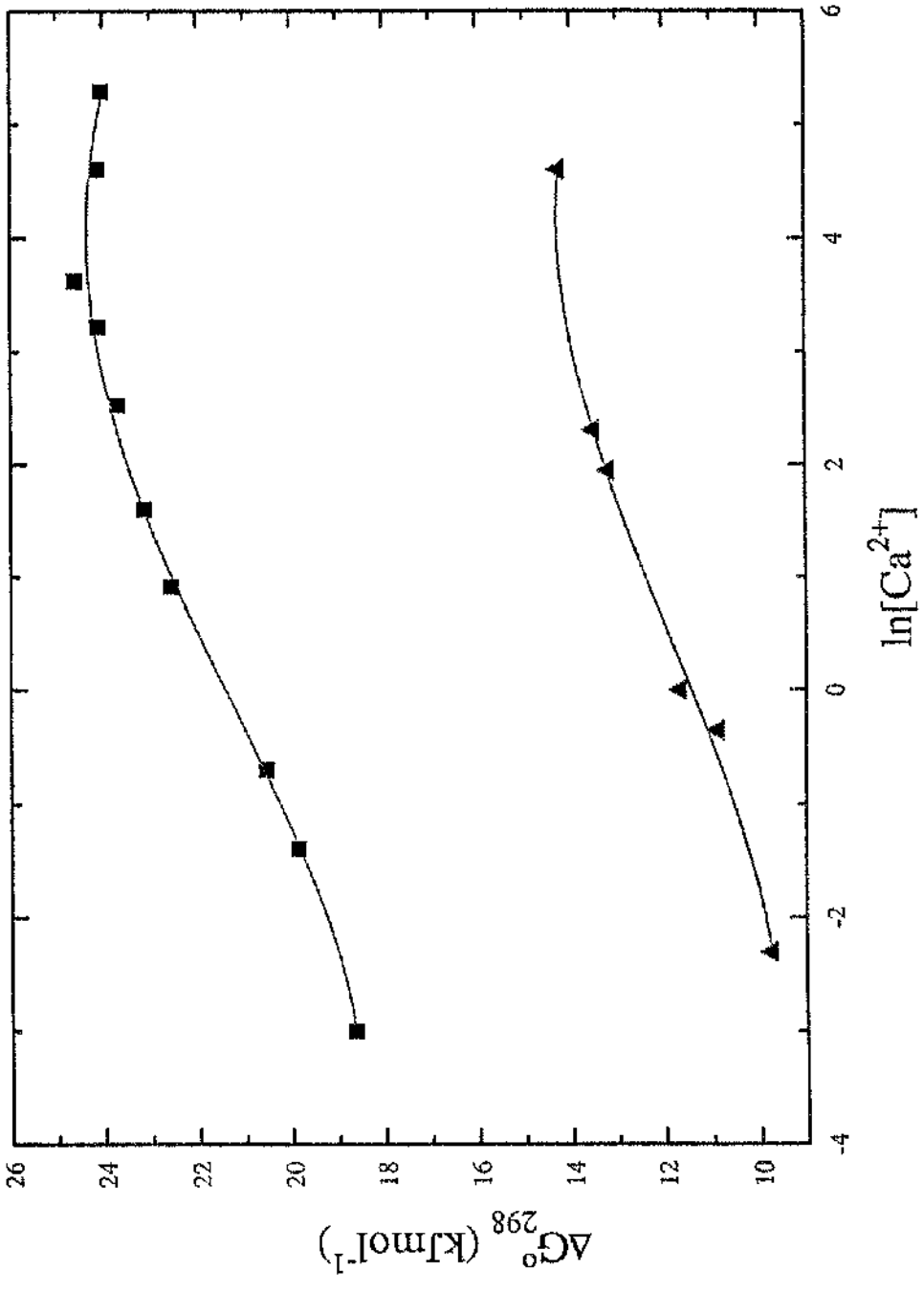

Figure 4.8 plot of natural log of calcium ion concentration versus free energy for WT (■) and 3SS (▲) α-lactalbumin.

Since the plots of ΔH_m versus T_m for WT- and 3SS-BLA are co-linear it is obvious that $\Delta\Delta H_m$ is zero, and that in fact the stabilisation results from an entirely entropic contribution. This allows us to calculate that $\Delta\Delta S = 32.2 \pm 4$ $JK^{-1}mol^{-1}$ from the pH dependence data. Similarly, from the calcium ion concentration data we can calculate a value of $\Delta\Delta S = 33.2 \pm 1$ $JK^{-1}mol^{-1}$. The significance of these results will be discussed in section 4.4.

4.3 Fluorescence spectroscopy

BLA contains four Trp residues. It is therefore possible to use these as fluorophores in intrinsic protein fluorescence and fluorescence quenching measurements. These give us a useful insight into the structure and dynamics of the protein.

The relative exposure of the Trp residues in a protein to the solvent will play a significant role in determining the wavelength of the emitted light in an intrinsic fluorescence measurement. This means that it is possible to use fluorescence measurements to determine rapidly whether or not a protein (e.g. BLA) adopts the native or molten globule state. The MG state has an expanded structure resulting in more penetration of the solvent, and therefore greater exposure of the Trp residues to water. this in turn results in the λ_{max} of the protein being red shifted. Lala and Kaul (1992) found that λ_{max} shifted from 328 nm in the N state to 343 nm in the MG state.

Intrinsic fluorescence measurements (figure 4.9 shows typical fluorescence spectra) were made in order to be sure that in the experiments above all of the thermally induced transitions were from the N to D states, with no interference from the MG state. It was found that the λ_{max} did not depend on the pH or calcium ion concentration employed either for WT- or 3SS-BLA and that hence these did not, by themselves, induce any gross structural changes in the natiove state. The λ_{max} observed for WT-BLA was 330 nm and that for 3SS-BLA was 334 nm, indicating that neither of these adopted the MG state under the conditions employed. The small difference between WT- and 3SS- forms of the protein possibly indicates that the protein lacking the crosslink has a slightly more open structure allowing further penetration of the solvent into the hydrophobic core of the protein.

At neutral pH complete removal of the bound calcium ion, by dialysis with buffer containing EDTA, resulted in a significant red shift at room temperature. This is in accordance with the reported observation that the apo- form of the protein adopts the MG state (Hiraoka & Sugai, 1984; Ikeguchi et al, 1986; Yutani et al, 1992).

The relative dynamics and solvent accessibilty of the two forms of the protein were examined using the technique of fluorescence quenching., with succinimide as quenching agent (Eftink & Ghiron, 1976; Eftink & Ghiron, 1984). Succinimide is fairly large and so diffuses into the protein interior more slowly than other quenchers (e.g. acrylamide), and its internal filtering effect is negligible. These factors make succinimide ideal as a probe for the dynamics of the BLAs.

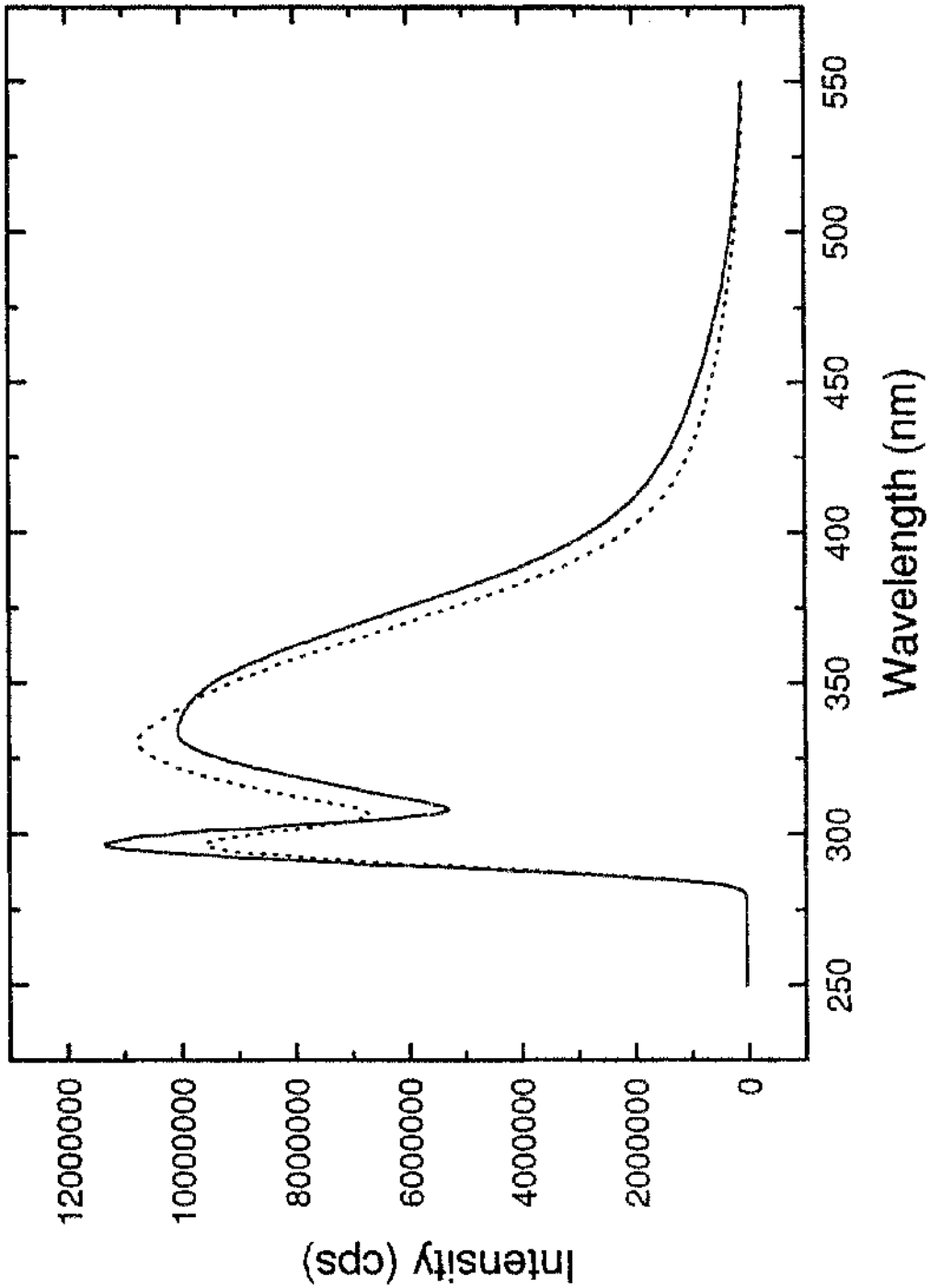

Figure 4.9 Emmission fluorescence spectra of WT (dotted) and 3SS (solid) α-lactalbumin (in acetate buffer pH 3.7, 100mM NaCl, 150mM $CaCl_2$.

As the protein is titrated with quencher the fluorescence intensity drops (figure 4.10). Care was taken to ensure that the protein concentration remained constant throughout and to eliminate any extraneous factors which may cause either a drop or an increase in intensity with addition of the quencher, such as temperature fluctuations or impurities in the succinimide (see section 3.5.2) thus ensuring that the observed intensity changes result purely from succinimide quenching of the Trp fluorescence.

BLA contains four tryptophan residues, each of which in principle may have different quantum yields, and different quenching constants. The result is non-linear Stern-Volmer plots (Figure 4.11). Because of this non-linearity the analysis of the data is rather complicated. However it is plain from simple observation of the quenching results (figure 4.10) that the Trp residues in 3SS- protein were quenched to a greater extent than the WT-form. In other words the results presented in figure 4.11 indicate that 3SS-BLA is significantly more dynamically flexible than WT-BLA.

In order to attempt to put this observation on a more quantitative footing the data were analysed in two ways. In one the curves were simply approximated by two straight lines (figure 4.11). At low quencher concentration, where quenching of the most easily accessible tryptophans is taking place, both 3SS- and WT-BLA appear to be quenched to the same extent ($K_{A,3SS} = 1.10 \pm 0.09$; $K_{A,WT} = 1.14 \pm 0.12$), however above 181 mM succinimide, where quenching of the more deeply buried tryptophan residues is taking place, the extent of quenching in 3SS-BLA is much greater than in WT-BLA ($K_{B3SS} = 0.613 \pm 0.002$; $K_{B,WT} = 0.402 \pm 0.001$), suggesting that at least part of the molecule is experiencing much larger and/or

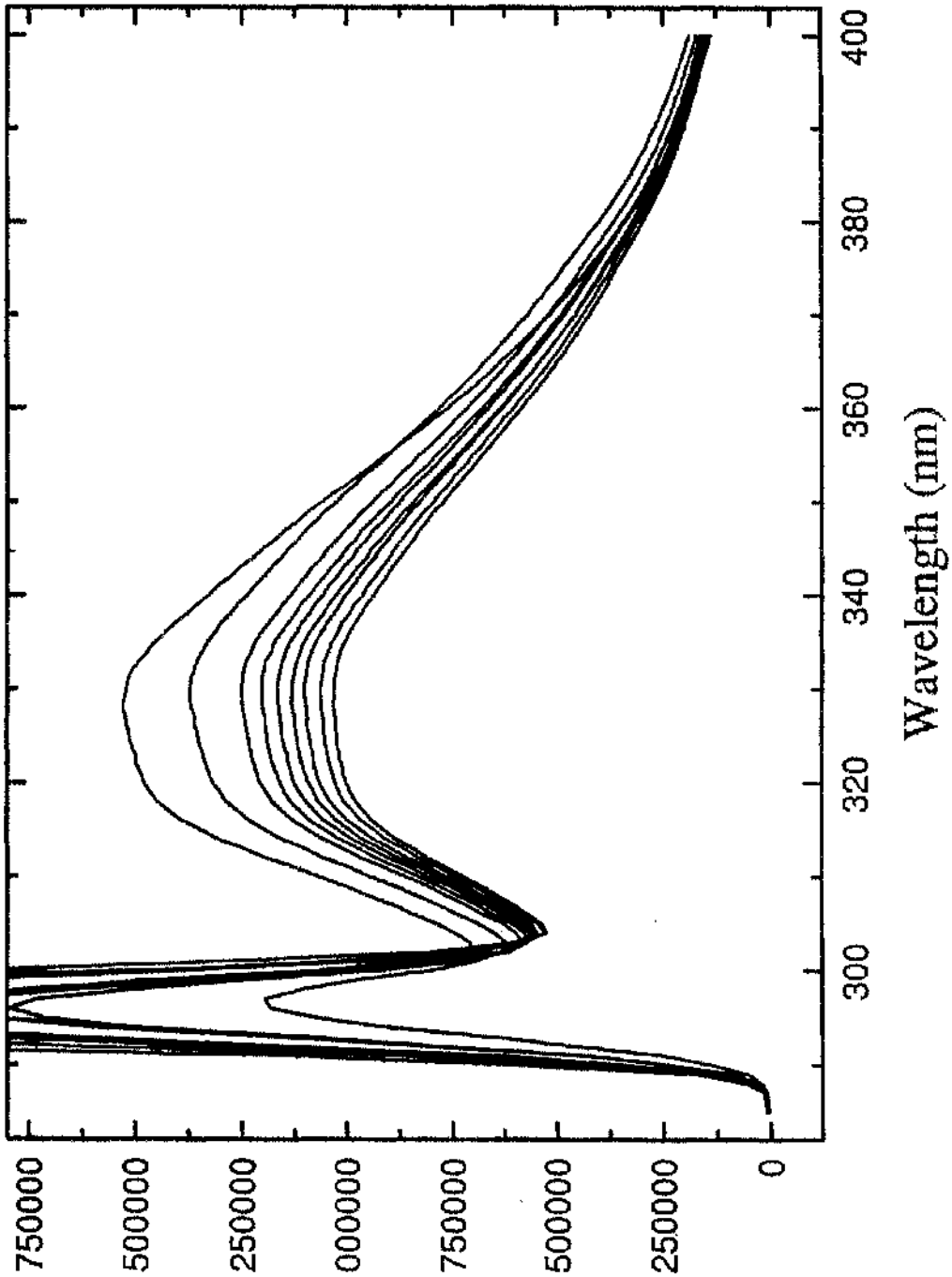

Figure 4.10 Series of emmission specra obtained for WT-α-lactalbumin with succinimide concentration increasing as you go down the series (conditions as before, water Raman band subtracted).

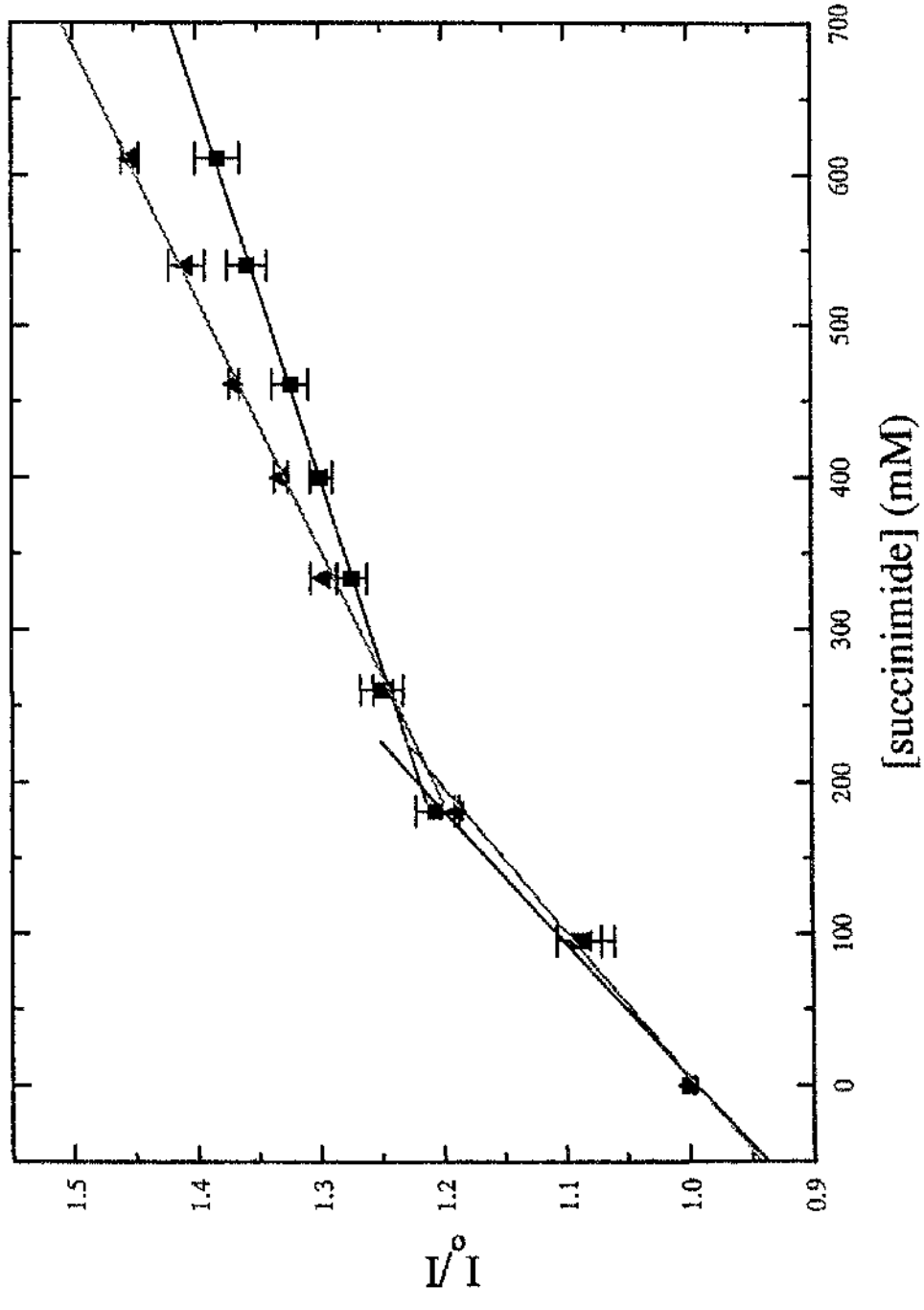

Figure 4.11 Stern-Volmer plot of mean data for WT (■) and 3SS (▲) α-lactalbumin.

more rapid thermal fluctuations. Indeed the quenching of the least accessible portion of the molecule appears to proceed 50% faster in the modified form of the protein.

Data were also analysed by making use of a modification of the Stern-Volmer approach to fluorescence quenching in which it was assumed that the four Trp residues could be able to be split into two populations with different, non-zero quenching constants. This is comparable to the modification described elsewhere (section 1.6.3; Lehrer, 1971) for two populations one with zero quenching constant. The following model was constructed on this basis and fitted to the data (figure 4.12).

$$\frac{I_0}{I} = \frac{(1+K_A[Q])(1+K_B[Q])}{1+(K_B + f_B(K_A - K_B))[Q]} \qquad (3.1)$$

Where [Q] is the quencher concentration, K_A and K_B are the quenching constants for population A and B respectively and f_B is the fraction of fluorescence intensity deriving from population B. The starting point in the derivation of the model was the assumption that all of the observed fluorescence intensity resulted from two Trp populations i.e.:

$$I_0 = I_{0,A} + I_{o,B} \qquad (4.9)$$

where I_o is the initial intensity (before addition of any quencher) and $I_{o,A}$ and $I_{o,B}$ are the intensities from population A and B respectively. While the intensity in the presence of quencher is simply:

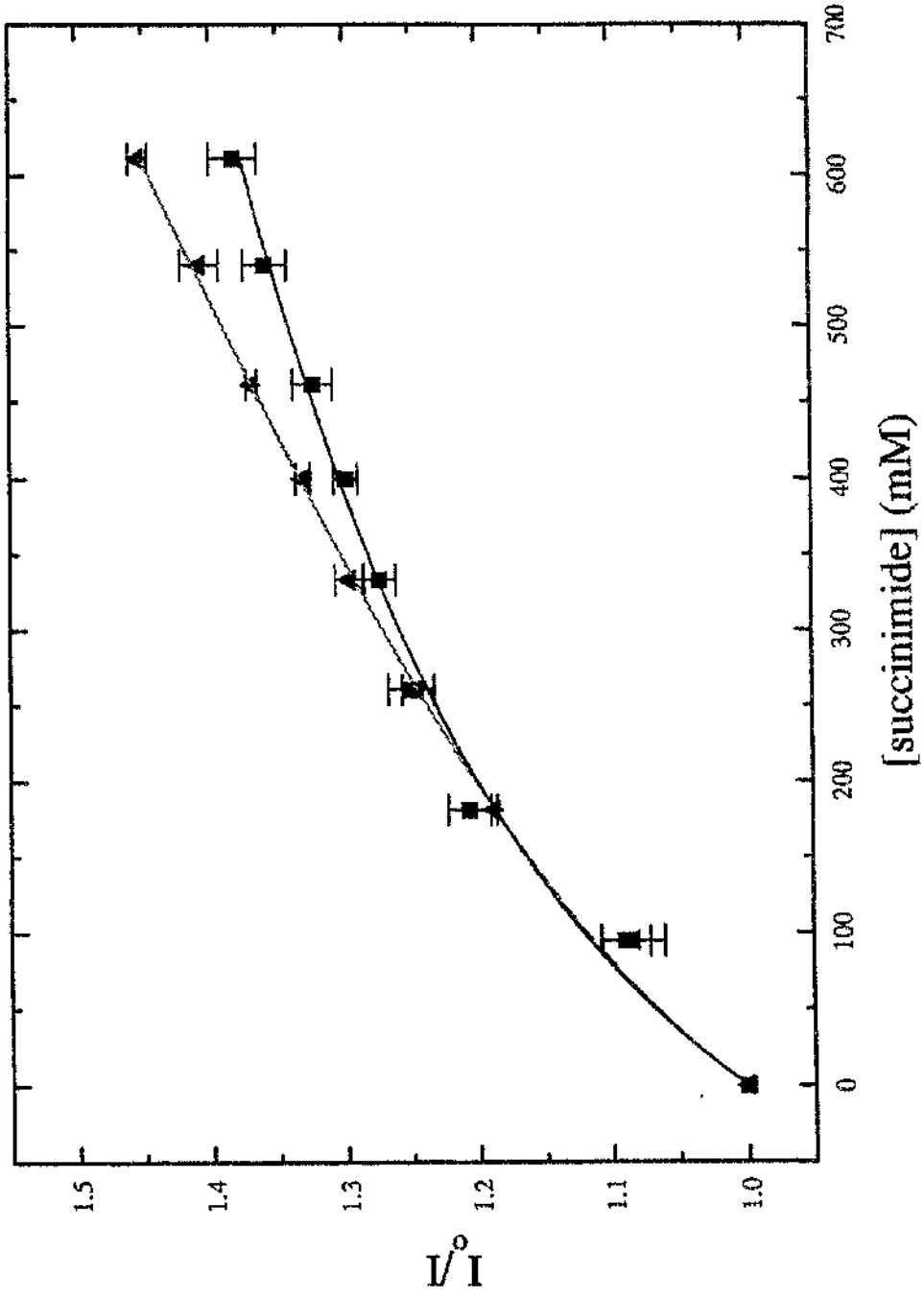

Figure 4.12 Stern-Volmer plot of mean data for WT (■) and 3SS (▲) α-lactalbumin with the model given by equation 3.1 fitted.

$$I = I_A + I_B \tag{4.10}$$

Recall equation 1.33:

$$\frac{I_o}{I} = 1 + K[Q] \tag{1.33}$$

Combination of equations 4.10 and 1.33 gives:

$$\frac{I}{I_0} = \frac{f_A}{1 + K_A[Q]} + \frac{f_B}{1 + K_B[Q]} \tag{4.11}$$

where f_A and f_B are the fractions of fluorescence intensity deriving from populations A and B respectively are defined by:

$$f_A = \frac{I_{0,A}}{I_{0,A} + I_{0,B}} \tag{4.12}$$

$$f_B = \frac{I_{0,B}}{I_{o,A} + I_{0,B}} = 1 - f_A \tag{4.13}$$

simplification and rearrangement of equation 4.11 gives the model, equation 3.1.

The model gave the following values when the fraction f_B was constrained to 0.75:

	K_A	K_B
3SS	1.83	0.251
WT	1.80	0.133

TABLE 4.9 Summary of model fitting parameters with $f_B = 0.75$.

Obviously this model is a gross simplification, and in fact it should be expanded to take into account all four tryptophans, however this would result in a model with too many variables to be of any real use.

Fluorescence quenching of WT-BLA (at 25 °C) was measured under different calcium ion concentrations. Under these conditions the T_m of the protein varied through the range 63-71 °C. Despite the changing T_m no effect was found on the Stern-Volmer quenching rate.

It is generally accepted that the mode of quenching by succinimide (and other quenching agents) is through global breathing movements and not through local or global unfolding of the protein molecule. Since the population unfolded at a given time and temperature should be seen to increase with decreasing T_m this mode of action would allow a greater rate of quenching under conditions when the T_m was low. However it was observed that the rate of quenching did not vary with the T_m of the protein. This observation adds credence to the accepted interpretation, showing global breathing movements to be the dominant mode of quenching.

4.4 Discussion of BLA results

It has previously been reported (Ikeguchi et al, 1992) that the difference in free energy of unfolding between WT- and 3SS-BLA is 13.0 ± 1.3 $kJmol^{-1}$ at 4°C, in the presence of 1mM Ca^{2+}. This is comparable to the mean value for $\Delta\Delta G^{o}$ estimated here of 9.7 ± 0.6 $kJmol^{-1}$.

The theoretical predictions of the increase in conformational entropy, and hence free energy of unfolding, based on equation 1.2, when the loop size is 115 amino acid residues are as follows. Schellman (1955) predicts $\Delta\Delta S_{298}$ = -92.2 $JK^{-1}mol^{-1}$ resulting in $\Delta\Delta G^{o}_{298}$=27.5 $kJmol^{-1}$, Poland and Scheraga (1965) predict $\Delta\Delta S_{298}$ = -88.0 $JK^{-1}mol^{-1}$ with $\Delta\Delta G^{o}_{298}$ = 26.2 $kJmol^{-1}$ and Pace et al (1988) predict $\Delta\Delta S_{298}$ = -68.0 $JK^{-1}mol^{-1}$ and $\Delta\Delta G^{o}_{298}$= 20.3 $kJmol^{-1}$. However the mean of the values measured for the removal of the 6-120 disulfide bond in BLA, $\Delta\Delta S_{298}$ = 32.7 $JK^{-1}mol^{-1}$, was significantly lower than even the lowest estimate based on the theory.

As observed the differences in the observed entropy change (ΔΔS) are measured relative to the absolute entropy of the native state, and for convenience the absolute entropy of the native state is generally assumed to be unaffected by the removal of the disulfide bridge. However it has been shown that this assumption may not be valid in all circumstances, and that the entropy of the native state may, sometimes, have to be considered explicitly (Karplus et al, 1987).

It is clear from the fluorescence quenching data that the removal of the disulfide bridge between residues 6-120 has an effect on the native state, resulting in the native state of the modified protein being significantly more dynamically flexible than the native state of the disulfide intact protein, consequently the native state of 3SS-BLA cannot be assumed to have the same absolute entropy as the native state of WT-BLA. In this case, $\Delta\Delta S$ would be smaller than predicted by any of the theory, at least in part because the entropy of the native state of 3SS-BLA is greater than that of the native state of disulfide intact BLA.

As discussed earlier, one of the underlying assumptions in the Poland & Scheraga type argument is that the change in unfolding entropy ($\Delta\Delta S$) resulting from removal of the crosslink arises solely from entropic effects upon the unfolded state. We propose here that this assumption is not valid in every case. Whilst in the case of the 6-120 disulfide bride in HEWL the assumption appears to hold true, the situation is quite different when the case of the 6-120 disulfide bridge in BLA is considered. Application of the relevant equation (1.2) results in a significant over estimation of $\Delta\Delta S$. We further propose, on the basis of the fluorescence quenching data presented that this is due to an increase in the dynamic nature of the native state.

"Thou art beside thyself; much learning doth make thee mad."

The Bible, The Acts of the Apostles.

Chapter Five

Solvent effects on lysozyme folding stability

5.1 Introduction

Lysozyme is probably one of the best known, and most widely studied proteins. In particular the ready availability of the lysozymes and their relative simplicity has made them an extremely common subject for all kinds of studies of the properties of proteins in general, for example the effects of metal ion binding (Kuroki et al, 1992b & 1992c), of pH (Yang & Honig, 1993), of disulphide bridges (Cooper et al, 1992; Eyles et al, 1994; Kidera et al, 1994) and of protein-protein aggregation (Persson & Gekas, 1994).

Hen egg white lysozyme, HEWL (figure 5.1) consists of 129 amino acid residues (with a molecular weight of 14,500), among which are 8 Cysteine (Cys or C) residues, giving rise to 4 disulphide bridges. Some members of the lysozyme family have a calcium binding function (e.g. equine lysozyme: Nitta et al, 1993), which is not present in HEWL.

In common with the other members of the lysozyme family, the function of HEWL is bacteriolysis (Fleming, 1922), and is classified as a 1,4-β-N-acetylmuramidase. It

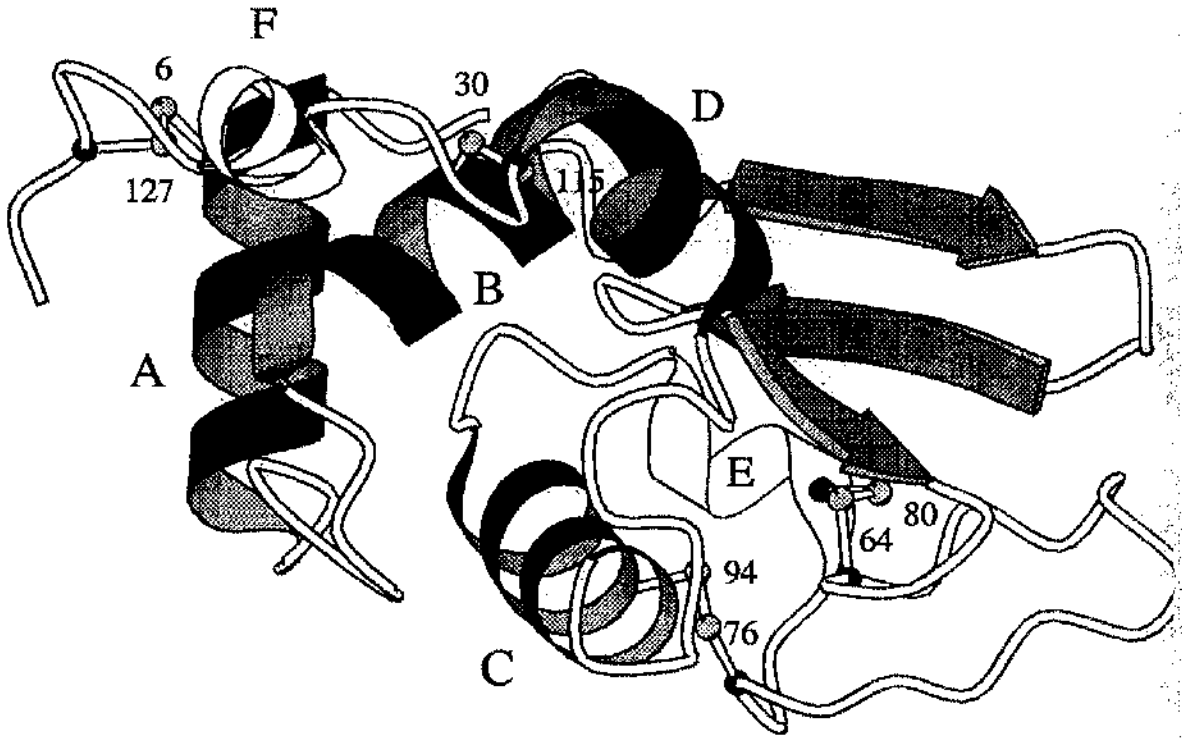

Figure 5.1 Three dimensional structure of hen egg white lysozyme.

catalyses cleavage of the glycosidic linkage between the C-1 of N-acetylmuramic acid and C-4 of N-acetylglucosamine in bacterial peptidoglycan (McKenzie & White, 1991).

Much of the work on HEWL and the lysozyme family in general has focused on the comparisons with the alpha-lactalbumin family (Kuwajima et al, 1985; McKenzie & White, 1991). The two families have similarities in structure, but differ in function, although they are believed to have evolved from a common ancestor (Brew et al, 1967; Browne et al, 1969). However despite the vast amount of work which has been carried out on HEWL there is still no real understanding of relative importance of the various interactions responsible for protein stability. Studies which begin to dissect these forces in the case of HEWL would give a valuable starting point for the formation of general principles as well as give specific information on HEWL itself.

Isotope exchange is one of the most subtle perturbations one can make to the system, but the effect is widespread. Hydrogen bonding and the hydrophobic effects depend heavily on the solvent, even electrostatic effects are influenced through the dielectric constant of the solvent. When proteins are prepared in heavy water, D_2O, most of the available, exchangeable hydrogens are replaced with deuterium within a few millionths of a second (Baldwin, 1993), with little apparent effect on structure (Finer-Moore et al, 1992; Kossiakoff et al, 1992; Connelly et al, 1993).

Much of the work carried out so far focuses on model compounds, their relative solubilities in light and heavy water and their heats of transfer from H_2O to D_2O. Nemethy

& Scheraga (1964) found that the O-D...O bond was stronger in D_2O than the equivalent O-H..O bond in H_2O. Kresheck et al (1965) studied the solubilities of the alkanes propane and butane in heavy and light water and found that both were more soluble in D_2O. They then went on to use this information to calculate unitary free energy, enthalpy and entropy changes, i.e. the thermodynamic parameters resulting only from solute-solvent interactions. They found that both ΔH_u and ΔS_u were more negative in D_2O than in H_2O. They also calculated the free energy of transfer from H_2O to D_2O from the unitary free energies in the following way:

$$\Delta G_t^o = \left(\Delta G_u^o\right)_{D_2O} - \left(\Delta G_u^o\right)_{H_2O} \qquad (5.1)$$

They found that ΔG_t° was small and negative (approximately -165 $Jmol^{-1}$ at 4°C) indicating that the hydrophobic interaction between pairs of these molecules was stronger in H_2O. As with all of the parameters measured or calculated ΔG_t° was found to be temperature dependent and dropped to near zero at 50°C (ΔG_t° = -10± 42 $Jmol^{-1}$ at 50°C).

Kresheck et al (1965) also studied the heats of solution of various amino acids, and found that, contrary to the hydrocarbon models, the free energy of transfer was positive and therefore concluded that the hydrophobic interactions between pairs of amino acid side chains was stronger in D_2O, although they point out that the efects of the polar portion of these molecules must be taken into account. In summary Kresheck et al (1965) conclude that "if the free energy of transfer of non-polar amino acid side chains were to become more positive with increasing temperature, like that for hydrocarbons... an immediate consequence would be that proteins would be more stable to thermal modifications in D_2O".

Ben-Naim et al (1973) again employed studies of solubility to determine the thermodynamic parameters relating to transfer of model compounds from H_2O to D_2O. In that study it was found that the hydrophobic interaction between ethane-ethane and methane-methane pairs was stronger in H_2O (in broad agreement with the study of Kresheck et al). In contrast the hydrophobic interaction between benzene-benzene pairs appeared to be stronger in D_2O.

Connelly et al (1993) used ultrasensitive isothermal titration calorimetry to study the binding of two largely non-polar ligands tacrolimus and rapamycin to the protein FK506 binding protein in heavy and light water. They found a large enthalpic destabilisation of binding in D_2O relative to H_2O, a fact which is ascribed to the increased heat required to dehydrate the non-polar portions of the ligand. This seems at odds with their results which show that rapamycin, which binds with the burial of more polar and less non-polar surface area than tacrolimus, is destabilised to a greater degree in D_2O ($\Delta\Delta H$ typically 14-18 $kJmol^{-1}$ for rapamycin and 7-8 $kJmol^{-1}$ for tacrolimus). They also found that binding is accompanied by a large negative change in heat capacity in both D_2O and H_2O, and that ΔC_p is less negative in D_2O. No explanation for the difference was offered, however it may be that it reflects increased structural order of the solvent in D_2O (Kresheck et al, 1965).

5.2 pH versus pD

pD is defined in an exactly analogous fashion to pH i.e.

$$pD = -\log[D] \tag{5.2}$$

where [D] is the concentration of free D^+ ions in the solution (strictly the activity of the D^+ ion should be employed but in general it is possible to ignore the small difference between activity and concentration) should. One question which must be addressed is how to compare experiments carried out in H_2O and D_2O. In view of the strong influence pH has on the stability of proteins it is critical that experiments carried out in the two solvents are compared like for like. In fact two questions must be considered. How does the pD compare with the apparent pH (pH_{app}) measured using a pH electrode? The standard electrodes used routinely for measuring pH are specific for H^+ ions and care has to be taken when interpreting the measurements of pD in D_2O solutions. Measurements have been made using a deuterium gas electrode and compared with the pH_{app} obtained using a standard pH electrode (Covington et al, 1968). The result is the correction:

$$pD = pH_{app} + 0.41 \tag{5.3}$$

In other words the standard pH electrode underestimates pD by approximately 0.41 pH (or pD) units. The second question which must be answered is as follows. Should conditions be considered equivalent when pH = pD? Ionization constants are known to differ between heavy and light water (Kalinichenko & Lobyshev, 1976) and so what effect, if any, might this have on the protein. In answer to this question, acid and alkaline titration of a number of proteins has shown that the pK of proteins in heavy water is approximately 0.4-0.5 pH units higher in D_2O than in H_2O, and therefore there is no electrostatic

equivalence when pH = pD (Kalinichenko & Lobyshev, 1976). In other words the net charge on the protein in H_2O at a certain pH does not equal that at the same pD in D_2O solutions. The point at which there is electrostatic equivalence is, somewhat surprisingly, when pH = pHapp. So it would appear that the conditions under which comparisons should be made is when the apparent pH of the D_2O solution is equal to the actual pH of the H_2O solution.

5.3 Hydrogen exchange in protein interior

When proteins are dissolved in D_2O it is possible for the hydrogen atoms of all O-H and N-H groups to undergo hydrogen-deuterium exchange, this occurs rapidly where these groups are accessible to the solvent. However, a certain fraction of the exchangeable hydrogen atoms are buried deep in the hydrophobic core of the protein and so are retained until the protein undergoes thermal unfolding in the DSC (or indeed until any other process which allows permeation of the solvent into the protein interior occurs). The enthalpy associated with exchange reaction in alcohols:

$$\text{R-O-H} + \text{D}_2\text{O} \rightleftharpoons \text{R-O-D} + \text{DOH} \qquad (5.4)$$

was found to be $\Delta H = -0.32$ kJmol^{-1} (Chand & Fenby, 1978) and similar values would be expected for exchange in proteins. This then must be added to the measured enthalpy of unfolding. This leads to a potentially significant effect on the measured enthalpy of

unfolding. It is, therefore, necessary to set up controls to asses exactly how significant these effects may be. The controls were set up as outlined below.

It is assumed that, when carrying out DSC in D_2O, only the first scan will be significantly affected by heat effects associated with hydrogen exchange and that second and subsequent scans of the sample should be free from this additional energy consideration and so would give a true estimate of the unfolding enthalpy. However thermal unfolding of lysozyme is only partially reversible. HEWL samples removed from the DSC after just one scan are turbid and so appear to have undergone protein-protein aggregation. Furthermore, the appearance and subsequent increase of a second, low temperature transition as the area of the main transition is decreased upon re-scanning is observed (figure 5.2) indicating possible refolding to non-native structures. These two together result in the effective reduction of the concentration of native protein (figure 5.2). Since the effective protein concentration is not known for subsequent scans it is not possible to use the directly measured ΔH_{cal} from these scans to obtain the unfolding enthalpy. However it is possible to get around this difficulty by obtaining ΔH_{cal} from a number of scans and extrapolating back to the first scan. In H_2O the first scan should lie directly on whatever curve is chosen to fit the data (see below) whereas the first scan will lie on the curve if and only if the heat effects of exchange occurring within the time scale of the experiment are insignificant.

Figure 5.2 shows the thermograms obtained for lysozyme in H_2O (0.1 M sodium acetate/acetic acid buffer pH 4.41). Deconvolution of the data for the second and subsequent scans allowed estimation of the apparent calorimetric enthalpy, as shown in the table below:

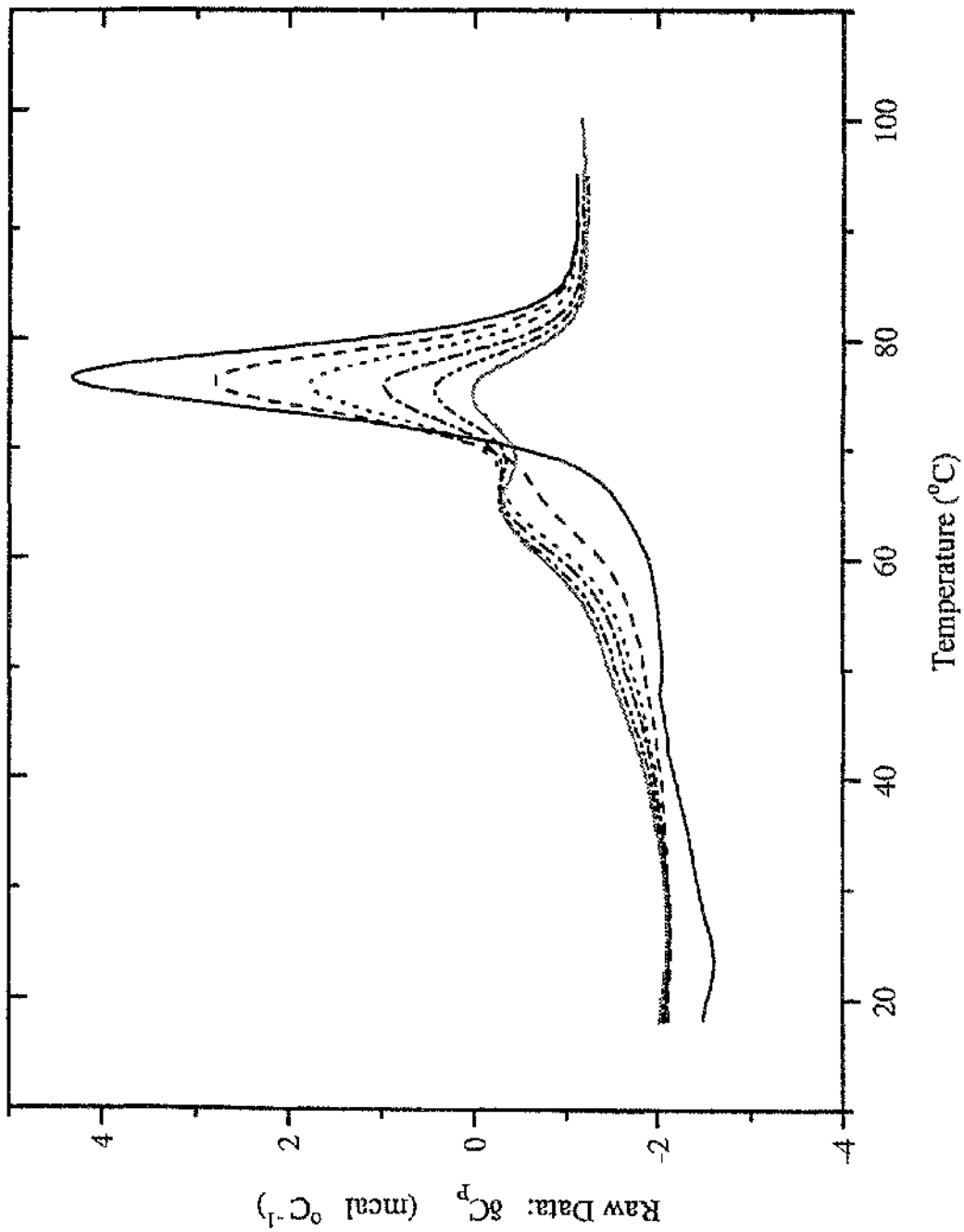

Figure 5.2 Multiple DSC traces of Lysozyme (H_2O, acetate buffer pH 4.41), re-scans results in reduction of the main (second) peak.

Scan	ΔH_{cal} ($kJmol^{-1}$)	$\ln\Delta H$
1	439	6.09
2	315	5.75
3	237	5.47
4	177	5.18
5	133	4.89
6	103	4.63

TABLE 5.1 Summary of enthalpies from re-scans of lysozyme in H_2O buffer.

If one makes the assumption that the same fraction of protein refolds to the native state each time then one finds that ΔH_{app} is related to the number in the following way:

$$\Delta H_{app} = \Delta H_0 F^S \tag{5.4}$$

where ΔH_o is the enthalpy of the zeroth scan, F is the fraction refolding and S is the scan number. Taking logs of equation 5.4 gives:

$$\ln \Delta H_{app} = \ln \Delta H_0 + S \ln F \tag{5.5}$$

In other words if the above assumption is correct then a plot of $\ln\Delta H_{app}$ versus scan number will give a straight line with slope = lnF. However, experience has lead to the belief that the fraction of the protein which refolds depends on the concentration of the protein, a

not unreasonable belief since in a more concentrated solution unfolded proteins are more likely to interact and therefore possibly aggregate prior to refolding. If this second assumption is correct and F increases with increasing protein concentration then the plot would be expected to follow a smooth curve, which is concave up.

When the data in table 5.1 for the unfolding in H_2O is plotted in this way it appears to produce a straight line within experimental error (figure 5.3) and the fraction refolding obtained from the slope is 0.7.

Data with D_2O buffer was obtained and treated in the same way (0.1 M sodium acetate/acetic acid buffer pH_{app} 4.14) as shown in table 5.2 and figure 5.3:

Scan	ΔH_{cal} ($kJmol^{-1}$)	$\ln\Delta H$
1	489	6.19
2	362	5.89
3	283	5.64
4	225	5.41
5	184	5.21
6	153	5.03
7	125	4.83
8	106	4.67

TABLE 5.2 Summary of enthalpies from re-scans of lysozyme in D_2O buffer.

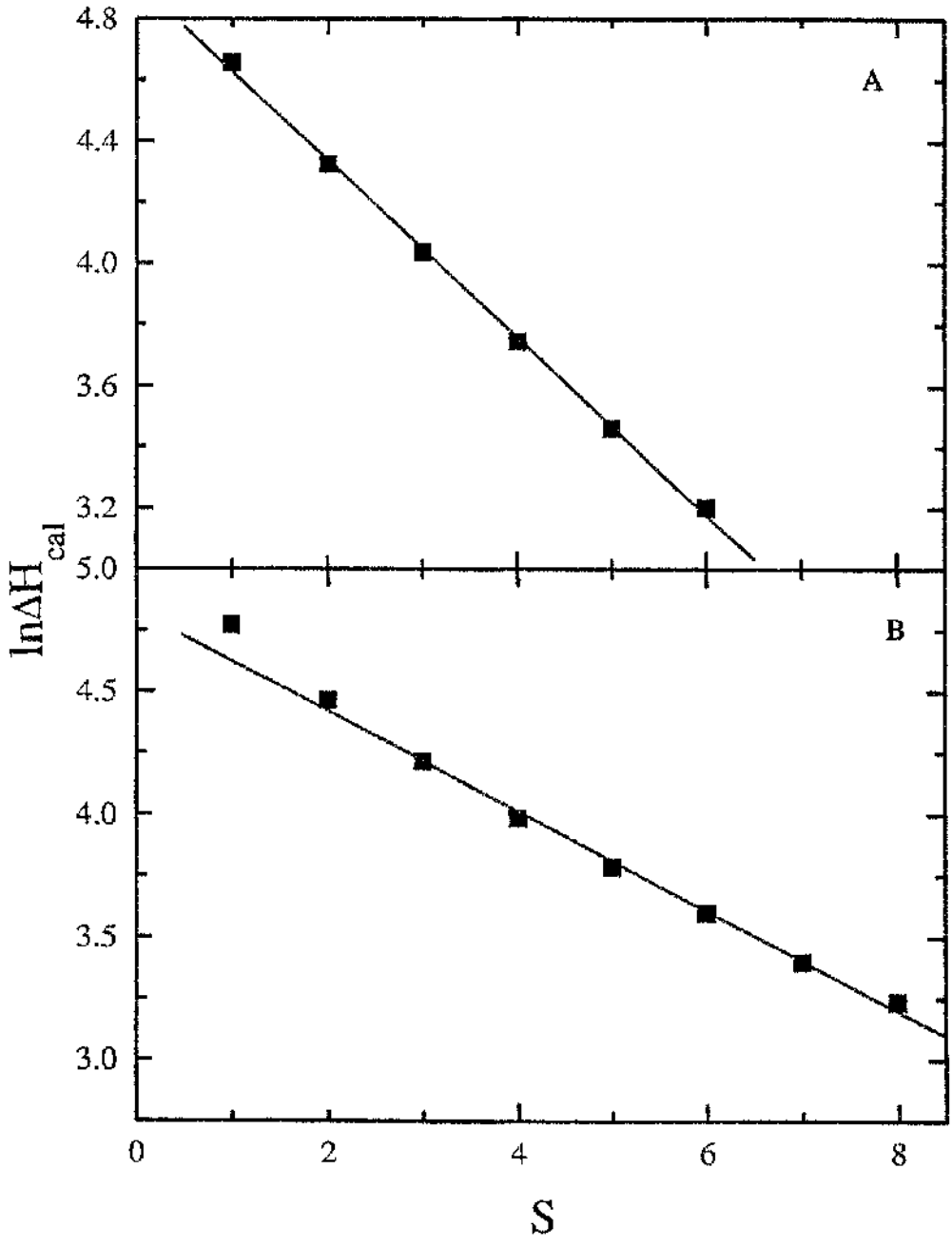

Figure 5.3 Plots of scan number versus natural log of ΔH_{cal} for lysozyme in H_2O (A) and D_2O (B) buffers.

The $\ln\Delta H_{app}$ versus scan number plot for lysozyme in D_2O buffer (figure 5.3) contains a larger number of data points and quite clearly has a significant curve, concave up (this curvature may have been present in the previous plot, but was not observed due to the fewer data points collected). This would seem to indicate that the second assumption holds and that the fraction refolding is some function of folded protein concentration. What is clear is that ΔH for the first scan does not appear to be at odds with the extrapolated value, so no significant heat effect accompanies the exchange upon unfolding (which, remember, was the point of the experiment in the first place). It is, therefore, safe to proceed and use the enthalpies observed for unfolding in heavy water without further correction.

5.4 Thermodynamics of lysozyme unfolding in H_2O and D_2O

DSC was carried out on lysozyme in various buffers over a range of pH. The buffers used (with ranges covered) were glycine/HCl (pH 1.5 - 3.5), sodium citrate/HCl (3.0 - 4.5) and sodium acetate/HCl (3.8 - 4.7). The following data were obtained:

Solvent	pH/pH$_{app}$	T_m (°C)
H_2O	1.50	47.4
H_2O	2.51	60.3
H_2O	3.03	63.6
H_2O	3.35	67.6
H_2O	3.52	69.9
H_2O	3.78	74.4
H_2O	4.03	75.4
H_2O	4.20	75.8
H_2O	4.41	76.2
H_2O	4.61	76.3
D_2O	4.14	78.1
D_2O	4.01	77.6

TABLE 5.3 Summary of data obtained from DSC of lysozyme in H_2O and D_2O buffers.

Clearly lysozyme is more stable in D_2O than in H_2O. Comparison of the enthalpies requires normalisation to some standard reference temperature using ΔC_p. The ΔC_p for lysozyme in D_2O is not known, and cannot be reliably estimated from the data here since there are too few data points which are too close together to construct a ΔH versus T_m plot, however ΔC_p for lysozyme in H_2O in known to be 6.3 $kJK^{-1}mol^{-1}$ (Cooper et al, 1992). Using this for both solvents gives the following results:

solvent	pH/pH_{app}	T_m (°C)	ΔH_{348} ($kJmol^{-1}$)	ΔS_{348} ($kJK^{-1}mol^{-1}$)
H_2O	4.03	75.4	434	1.38
D_2O	4.01	77.6	456	1.45
H_2O	4.20	75.8	469	1.49
D_2O	4.14	78.1	470	1.50

TABLE 5.4 Summary of ΔH and ΔS at 75 °C (348K) for lysozyme in H_2O and D_2O buffers.

5.5 Discussion of lysozyme results

DSC shows that lysozyme is more stable in D_2O buffer than in H_2O buffer by approximately 2°C in terms of T_m at pH (pD) 4.0 and 4.2. The unfolding enthalpies were normalised to the common temperature of 75°C using ΔC_p calculated in H_2O by Cooper et al (1992), the temperature of 75°C was chosen in order to minimise potential errors which would arise if the ΔC_p in D_2O is different from that in H_2O. It is not clear that it is a valid assumption that the ΔC_p can be taken to be the same in both solvents. Potentially significant ΔC_p effects have been indicated by model compound studies (Kresheck et al, 1965). Further work which permits calculation of ΔC_p for lysozyme in D_2O should be carried out. In line with the work of Connelly et al (1993) it would appear that lysozyme in D_2O undergoes an enthalpic destabilisation, this however is more than offset by an entropic stabilisation. This

is also in accord with the findings from the model compound studies which suggested that hydrophobic effects would be entropically weakened in heavy water, due to increased solvent structure, while hydrogen bonding interactions have increased strength, leading to a net stabilisation of proteins in heavy water relative to light water.

Along the way we have perhaps learned something about irreversibility in DSC experiments, certainly the long held suspicion that the fraction refolding after a transition depends on protein concentration has received some empirical backing. Similar re-scanning experiments with other proteins would provide an interesting comparison.

"Anyone can do any amount of work, provided it isn't the work he is supposed to be doing at that moment."

Robert Benchley.

"Work is the curse of the drinking classes."

Oscar Wilde.

Chapter six

Pf1 gene five protein

6.1 Introduction

One important class of proteins is the single stranded DNA (ssDNA)-binding proteins, as exemplified by the gene 5 protein (gVp) of the filamentous bacteriophages such as Fd and Pf1 (Alberts et al, 1972; Maeda et al, 1982). The protein performs two roles in the life cycle of the virus. It is the major component of the intracellular nucleic acid-protein complex, which takes the form of a regular helical assembly (Kneale et al, 1982; Gray, 1989), a process which appears to be a necessary step towards assembly of the virion (Pratt et al, 1974). It also acts as a translational repressor, binding to the gene 2 mRNA (Michel & Zinder, 1989). By a combination of these two activities the gVp appear to switch off double stranded DNA synthesis and switch on viral, ssDNA synthesis (Kneale, 1992).

The gVp's of the filamentous bacteriophages Fd, M13 and f1 have identical sequences and that of Ike has 40 % homology, however despite the obvious functional homology to the Pf1 gVp (Pf1-gVp) there is no overall sequence homology (Kneale, 1992). There is some very limited homology between the N-terminal regions of the two proteins, where the DNA binding wing of Fd-gVp is located. A similar role is proposed for this region of Pf1-gVp (Plyte & Kneale, 1991).

Pf1-gVp is 144 residues in length (which is 57 residues longer than the equivalent proteins in the other phages such as Fd) and has a molecular weight of 15,400. The protein has a globular N-terminal domain, which is proposed to be similar to that of the gVp's of the other filamentous bacteriophages (Kneale, 1992), and a flexible C-terminal domain from which it is possible to proteolytically cleave 32 residues and retain the protein's DNA binding function (Plyte & Kneale, 1993).

The gene five protein of the filamentous bacteriophages bind to the viral DNA to form a long helical nucleoprotein complex. These complexes tend to be extremely large with molecular weights of the order of 10^7 (Fd-gVc has molecular weight 1.8×10^7 and Pf1-gVc has molecular weight 3.1×10^7)(Gray et al, 1982; Kneale & Marvin, 1982). The NMR spectrum of the Fd-gVc has no detectable peaks originating from protein or DNA (Davis et al, 1995). This is a common feature of NMR on such large supramolecular complexes where slow tumbling of the complex in solution produces extremely broad peaks, resulting in the featureless spectra.

The NMR spectrum obtained for Pf1-gVc was unlike that for Fd-gVc in that a number of sharp peaks were observed with chemical shifts (δ) of 0-4.5 ppm. i.e. in the aliphatic region of the spectrum. Resolvable peaks in the spectra of such large structures generally arise from regions which enjoy a large degree of flexibility. It was proposed (Davis et al, 1995) that resonances arising from the side chains of the amino acids Ala, Gln and Pro could be identified with peaks at 1.40, 2.37 and 3.58 ppm respectively. These three amino acids make up twenty one of the thirty two amino acid C-terminal domain (8 Ala, 8 Gln and 5 Pro). Other peaks were assigned as follows:

δ (ppm)	contributing amino acid resonances
1.40	Ala, Lys
1.70	Lys, Arg
2.00	Gln, Arg
2.37	Gln
2.65	Asp
3.02	Lys
3.58	Pro
3.90	Ser
4.32	Ala, Arg, Lys

TABLE 6.1 Summary of the assignment of the nine major peaks in the NMR spectrum of Pf1-gVc (Davis et al, 1995).

All of the amino acid resonances are assigned to peaks within 0.05 ppm of their random coil values, and all are to be found in the C-terminal domain, leading to the assertion that all of the observable NMR signals arise from this region. It is proposed that the C-terminal domain forms a highly flexible, tail like structure.

DSC data (Davis et al, 1995) from the thermal unfolding of Fd-gVp and Fd-gVc (figure 6.1) showed that the formation of the complex resulted in a significant stabilisation of the protein. Free protein was found to have a T_m of approximately 51 °C while upon complexation with the viral ss-DNA the T_m increased to approximately 59 °C, which constitutes a stabilisation of 8 °C. The unfolding transitions of both complex and free protein were found to be irreversible due, most probably, to protein-protein aggregation in the unfolded state. For this reason T_ms could not be more accurately determined and no reliable estimate of the unfolding enthalpies could be made. It was clear, however, from the sharpness of the nucleoprotein complex thermogram that the unfolding transition was highly cooperative, suggesting some degree of protein-protein interaction along the axis of the DNA, perhaps analogous to interactions betwen adjacent groups on a string of beads.

DSC was carried out on the intact complex and free, intact Pf1-gV protein (figure 6.1) (Davis et al, 1995). The results showed that the free protein had a T_m of 45.4 °C while the complex had a significantly higher T_m of 53 °C, representing a 7.6 °C stabilisation of the protein upon formation of the complex. As with the Fd-gVp and Fd-gVc the thermal unfolding transition of Pf1-gVp was accompanied by apparent protein-protein exothermic aggregation. The thermogram obtained for the complex appeared to be free from this allowing calculation of the calorimetric and van't Hoff enthalpies which were found to be

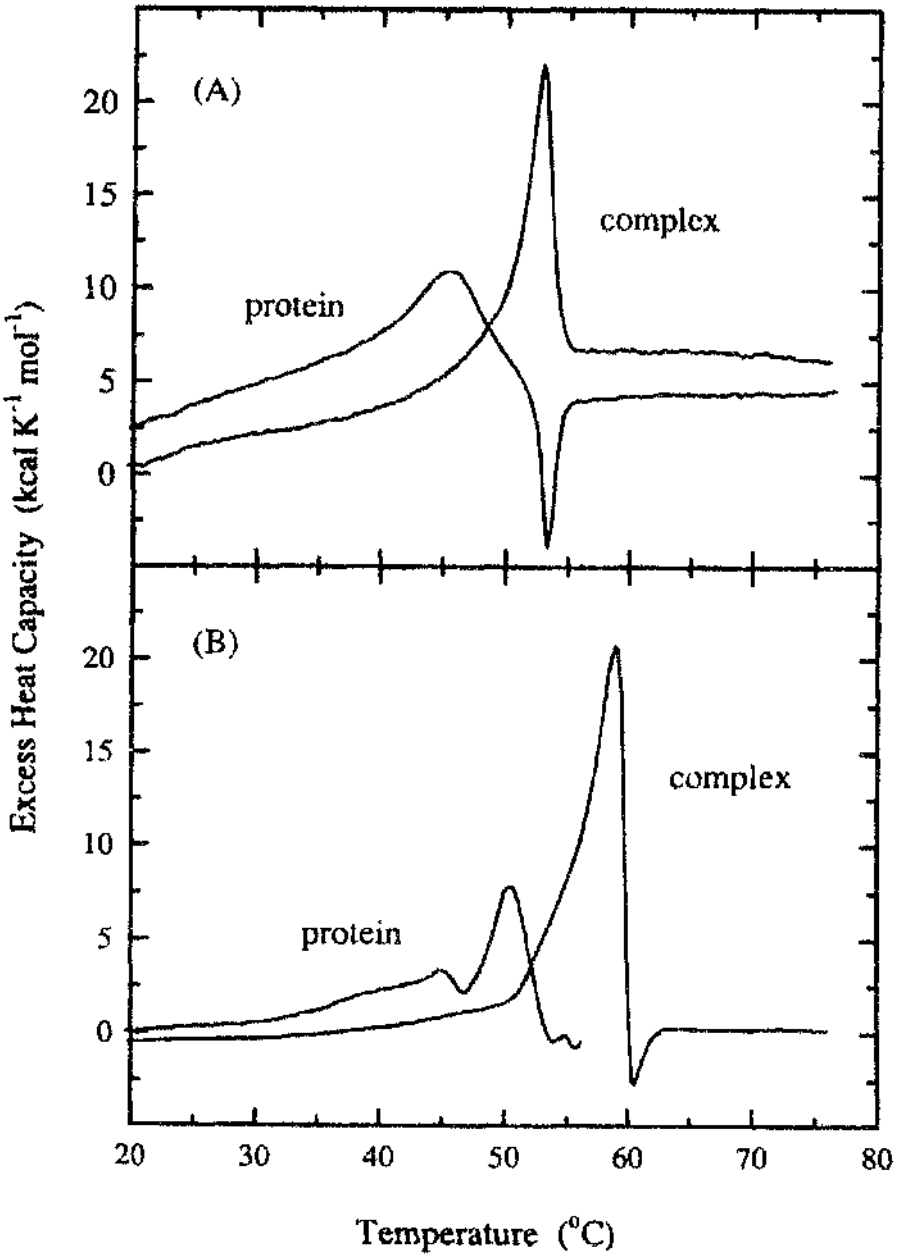

Figure 6.1 Normalized DSC traces of Pf1 (A) and Fd (B) gene five protein and complex (10mM Tris buffer pH 7.5).

ΔH_{cal} = 188.3 kJmol^{-1} and ΔH_{vH} = 1130 kJmol^{-1}. Using these two estimates it was possible to calculate that the cooperative unit contained six protein monomers (see section 2.2.2 equation 2.3).

This tail is highly susceptible to proteolytic cleavage, and can be selectively removed from the globular N-terminal domain of the protein (Plyte & Kneale, 1993), which retains its original structure. The NMR spectrum obtained for this partially proteolyzed complex shows that nearly all of the peaks are lost, supporting the assertion that these peaks arise from the amino acids which constitute the tail.

It is not clear what the precise role of the flexible tail like region of the protein is. The following questions still need to be addressed. Does tail actually stabilise the complex and if so how? Is the tail involved in protein-protein interactions along the DNA axis or does it interact directly with the nucleic acid? In order to further probe these questions a DSC study of partially proteolyzed Pf1 gene five nucleoprotein complex (Pf1-gVc) was undertaken in parallel with NMR studies in collaboration with K.G. Davis, S.E. Plyte and G.G. Kneale at the University of Portsmouth and the results examined in the light of recent work on both Pf1 and Fd nucleoprotein complexes.

6.2 Thermodynamics of the unfolding of proteolyzed Pf1-gVc

6.2.1 Thermal stability of the proteolyzed Pf1-gVc

DSC of the partially proteolyzed nucleoprotein complex was carried out and compared to the previous results for the intact complex. It was found that removal of the flexible C-terminal domain did not significantly affect the thermal stability of the complex. It was previously found that the transitions of Fd-gVc and intact Pf1-gVc were irreversible, and it was similarly found that cooling and re-scanning of the proteolyzed Pf1-gVc gave featureless thermograms (a typical example is shown in figure 6.2), indicating that the unfolding transition for the proteolyzed sample was also irreversible. This complicates analysis somewhat and it should be borne in mind when considering the data (the effect of aggregation is discussed in section 6.3).

6.2.2 Concentration dependence of thermal transitions

DSC was carried out on Proteolyzed Pf1-gVc over a roughly six fold concentration range and it was found that as the concentration was reduced so too was the observed T_m of the unfolding transition. Table 6.2 shows how the T_m varied:

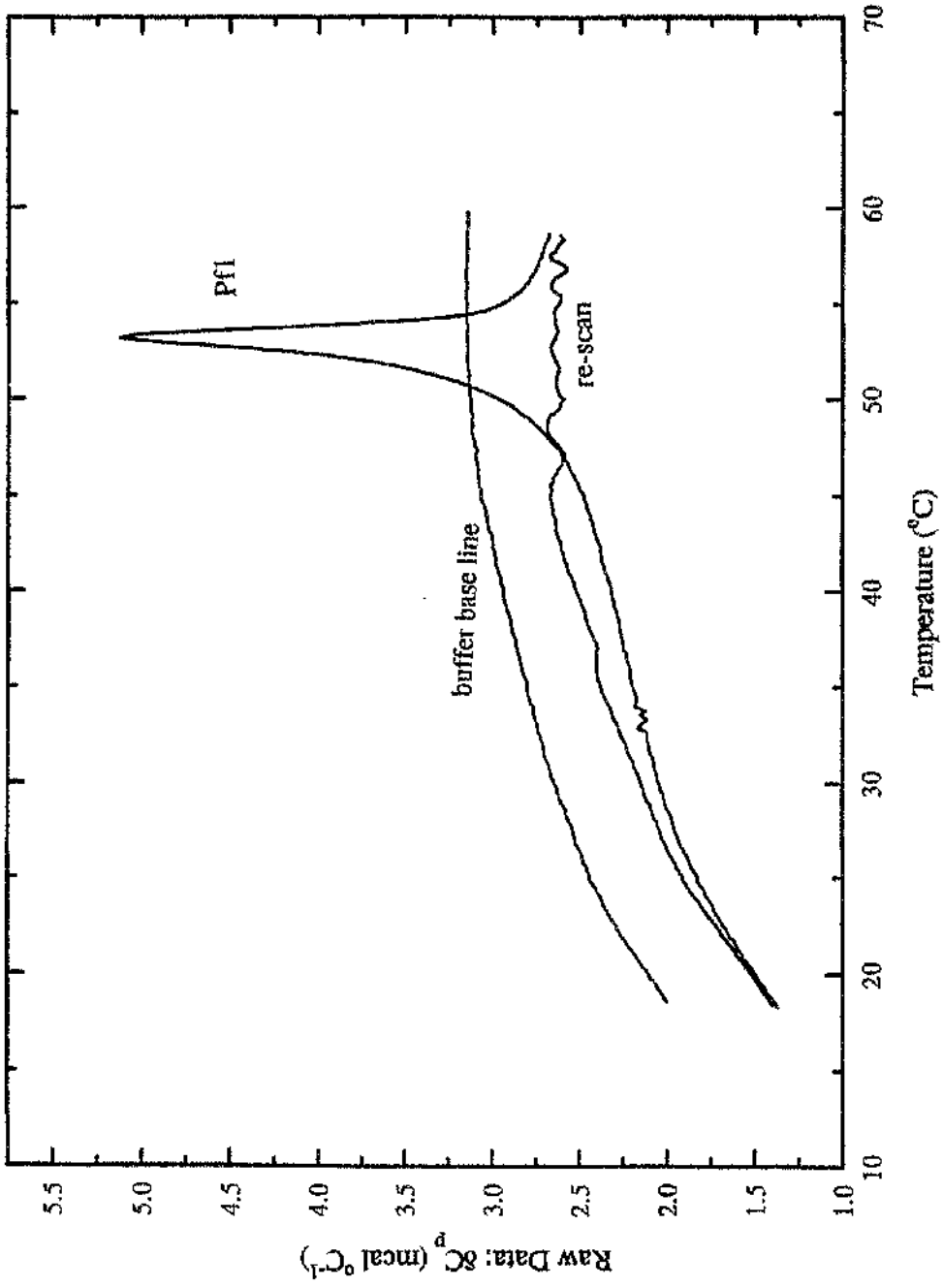

Figure 6.2 DSC thermogram of partially proteolysed Pf1 gene five complex showing re-scan and buffer base line (buffer conditions as for figure 6.1).

Protein monomer concentration (mM)	T_m(°C)
0.1160	54.0
0.0610	53.5
0.0302	53.0
0.0189	51.5

TABLE 6.2 Summary of concentration dependence of T_m for Pf1-gVc.

The shape of the transition also varied with concentration. The transitions observed with the most concentrated samples were sharp with a very pronounced asymmetry in their shape, while those observed at lower concentrations were much broader with much more symmetrical shapes (figure 6.3), although even at the lowest concentration the transition still deviated significantly from the normal symmetrical shape (figure 6.4).

6.2.3 Analysis of Pf1-gVc data

Both the temperature dependence of the T_m and the asymmetry observed (especially in the higher concentration experiments) are consistent with a decrease in the degree of oligomerization during the unfolding transition (Fukada et al, 1983; Manly et al, 1985; Bae et al, 1988). Furthermore it would appear that this reversible association of the protein

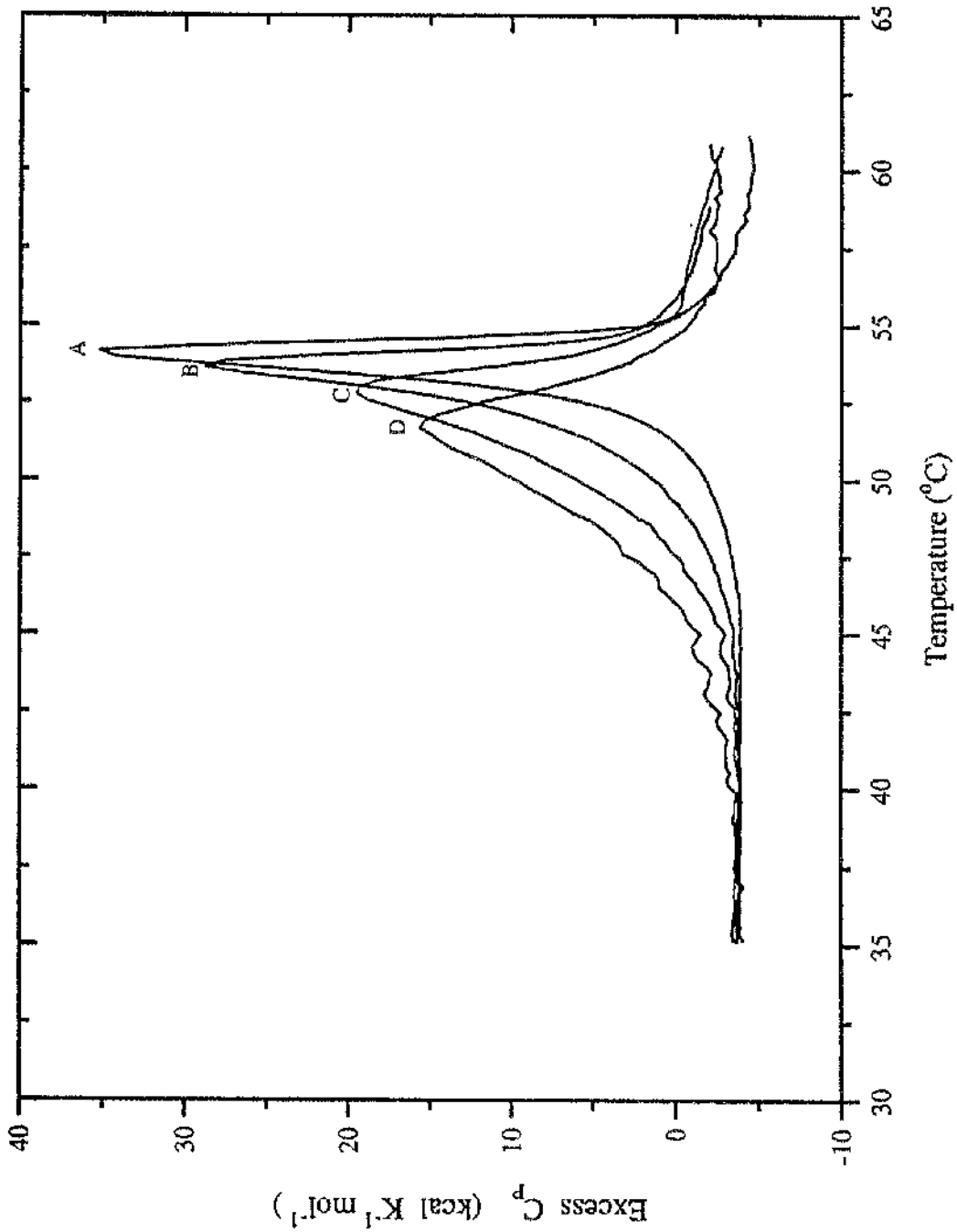

Figure 6.3 Normalized DSC thermograms of partially proteolysed Pf1-gVc at concentrations of (A) 0.166, (B) 0.061, (C) 0.0302 & (D) 0.0189 mM (buffer conditions as for figure 6.1).

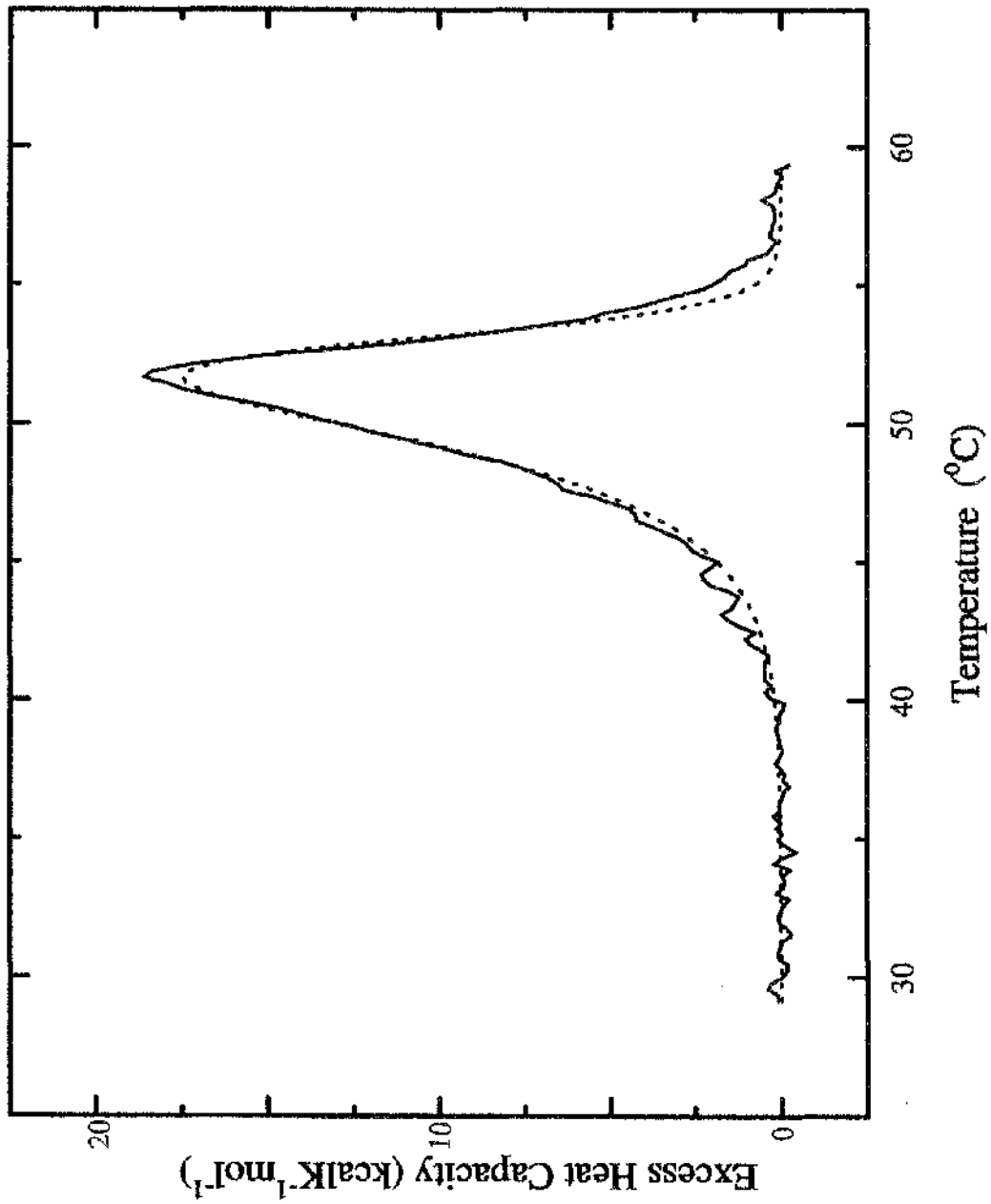

Figure 6.4 The lowest concentration scan (figure 6.3, D) shown with model which assumes four protein monomers dissociate and unfold cooperatively fitted (dotted line).

dimers with the ss-DNA forms increasingly larger cooperative units as the concentration is increased. Therefore data were analyzed in terms of the dissociation-equilibrium model:

$$nN \rightleftharpoons Nn \rightleftharpoons nU \qquad (6.1)$$

In this model subunit dissociation occurs simultaneously with thermal unfolding. The Microcal Origin software package is able to fit such a model to the data and this option was chosen. The results of this model fitting procedure were more satisfactory for the lower concentration thermograms, probably due to the greater degree of oligomerization affecting those at higher concentration. At the lowest concentration the model fit is good (figure 6.4) and indicates a cooperative unit of four protein monomers. The estimate of the size of the cooperative unit rises to 10-20 (or more) at higher nucleoprotein complex concentrations. These estimates are consistent with those made from the ratio of van't Hoff to calorimetric enthalpies (equation 2.3).

In general, for a system which undergoes a change in the degree of oligomerization the following equation applies:

$$(n-1)\ln[N] = -\frac{\Delta H_{vH}}{RT_m} + c \qquad (6.2)$$

where [N] is the concentration of protein monomers and c is a constant. In other words if the size of the cooperative unit, n, is a constant then a plot of ln[N] versus $1/T_m$ will give a straight line with the slope, S:

$$S = \frac{\Delta H_{vH}}{(n-1)R} \qquad (6.3)$$

If however n is in fact a function of concentration (in this case n apparently increases with increasing concentration) then plotting ln[N] versus $1/T_m$ would give a curve. With n increasing with increasing concentration the curve would be expected to be concave up. This is indeed what is observed when this plot is constructed (figure 6.5). The curve in figure 6.5 is there as a guide to the eye only, it is not possible to fit a model to this data since the nature of the function which relates n to concentration is not known.

DSC of proteolyzed Pf1-gVc shows that in this system the unfolding reaction is accompanied by a decrease in the degree of oligomerization. This is not unusual, the effect is observed in a number of other systems and the DSC results arising from such a system are well described (Fukada et al, 1983; Manly et al, 1985; Sturtevant, 1987; Bae et al, 1988). The typical DSC trace for such a system displays a marked assymetry in the transition, being sharper on the high temperature side. Such assymetry is observed in the thermograms of Pf1 (figures 6.3 and 6.4). Another feature observed for Pf1 and typical of dissociating systems is the temperature dependence of T_m with increasing concentration. (associating systems are assymetric in the opposite sence and have T_m decreasing with increasing concentration) Where Pf1-gVc is unusual is in the fact that n, the size of the cooperative unit, is also apparently dependant on concentration. This results in the curvature noticed in the plot of ln[N] versus inverse T_m.

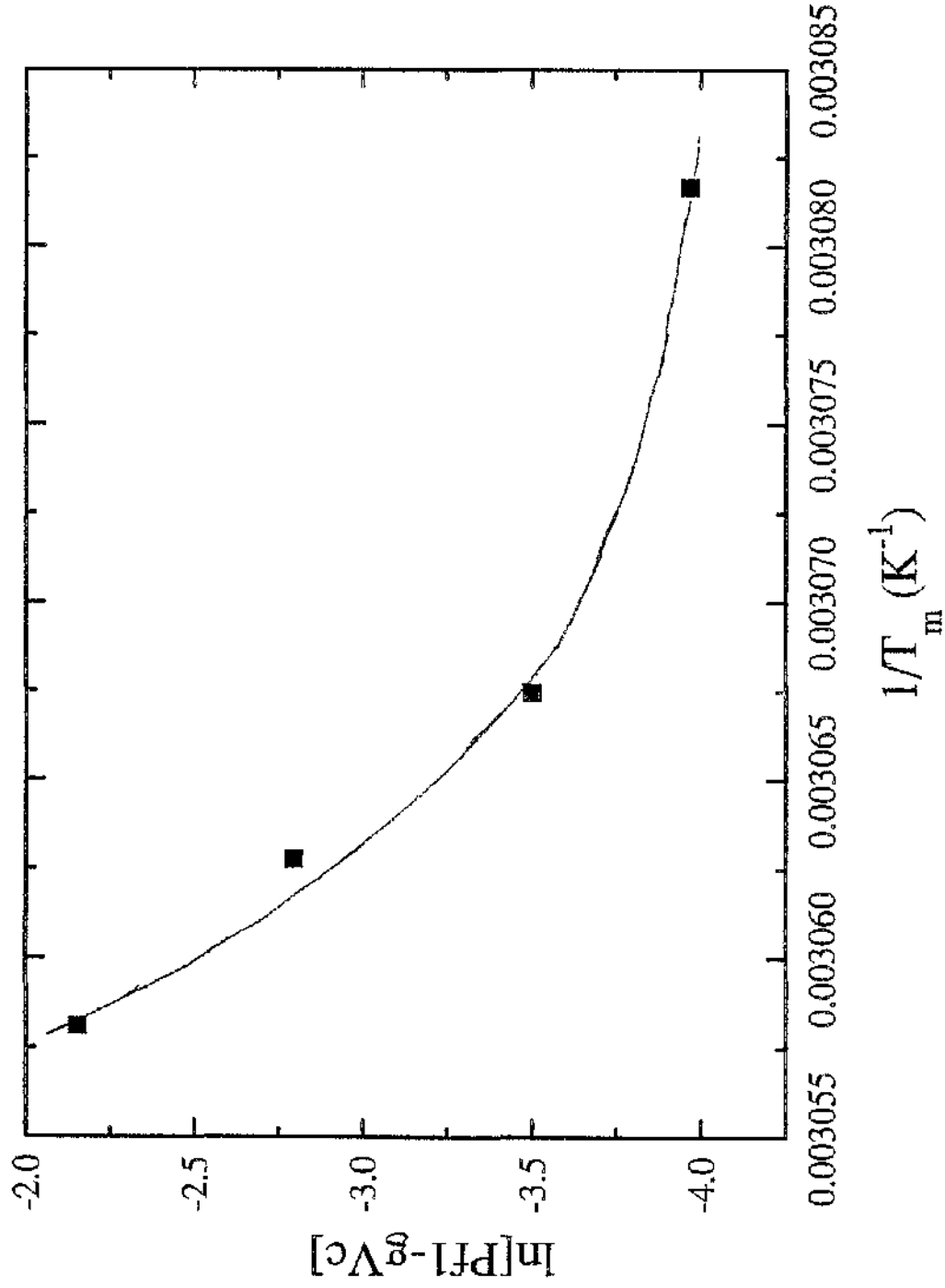

Figure 6.5 Plot of inverse T_m versus natural log of protein concentration for Pf1-gVc.

Comparison of the DSC of proteolyzed complex with that of intact complex indicates that cleavage of the flexible C-terminal domain has no effect on the thermal stability of the complex. Plyte and Kneale (Plyte & Kneale, 1993) have shown that the globular N-terminal domain left after proteolysis and intact protein are able to bind the ss-DNA with a similar affinity. However, the complex formed is kinetically less stable when the tail is removed, as judged by gel retardation studies which showed that the complex formed has a relatively short lifetime such that the equilibrium is disturbed in the time taken to run the gel (Fried, 1989; Plyte & Kneale, 1993). Two possibilities exist for the role played by the tail. The first is that it is involved in protein-protein interactions along the ss-DNA. This appears to be refuted by the results presented here which show that even when the tail is absent there is a considerable degree of association along the DNA. If the tail were involved in protein-protein contacts there could be no such association. Also there is little evidence for the loss of thermal stability of the complex upon removal of the tail, which would expected if there was a significant stabilising interaction. The second, and apparently more likely, possibility is that the highly flexible tail acts as a kinetic clamp, looping round the ss-DNA to interact with the same protein dimer. In order for the DNA to gain access to the binding site it would be necessary for the clamp to be a dynamic one, the clamp would have to open simultaneously on many subunits before the DNA could be released.

6.3 Effect of aggregation on thermal transitions

The effects of protein-protein aggregation and other such exothermic processes are not taken explicitly into account in the above analysis of the DSC results, however it should

be remembered that these transitions are irreversible and appear to be affected by some kind of exothermic rate-limited step, especially at higher concentration. The possible effects of this will now be considered in relation to the data and analysis presented above.

In the endotherms collected it is apparent that the onset of any aggregation which could cause the T_m to be incorrectly estimated occurs after the transition midpoint. It would be expected then that this would tend to cause the T_m to be under estimated and where more aggregation occurred (i.e. at higher concentration) the error would be greater. In other words one of the effects of aggregation would be to cause an apparent decrease in T_m with increasing concentration. Clearly any artifact caused by the aggregation of the protein would be contrary to the observation that T_m increases with increasing protein concentration.

Aggregation might also be expected to produce an artificial sharpening of the high temperature portion of the transition, which could cause the kind of asymmetry observed and could, in principle, account for the relative sharpness of the thermograms obtained at higher concentration. The effect of aggregation on its own would be simply to reduce the apparent width of the peak, however the peak height would be unaffected. The data shown in figure 6.3 are all normalized with respect to concentration and so direct comparison is possible. Clearly the thermograms corresponding to higher concentrations are significantly higher. In other words the transitions are additionally sharpened by factors other than aggregation. Analysis of the measured enthalpies suggests that thermograms obtained at higher concentration are sharper than those at lower concentration due to increased size of the cooperative unit.

It is clear then that the interpretation of the data presented in the previous sections is not greatly affected by the possible protein-protein aggregation. Most significantly the temperature dependence of T_m would be, if anything, under estimated as a result of aggregation and therefore the magnitude of the effects discussed would also be under estimated.

"The average PhD thesis is nothing but the transference of bones from one graveyard to another."

J. Frank Dobie.

Chapter Seven

MS2 and AhrC

7.1 MS2 viral coat protein

7.1.1 Introduction

Male specific bacteriophage, fraction 2 (MS2) is a positive sense, single stranded RNA bacteriophage which infects male *E. coli* cells (Brinton et al, 1964; Crawford et al, 1964). Its entire genome, consisting of 3,569 nucleotides has been sequenced (Min-Jou et al, 1972; Vandenberghe et al, 1975; Fiers, 1975; Fiers, 1976). Like all viruses the nucleic acid is surrounded and protected by encapsidation with a coat protein while exiting and entering new cells. The capsid is made up of 180 copies of the coat protein (molecular weight 13,700) and one copy of a protein known as the A protein (molecular weight 44,000), which is believed to be important for gaining entry into the cell (Steitz, 1968).

When formed the capsids are icosahedral in shape with a triangulation number (T) of three (Valegård et al, 1990). The three dimensional structure of the capsids has been solved (Valegård et al, 1990; Golmohammadi et al, 1993) and refined to 2.8 Å and the structure of a complex between capsids and an RNA 19mer has been reported to 3.0 Å (Valegård et al, 1994).

The viral coat protein monomer consists of a five-stranded anti-parallel Beta-sheet with two anti-parallel β-strands at the N terminus and a kinked α-helix at the C terminus. It can adopt one of three possible structures in the capsid, A, B or C, depending on position. The major difference between the three structures is the relative positions of the loop joining the F and G strands of the β-sheet (known as the FG loop: figure 7.1). The monomers form two types of dimer AB and CC in a ratio of two to one. Each triangular face of the icosahedron upon which the structure is based is made up of one of each of the three possible structures, with their dimer counterparts being part of an adjacent face (figure 7.2).

Nature is, by nature, conservative (with a small c). Its frugality is shown by its double use of the coat protein. At low protein to RNA ratios a complex of one coat protein dimer bound to a 21 nucleotide long hairpin loop (complex 1) is formed. Located on this loop is the initiation codon for the viral replicase gene (Bernardi & Spahr, 1972; Beckett & Uhlenbeck, 1988) and in this way acts as a translational repressor (Sugiyama & Nakada, 1967), switching off replication of the viral genome in preparation for exiting the host cell and infecting new cells. In intact virus, the formation of this complex has been shown to be a trigger for formation of complete capsids (Hohn, 1969; Hung et al, 1969; Beckett & Uhlenbeck, 1988). Addition of coat protein to complex 1 leads to formation of complex 2,

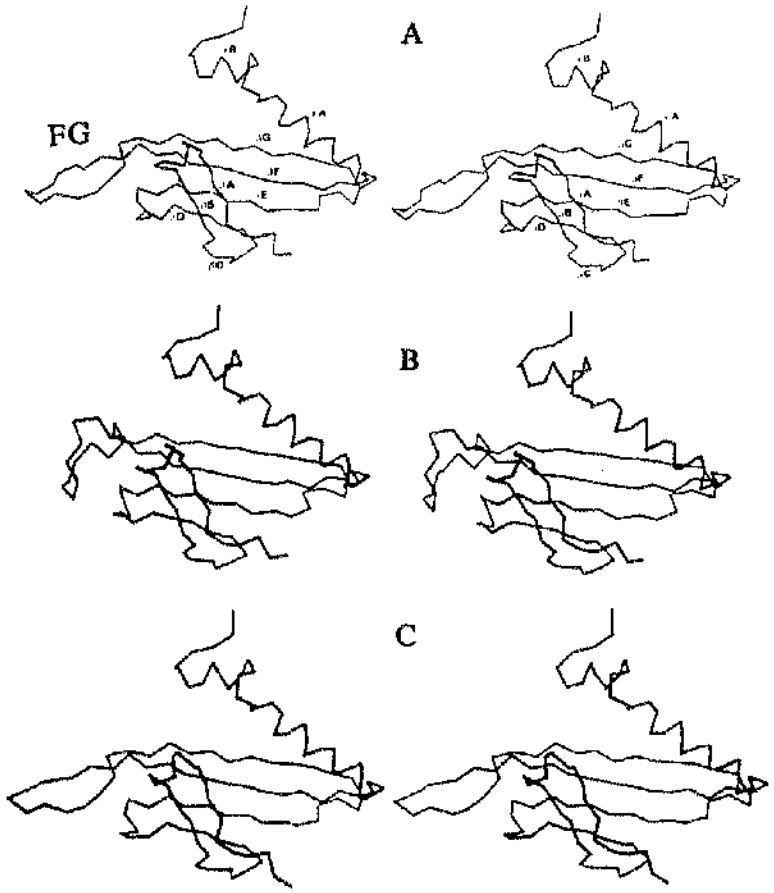

Figure 7.1 Stereograms of the three conformations adopted by the MS2 coat protein, showing the orientation of the FG loop (copied from Valegård et al, 1990).

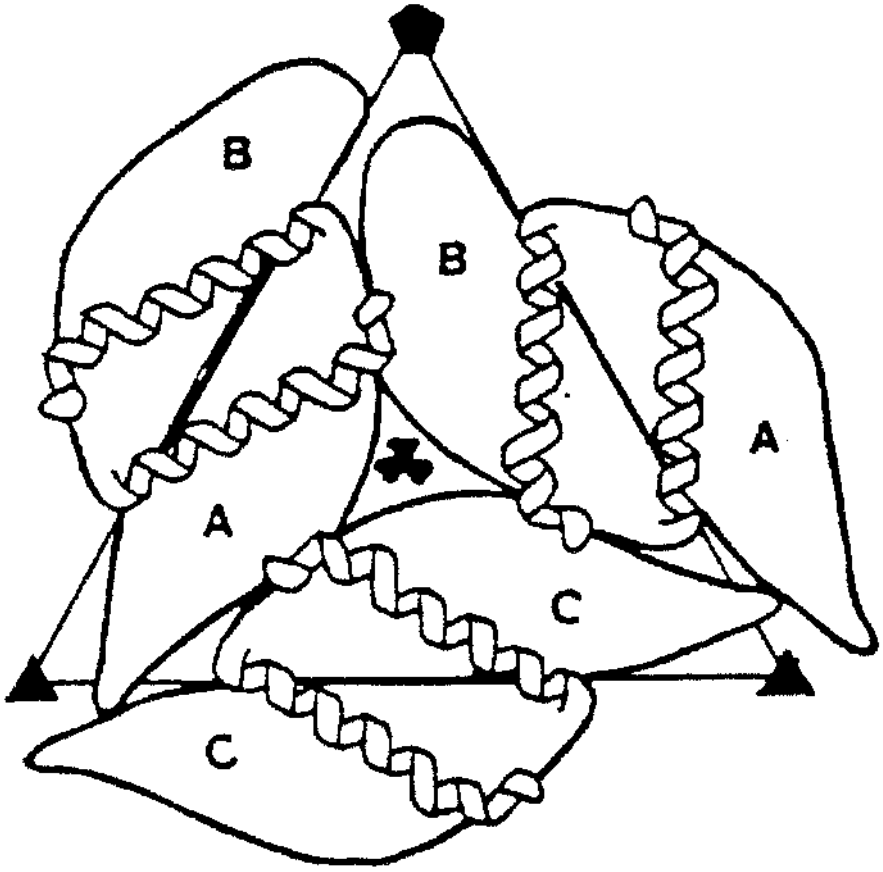

Figure 7.2 Diagram of dimer contacts made within the capsid (copied from Valegård et al, 1990).

which resembles the complete capsids but is lacking in the A protein. This results in the RNA being partially exposed, and therefore accessible to nucleases (Heisenberg, 1966).

Studies on the thermal stability of the empty viral capsids, employing transmission electron microscopy (TEM: Stonehouse & Stockley, 1993), have found that capsids incubated at 68°C for 10 minutes remained intact but those incubated at 69 °C have apparently melted and only amorphous masses could be visualised. Upon cooling of these samples a small number of capsids appeared to reform but the majority of the material remained amorphous. In samples incubated to 70.5°C no reassembly appeared to take place at all. It was found that intact virus and empty capsids underwent melting at approximately the same temperature, indicating that the presence of viral RNA confers no increased stability upon the capsid.

The crystal structure of the MS2 bacteriophage capsids has been solved, so too has the structure of the coat protein-operator complex (figures 7.1 & 7.2). The thermal stability of capsids lacking in genomic RNA has been investigated using transmission electron microscopy, TEM (Stonehouse & Stockley, 1993) and in parallel studies we have investigated the thermal stability of the coat protein using DSC and have begun to study the binding of a specific RNA 19-mer using ITC.

7.1.2 Thermodynamics of the unfolding of MS2 viral coat protein

A typical DSC thermogram of MS2 viral coat protein is shown in figure 7.3. The protein undergoes a very sharp single endothermic transition from native to denatured states, followed by an exothermic process which is probably protein-protein aggregation. The unfolding was irreversible as can be seen from the re-scan shown in figure 7.3. The transition has an asymmetric shape, which may indicate a reduction in the degree of oligomerization upon unfolding (see chapter six), however it is more likely that this is a spurious effect caused by the artificial sharpening of the transition by aggregation.

The data were fitted to a non-two state model using the Origin software package from Microcal (figure 7.4). This model gives two estimates of the enthalpy of unfolding, the calorimetric and van't Hoff enthalpies. However, enthalpy is an extensive property of the sample and so depends on the quantity of material present. It proved to be extremely difficult to obtain reliable estimates of the protein concentration. This is because the preparation and purification method used does not remove all of the nucleic acid present in the sample, a variable quantity of which is encapsidated by the protein. Nucleic acids have an absorbance maximum at 260 nm which overlaps with the protein maximum at 280 nm, prohibiting the use of simple UV measurements to estimate concentration. Instead two methods were employed as detailed in chapter three (sections 3.6.3 and 3.6.4), these were the coomassie brilliant blue assay and the method of Warburg and Christian (1942), which relies upon taking a ratio of the absorbance at 280 nm to that at 260 nm. These gave two different estimates of protein concentration, either or neither of which might be correct. This means that the calorimetric enthalpy calculated from the model fit depends on the method

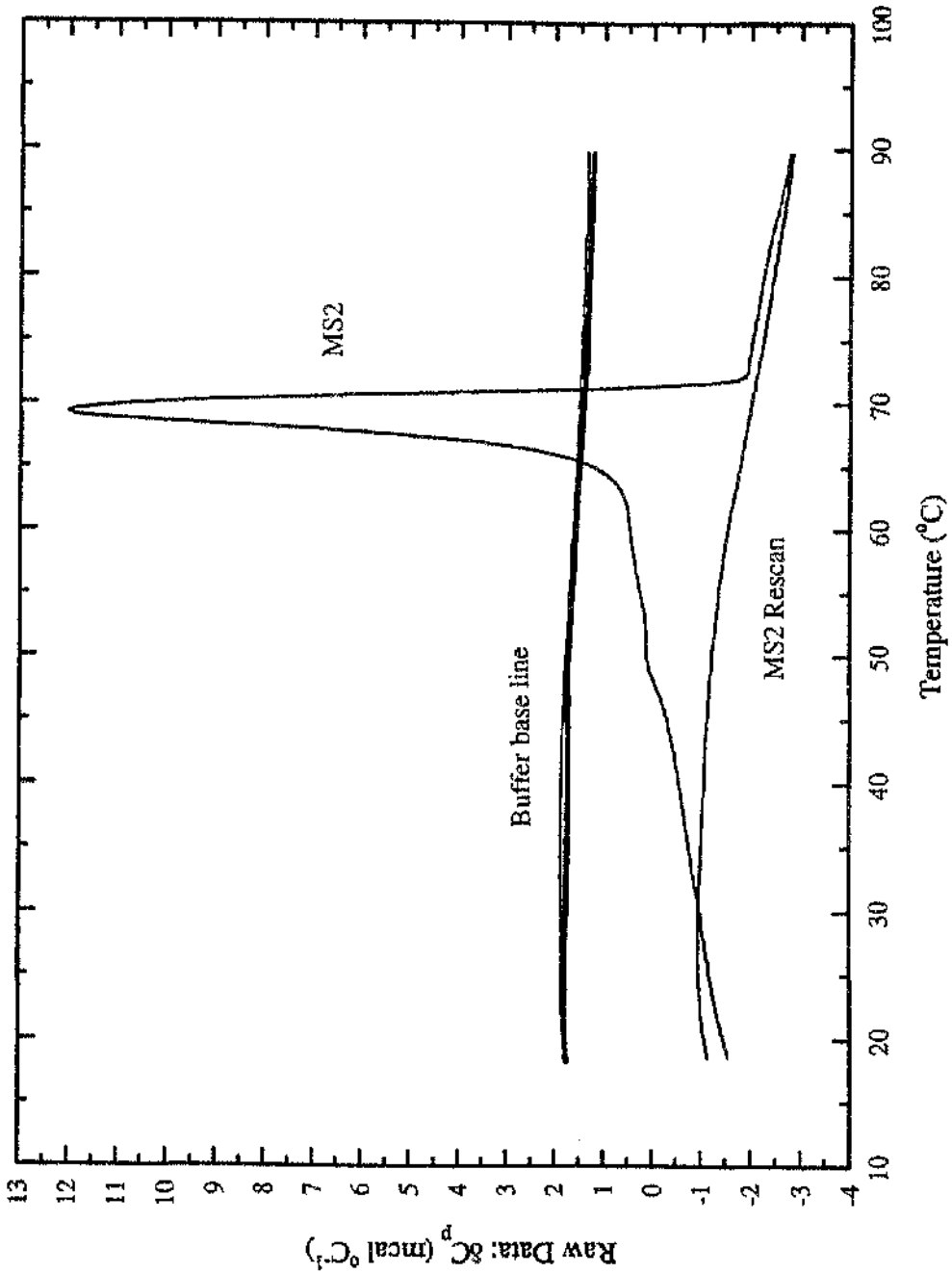

Figure 7.3 DSC thermogram of MS2 viral coat protein showing re-scan and buffer baseline (10mM HEPES pH 7.4, 1mM EDTA, 100mM NaCl).

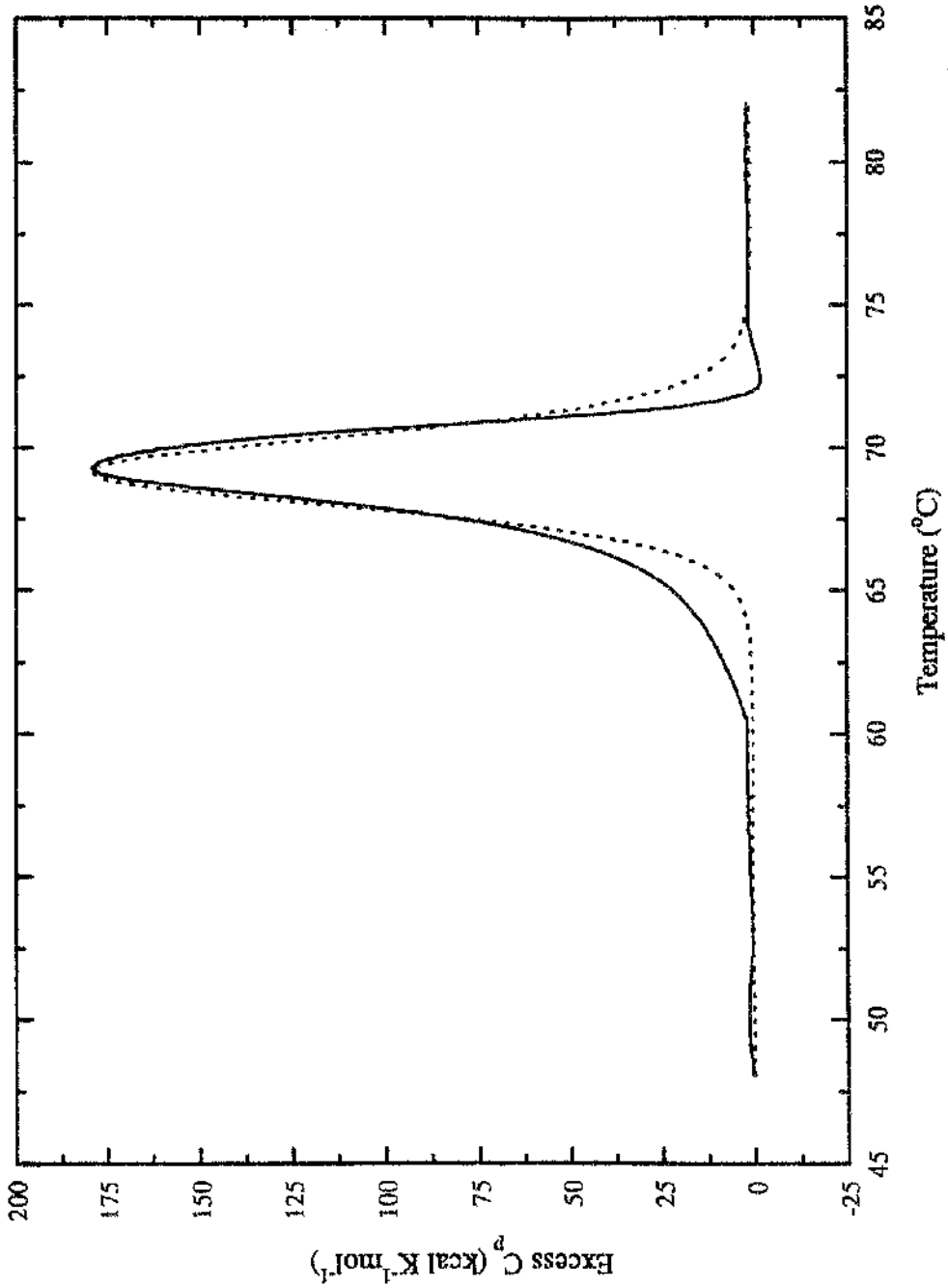

Figure 7.4 Normalized DSC thermogram of MS2 viral coat protein with buffer baseline subtracted (buffer conditions as for figure 7.3) and model fit shown (dotted line).

used for measuring the protein concentration. The van't Hoff enthalpy is independent of concentration and so the same value (within experimental error) is obtained. DSC of three identical samples, from the same original stock, were carried out with results shown in table 7.1 below for each estimate of protein concentration:

Sample	method	Conc.(mM)	ΔH_{cal} ($kJmol^{-1}$)	ΔH_{vH} ($kJmol^{-1}$)	Tm(°C)
Sample 1	Coomassie	0.0588	2415	1174	69.2
Sample 1	A280/A260	0.595	246	1162	69.2
Sample 2	Coomassie	0.0344	862	1205	69.1
Sample 2	A280/A260	0.0555	544	1184	69.1
Sample 3	Coomassie	0.0113	947	1149	68.8
Sample 3	A280/A260	0.0403	274	1129	68.8

TABLE 7.1 Summary of DSC results for MS2 viral coat protein in 10mM HEPES, 1mM EDTA, 100mM sodium chloride titrated to pH 7.4 with concentrated sodium hydroxide.

Clearly the problem of concentration estimation will have to be addressed before any meaningful thermodynamic data can be obtained. The transition midpoint temperatures are in good agreement with the findings from the TEM studies (Stonehouse & Stockley, 1993) which found that at 68°C the capsids remained intact, but raising the temperature by just one degree resulted in complete destruction of the capsids.

7.1.3 Thermodynamics of RNA binding by MS2 viral coat protein

Even in the absence of RNA the coat protein on its own forms capsids, with only a small opening created by the absence of the single copy of the A protein. It was decided to investigate the possibility of using isothermal titration calorimetry to study the binding of a specific RNA oligonucleotide to the capsid. The oligonucleotide used was a synthetic 19-mer which forms the hairpin loop to which the coat protein is known to bind (figure 7.5). Given that the RNA must first find its way into the interior of the capsid it was necessary to take into account the possibility that there would be a significant kinetic barrier to be overcome. For this reason an injection schedule in which each injection lasted for ten seconds with a five minute delay between was employed. The experiment was set up with coat protein in the cell at a concentration of 0.12 mM (from a coomassie assay, see previous section) and synthetic RNA 19-mer in the syringe at a concentration of 0.24 mM. The actual binding isotherm obtained was very noisy and is shown in figure 7.6.

As can be seen the binding reached saturation after the fifth injection. It was therefore decided to terminate the experiment early, after the eleventh injection, in order to preserve some material. Given the short supply of the material it was decided to carry out the necessary control experiments in the following way. First the buffer into buffer control was obtained, the cell was then filled with coat protein and three injections of buffer into coat protein were carried out, these were small enough to be ignored. The sample in the cell was not disturbed and the concentration was corrected for the buffer injected into it before titration with the RNA 19-mer was started. After this was terminated the sample in the cell was removed and injections of 19-mer into buffer were carried out. This was the largest

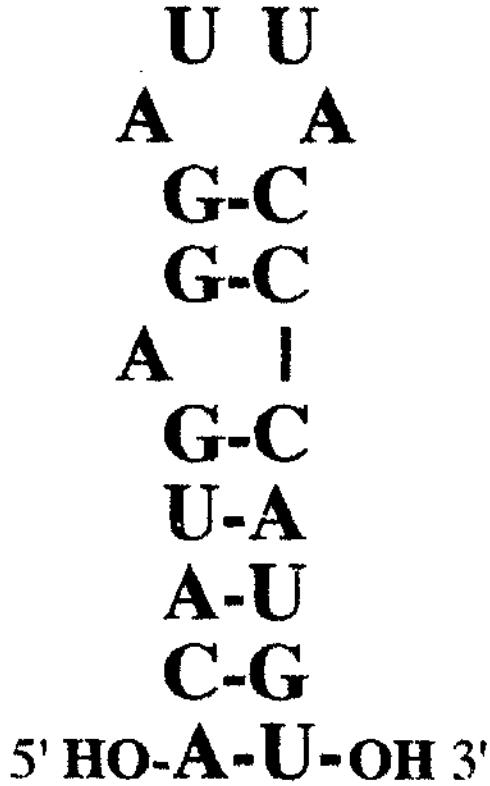

Figure 7.5 RNA 19-mer used in ITC experiments.

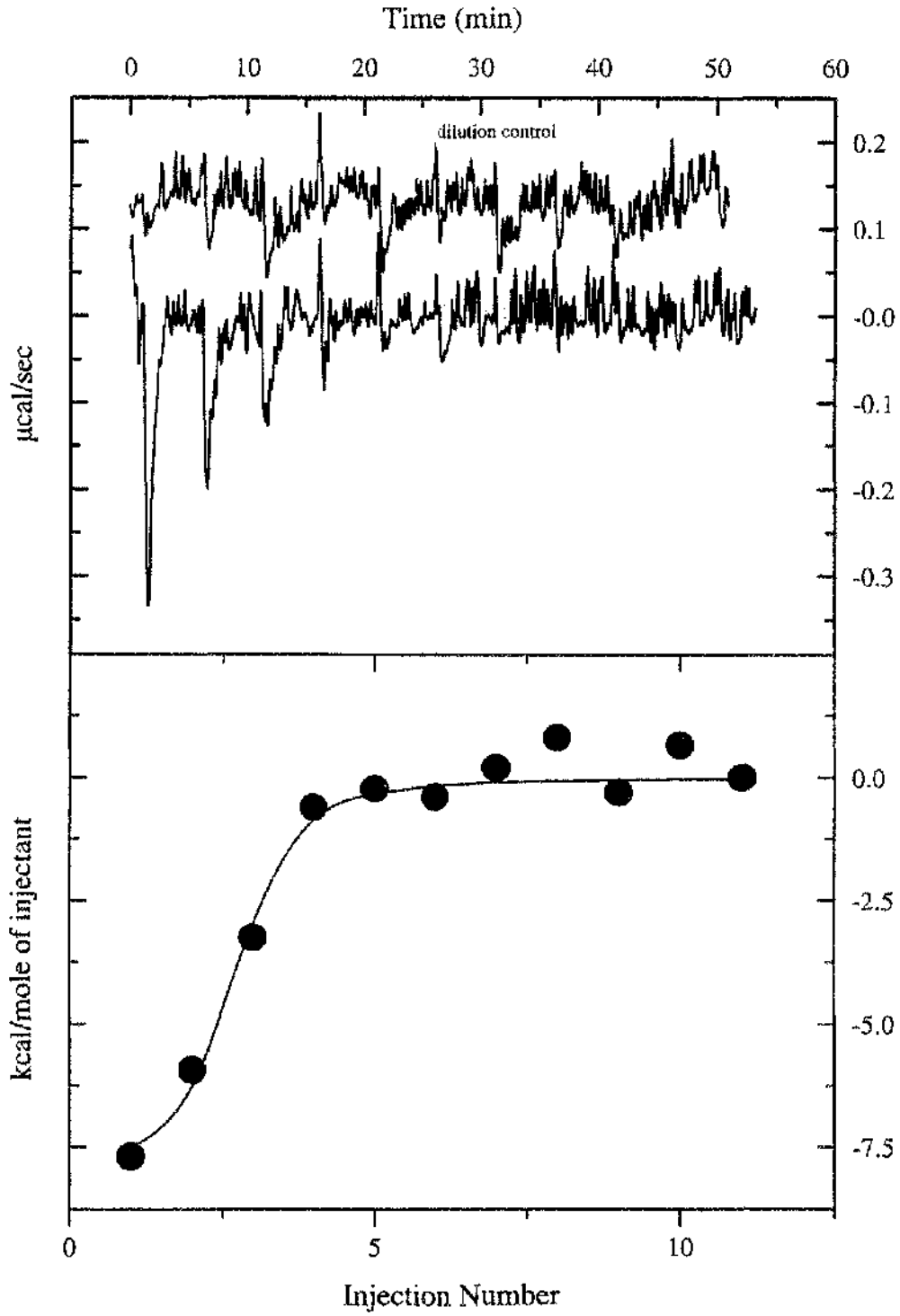

Figure 7.6 Typical ITC results for MS2-RNA binding, controls as shown (buffer conditions as before).

control and as can be seen from figure 7.6 it was very small in comparison with the noise and so was not subtracted from the titration data.

A model for one set of sites was chosen for fitting to the data, the fitting parameters obtained were N = 0.028, K_b = 6.7 x 10^6 and ΔH = -33.7 $kJmol^{-1}$. However considering the quality of the data and the difficulty described above in determining the concentration of the coat protein these figures are given for completeness only. Only qualitative statements can be made. However despite the relatively inconclusive nature of the data obtained so far these are, to the best of our knowledge, the first direct calorimetric observations of insertion of RNA into an intact viral capsid.

Microcalorimetry may yet prove to be a useful tool for the study of this system, but if and only if the problems in estimating the concentration of the protein can be overcome. If these problems are overcome DSC of the coat protein in the presence and absence of RNA should be carried out, with reliable concentrations it would be possible to estimate the size of the co-operative unit. Also a large number of mutant forms of the protein have been produced by the group in Leeds (Stonehouse & Stockley, 1993) and DSC and ITC experiments using these could provide useful thermodynamic data on protein stability, capsid assembly and protein-RNA interactions.

7.2 AhrC

7.2.1 Introduction

There is a large number of families of DNA-binding proteins, with each member of the family containing a recognition motif, the best known being the helix-turn-helix motif (Pabo & Sauer, 1984; Brennan & Matthews, 1989; Harrison & Aggarwal, 1990). The arginine (Arg or R) repressors of *Escherichia coli* and *Bacillus subtilis* are believed to belong to a new family because there is no homology between these and any other known DNA-binding protein (North et al, 1989), although neither of these proteins have had their crystal structures solved so far.

Again the economy of nature is revealed in *B. subtilis*. The protein AhrC both switches off the transcription of the enzymes responsible for Arg synthesis and switches on synthesis of those responsible for catabolism (Czaplewski et al, 1992). It is only in the presence of Arg that the binding is possible, in this way nature achieves a feedback system where excess Arg in the bacterium turns off its own production, until the levels fall too low for binding of the repressor and synthesis is allowed to continue.

AhrC is believed to bind to the *B. subtilis* argC promoter region with a stoichiometry of six to one (Czaplewski et al, 1992), in other words it is believed that an AhrC hexamer is responsible for binding to the operator. AhrC is also able to bind with

E. coli ARG boxes, which is the binding site for the functionally homologous ArgR protein in *E. coli*. Despite this close functional homology the sequences only show 27% amino acid identity. It is in the C-terminus of both proteins that the sequences contain most identity with one another, giving rise to the belief that the two share a common binding domain (North et al, 1989).

In a preliminary piece of work it was decided to investigate whether titration microcalorimetry was able to give useful information concerning the binding of Arg to AhrC.

7.2.2 Thermodynamics of Arg binding by AhrC

Isothermal titration calorimetry experiments on the binding of Arg to AhrC indicates exothermic complex formation (figure 7.7). In the experiment illustrated in figure 7.7 the reaction cell was filled with AhrC at a monomer concentration of 0.1398 mM and injection syringe filled with L-Arg at a concentration of 10.02 mM. As with all AhrC experiments, the buffer used contained 20 mM TRIS, 10 mM magnesium chloride, 250 mM sodium chloride and 1mM DTT, titrated to pH 7.4 with concentrated hydrochloric acid and an injection schedule of twenty five injections of 10 μL each with a three minute delay between was employed. As can be seen in figure 7.7 the binding is quickly saturated at these concentrations. Repeating the experiment with reduced concentration of Arg in the syringe (5.02 mM) but otherwise identical conditions gives a titration isotherm from which much more information can be gleaned (figure 7.8).

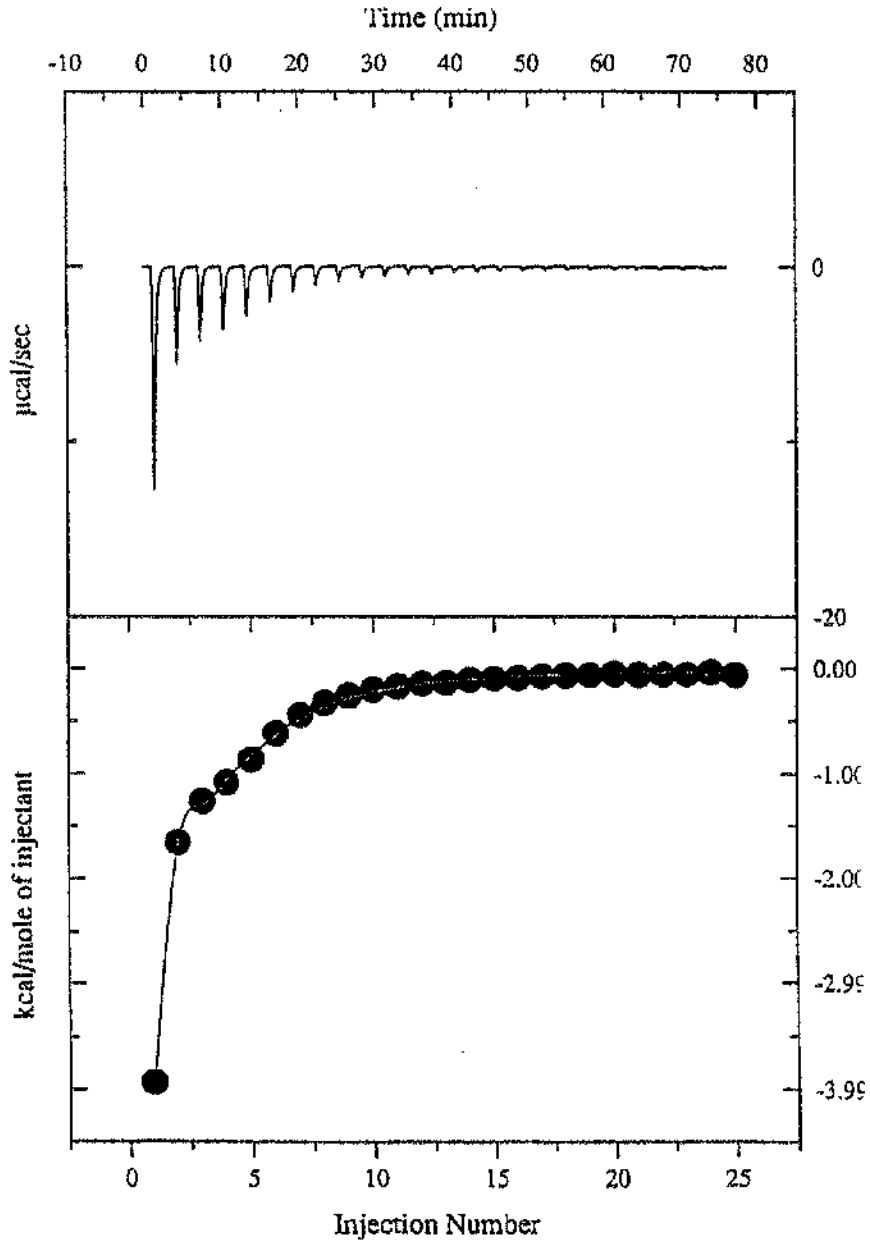

Figure 7.7 ITC data for titration of AhrC (0.1398 mM) with L-Arg (10.02 mM) showing rapid saturation of one strong set (within one or two injections) of sites followed by slow saturation of a second weaker set (20mM tris pH 7.4, 10 mM $MgCl_2$, 250 mM NaCl, 1 mM DTT).

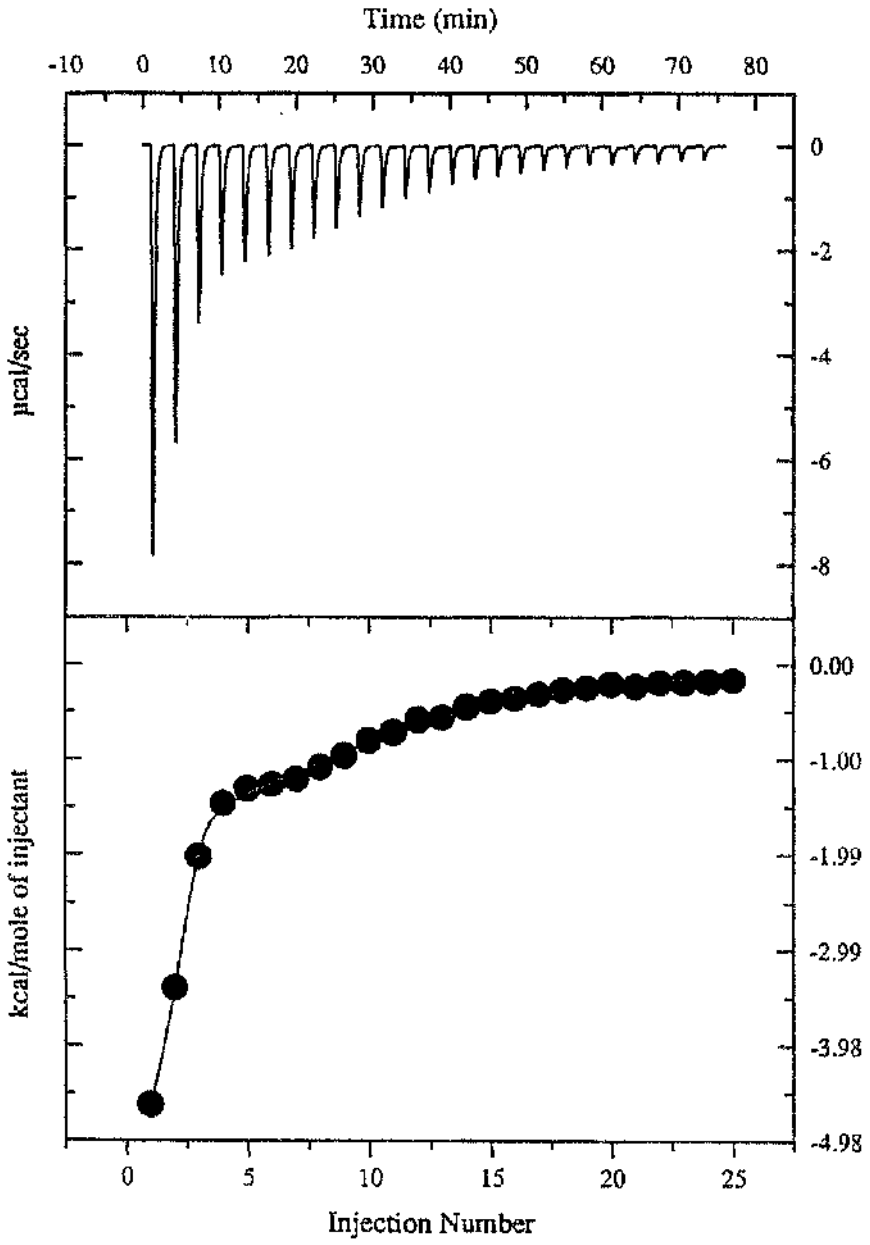

Figure 7.8 ITC data for titration of AhrC (5.02 mM) with L-Arg (10.02 mM) showing less rapid saturation of the strong set of sites allowing some limited analysis to be carried out (buffer conditions as figure 7.7).

The size of injections drops off rapidly at first but after the first three the rate of decrease in size falls, indicating that there are multiple sets of sites. It appears from simple observation of the data that the AhrC hexamer has one set of high affinity Arg binding sites which rapidly becomes saturated and one or more sets of non-specific, weaker binding sites. The data were analysed by model fitting using the Origin software package. The model used was one for two sets of non-interacting sites. Unfortunately the strong binding site becomes saturated too quickly for a good model fit to be attained, however, the following fitting parameters were obtained:

Site	N	K_h (M^{-1})	ΔH(kJmol^{-1})
1	2.62	10^6	-21.13
2	15.1	9812	-8.00

TABLE 7.2 Summary of ITC fitting parameters for Arg binding to AhrC.

Since the strong binding site became saturated so quickly it was decided to repeat the titration, but with much reduced Arg concentration. For purely pragmatic reasons it was decided to also reduce the concentration of AhrC used (it was all of the sample that was left!). This enabled us to investigate how little material was required for future experiments. The concentration of AhrC in the cell was 0.045 mM monomer and the concentration of Arg in the syringe was 0.57 mM, an effective 3 fold dilution of Arg compared when with the previous experiment. Analysis of the titration isotherm (figure 7.9) from this experiment is not realistic since it consists of binding to two sets of sites, but the second set is not

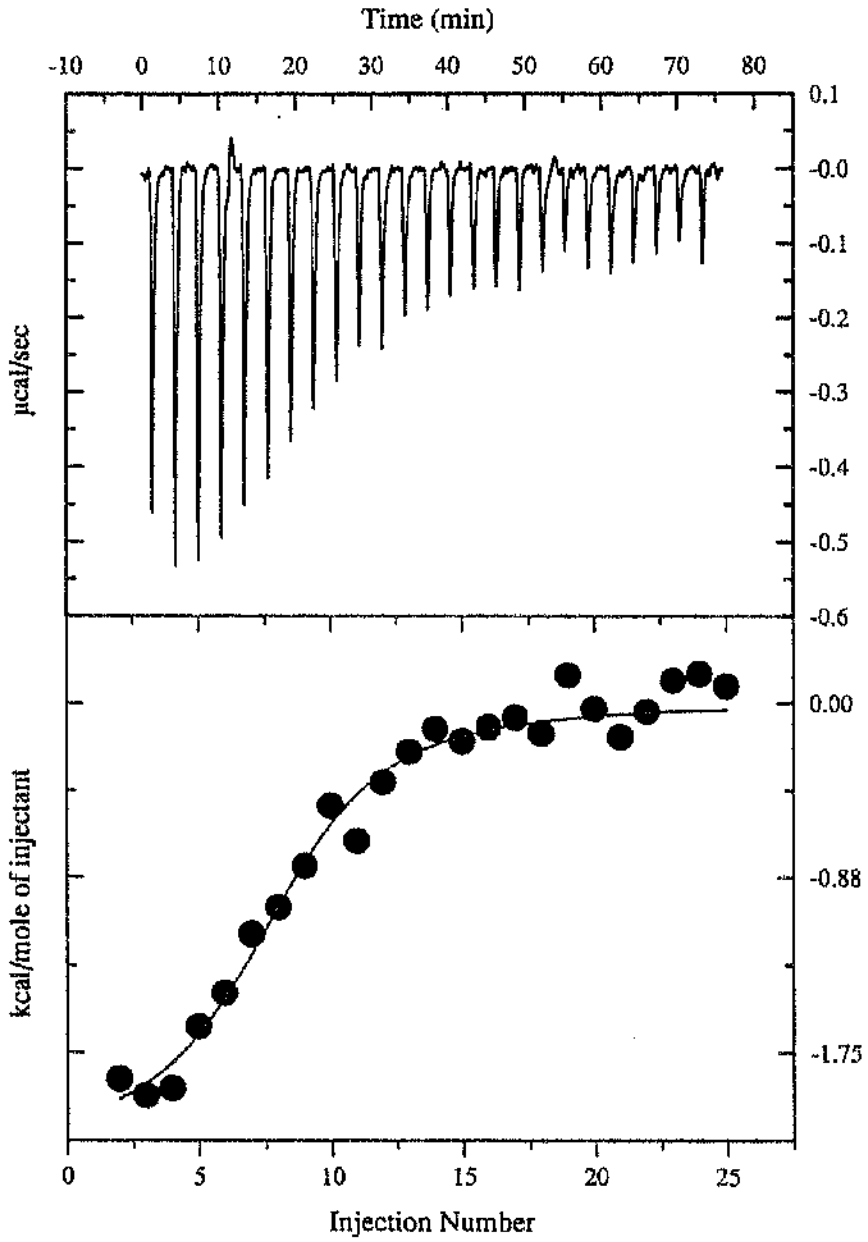

Figure 7.9 ITC data for titration of AhrC (0.045 mM) with L-Arg (0.57 mM) showing only rapid saturation of the strong set of sites (buffer conditions as figure 7.7).

saturated by the end of the experiment. This experiment is useful in that it allows us to see what may be the most fruitful course for future experiments. The data are relatively noisy, as would be expected at such low concentrations, but the level of noise is not so bad as to preclude future experiments at similar concentrations.

7.2.3 Discussion of AhrC results

While the values obtained by model fitting to the data were less than satisfactory some useful, qualitative observations can be made. It is clear that the AhrC hexamer has two sets of Arg binding sites. There appears to be one set of high affinity sites with a binding stoichiometry of three Arg to one hexamer. The homologous ArgR protein from *E. coli* has been found to have this binding stoichiometry also. AhrC also appears to have a large number of, probably, non-specific Arg binding sites.

This work has shown that ITC could be usefully employed to give data on the binding of Arg to the regulatory protein, AhrC, even at relatively low protein concentrations. Any future experiments will have to strike a careful balance in order to account for the rapidity with which the high affinity sites become saturated and the need to carry on the titration until the weaker sites reach saturation. It may be that experiments with slightly more concentrated Arg (relative to AhrC) and a larger number of injections of smaller volume would yield data which can be more accurately fitted by the model. Once the conditions are achieved which allow complete analysis of the data experiments with mutant forms of the protein should give a great deal of information. DSC experiments were

attempted here with no success, unfolding endotherms were completely obscured by non-equilibrium exothermic processes. If, however, appropriate conditions could be found DSC of AhrC with and without Arg may also provide useful data.

"What is written is merely the dregs of experience."

Franz Kafka.

Chapter Eight

Concluding remarks

The aims of this thesis were to study protein interactions and stability using the ultra sensitive microcalorimeteric techniques of DSC and ITC. Studies of the viral coat protein of the bacteriophage MS2 proved to be hampered by the need for accurate estimates of protein concentration, which were not possible with this system at the time. None the less qualitative information on the thermal stability of the protein was obtained which was in good agreement with published work. DSC gave a T_m for the coat protein of 69°C. Preliminary ITC experiments showed that titration of intact capsids with specific RNA oligonucleotide could give data on the binding, and that the RNA rapidly found its way into the interior of the coat protein capsid.

ITC was also carried out on another translational repressor system, the Arg repressor of *B. subtilis* AhrC. It was shown that the AhrC hexamer has three equivalent, high affinity Arg binding sites and a large number of possibly non-specific, low affinity binding sites. In both the MS2 and AhrC systems the results obtained here have helped to direct future effort.

Studies were also carried out on an ss-DNA binding protein from the filamentous bacteriophage Pf1; the so called gene five protein. It was found that the flexible C-terminal

domain of the protein was neither thermodynamically beneficial nor detrimental for binding to ss-DNA, rather by implication it was suggested that the tail performed a kinetic role, forming a clamp over the ss-DNA. The unusual concentration dependence of Pf1-gVc's unfolding transition and the asymmetry of the transition lead to the conclusion that the complex undergoes a decrease in the degree of oligomerization upon unfolding, and that the size of the cooperative unit was itself dependent on concentration.

Studies of the isotope effect on the thermal unfolding of lysozyme found that the protein was significantly more stable in D_2O, and it was tentatively suggested that this stabilisation was entropic in origin and that therefore the isotopic influence on the hydrophobic effect was the dominant factor in the stabilisation. Along the way some observations on the partial irreversibility of thermal unfolding were made which may prove useful for future DSC work. It was noticed that the fraction of the protein which refolds depends on the effective concentration in the cell.

Further studies on protein stability focused on the role of the disulfide bridge. It was found that while the 6-120 disulfide bridge of BLA stabilises the protein entropically, as predicted by theory, it does so to a lesser extent than expected. This is explained in terms of the effect removal has on the dynamics and therefore absolute entropy of the native state, which it is generally assumed is unaffected by the removal of the disulfide. Fluorescence quenching showed that in fact the 3SS-BLA is significantly more flexible that the wild type, a result which is in conflict with the findings from the crystal state of lysozyme and a mutant lacking the 77-95 disulfide bridge (Kidera et al, 1994).

Microcalorimetry, then, is a very effective tool for the study of protein stability and interactions, particularly when the results are taken hand in hand with those from other techniques such as fluorescence spectroscopy.

"It was a book to kill time for those who like it better dead."

Rose Macauley.

Appendix

Abbreviations

1 The amino acids

amino acid	3 letter code	1 letter code	amino acid	3 letter code	1 letter code
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamic acid	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

2 other abbreviations

AMP	ammonium persulphate
BLA	α-lactalbumin.
WT-BLA	wild type α-lactalbumin.
3SS-BLA	three disulfide α-lactalbumin.
apo-BLA	apo-α-lactalbumin.
DSC	differential scanning calorimetry.
DTT	dithiothreitol.
EDTA	ethylenediaminetetra acetic acid.
Hepes	N-[2-hydroxyethyl]piperazine-N'-[2-ethane sulfonic acid]
HEWL	hen egg white lysozyme.
IPTG	isopropyl-β-D-thiogalactoside.
ITC	isothermal titration calorimetry.
PMSF	Phenylmethylsulfonylfluoride.
TEMED	N,N,N',N'-tetramethyl-diamine
TPCK	N-tosyl-L-phenylalanine-methyl ketone
TRIS	tris(hydroxymethyl)methylamine

"From the moment I picked up your book until I laid it down again I was convulsed with laughter. Some day I intend reading it."

Groucho Marx.

References.

Alberts, B., Frey, L. & Delius, H. (1972) *J. Mol. Biol.*, **68**, 139.

Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J. & Dobson, C.M. (1993) *Biochemistry*, **32**, 1707 .

Anfinsen, C.B. & Scheraga, H.A. (1975) *Advan. Protein Chem.*, **29**, 205.

Aronow, R.H. & Witten, L. (1960) *J. Phys. Chem.*, **64**, 1643.

Badley, R.A. (1975) *Biochim. Biophys Acta*, **379**, 517.

Bae, S.-J., Chou, W.-Y., Matthews, K. & Sturtevant, J.M. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6731.

Baldwin, R.L. (1993) *Curr. Opin. Struct. Biol.*, **3**, 84.

Barlow, D.J. & Thornton, J.M. (1983) *J. Mol. Biol.*, **168**, 867.

Beckett, D. & Uhlenbeck, O.C. (1988) *J. Mol. Biol.*, **204**, 927.

Ben-Naim, A., Wilf, J. & Yaacobi, M. (1973) *J. Phys. Chem.*, **77**, 95.

Bernardi, A. & Spahr, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 3033.

Betz, S.F. (1993) *Protein Science*, **2**, 1551.

Brennan, R.G. & Matthews, B.W. (1989) *J. Biol. Chem.*, **264**, 1903.

Brew, K., Vanaman, T.C. & Hill, R.L. (1967) *J. Biol. Chem.*, **242**, 3747.

Brew, K., Vanaman, T.C. & Hill, R.L. (1968) *Proc. Natl. Acad. Sci. U.S.A.*, **59**, 491.

Brew, K., Castellino, F.J., Vanaman, T.C. & Hill, R.L. (1970) *J. Biol. Chem.*, **245**, 4570.

Brinton, C.C., Gemski, P. & Carnahan, J. (1964) *Proc. Natl. Acad. Sci. U.S.A.*, **52**, 776.

Brooks, C.L., Karplus, M. & Pettitt, B.M. (1988) *Adv. Chem. Phys.*, **71**,

Browne, W.J., North A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C. & Hill, R.L. (1969) *J. Mol. Biol.*, **42**, 65.

Bryant, D.T.W. & Andrews, P. (1983) *Biochem. J.*, **220**, 617.

Campbell, I.D. & Dwek, R.A. (1984) *Biological Spectroscopy*, Cummings, California, pp 91-121.

Chand, A. & Fenby, D.V. (1978) *J. Chem. Thermodyn.*, **10**, 997.

Chowdhry, B.Z. & Cole, S.C. (1989) *TIBTECH*, **7**, 11.

Cohn, E.J., McMeekin, T.L., Edsall, J.T. & Blanchard, M.H. (1933) *J. Biol. Chem.*, **100**, xxviii.

Connelly, P.R., Thomson, J.A., Fitzgibbon, M.J. & Bruzzese, F.J. (1993) *Biochemistry*, **32**, 5583.

Cooper, A. (1976) *Proc. Natl. Acad. Sci. U.S.A.*, **73**, 2740.

Cooper, A. (1980) *Science Progress*, **66**, 473.

Cooper, A. (1984) *Prog. Biophys. Molec. Biol.*, **44**, 181.

Cooper, A. (1992) *J. Am. Chem. Soc.*, 114, 9208.

Cooper, A., Eyles, S.J., Radford, S.E. & Dobson, C.M. (1992) *J. Mol. Biol.*, **225**, 930.

Cooper, A. & Johnson, C (1992) in *Methods in Molecular Biology: Physical Methods of Analysis*. Ed Jones C., Mulloy, B. & Thomas, A.H., Humana Press, Clifton, N.J.

Cooper, A., McAlpine, A.S. & Stockley, P.G. (1994) *FEBS Lett.*, **348**, 41.

Covington, A.K., Paabo, M., Robinson, R.A. & Bates, R.G. (1968) *Anal. Chem.*, **40**, 700.

Crawford, E.M. & Gesteland, R.F. (1964) *Virology*, **22**, 165.

Creighton, T.E. (1975) *J. Mol. Biol.*, **96**, 767.

Creighton, T.E. (1983) *Biopolymers*, **22**, 49.

Creighton, T.E. & Goldenberg, D.P. (1984) *J. Mol. Biol.*, **179**, 497.

Creighton, T.E. (1988) *Bioessays*, **8**, 57.

Czaplewski, L.G., North, A.K., Smith, M.C.M., Baumberg, S. & Stockley, P.G. (1992) *Mol. Microbiol.*, **6**, 267.

Damaschun, G., Gernat, C., Damaschun, H., Bychokova, V.E. & Ptitsyn, O.B. (1980) *Int. J. Biol. Macromol.*, **8**, 226.

Davis, K.G., Plyte, S.E., Robertson, S.R., Cooper, A. & Kneale, G.G. (1995) *Biochemistry*, **34**, 148.

Desmet, J., Hanssens, I. & van Cauwelaert, F. (1987) *Biochim. Biophys. Acta*, **912**, 211.

Dill, K.A. (1990) *Biochemistry*, **29**, 7133.

Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychokova, V.E., Semisotnov, G.V., Venyaminov, S.Y. & Ptitsyn, O.B. (1981) *FEBS Lett.*, **136**, 311.

Dolgikh, D.A., Abaturov, L.V., Bolotina, I.A., Brazhnikov, E.V., Bushvev, V.N., Bychokova, V.E., Gilmanshin, R.I., Levedev, Y.O., Semisotnov, G.V., Tiktopulo, E.I. & Ptitsyn, O.B. (1985) *Eur. Biophys. J.*, **13**, 109.

Doig, A.J. & Williams, D.H. (1991) *J. Mol. Biol.*, **217**, 389.

Doig, A.J. & Williams, D.H. (1992) *J. Am. Chem. Soc.*, **114**, 338.

Ebner, K.E., Denton, W.L. & Brodbeck, V. (1966) *Biochem. Biophys. Res. Commun.*, **24**, 232.

Eftink, M. & Biltonen, R. (1980) In *Biological Microcalorimetry*, Ed Beezer, A.E., Academic Press, New York, pp 343-412.

Eftink, M.R. & Ghiron, C.A. (1976a) *Biochemistry*, **15**, 672.

Eftink, M.R. & Ghiron, C.A. (1976b) *J. Phys. Chem.*, **80**, 486.

Eftink, M.R. & Ghiron, C.A. (1981) *Anal. Biochem.*, **114**, 199.

Eftink, M.R. & Ghiron, C.A. (1984) *Biochemistry*, **23**, 3891.

Ewbank, J.J. & Creighton, T.E. (1991) *Nature*, **350**, 518.

Ewbank, J.J. & Creighton, T.E. (1993) *Biochemistry*, **32**, 3694.

Eyles, S.J., Radford, S.E., Robinson, C.V. & Dobson, C.M. (1994) *Biochemistry*, **33**, 13038.

Eyring, H. & Stearn, A.E. (1939) *Chem. Rev.*, **24**, 253.

Finer-Moore, J.S., Kossiakoff, J.A., Hurley, J.H., Earnest, T. & Stroud, R. (1992) *Proteins: Struct Funct Genet.*, **12**, 203.

Fersht, A.R., Shi, J.-P., Jack, K.-J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y. & Winter, G. (1985) *Nature*, **314**, 235.

Fiers, W., Contreras, R., Duerinck, F., Haegmean, G., Merregaert, J., Min Jou, W., Raeymakers, A., Volckaert, G., Ysebaert, M., Van de Kerckhove, J., Nolf, F. & Van Montagu, M. (1975) *Nature*, **256**, 273.

Fiers, W., Contreras, R., Duerinck, F., Haegmean, G., Iserentant, D., Merregaert, J., Min Jou, W., Malemans, A., Raeymakers, A., Van den Berghe, A., Volckaert, G. & Ysebaert, M. (1976) *Nature*, **260**, 500.

Finklestein, A.V. & Shakhnovich, E.I., (1989) *Biopolymers*, **28**, 1681.

Flory, P.J. (1956) *J. Am. Chem. Soc.*, **78**, 5222.

Fleming, A (1922) *Proc. R. Soc. London, Ser. B*, **93**, 306.

Ford, S.J., Cooper, A., Hecht, L., Wilson, G. & Barron, L.D. (1995) *J. Chem. Soc. Farady Trans.*, **91**, 2087.

Frank, H.S. & Evans, M.W. (1945) *J. Chem. Phys.*, **13**, 507.

Fried, M. (1989) *Electrophoresis*, **10**, 366.

Fukada, H., Sturtevant, J.M. & Quiocho, F.A. (1983) *J. Biol. Chem.*, **258**, 13193.

Gast, K., Ziwer, D., Welfe, H., Bychokova, V.E. & Ptitsyn, O.B. (1986) *Int. J. Biol. Macromol.*, **8**, 1419.

Gill, S.C. & von Hippel, P.H. (1989) *Anal. Biochem.*, **182**, 319.

Goldenberg, D.P. & Creighton, T.E. (1985) *Biopolymers*, **24**, 167.

Golmohammadi, R., Valegård, K., Fridborg, K. & Liljas, L. (1993) *J. Mol. Biol.*, **234**, 620.

Gray, C.W., Kneale, G.G., Leonard, K.R., Siegrist, H. & Marvin, D.A. (1982) *Virology*, **116**, 40.

Gray, C.W. (1989) *J. Mol. Biol.*, **208**, 57.

Haynie, D.T. & Freire, E. (1994) *Anal. Biochem.*, **216**, 33.

Harrison, S.C. & Aggarwal, A.K. (1990) *Annu. Rev. Biochem.*, **59**, 933.

Heisenberg, M.J. (1966) *J. Mol. Biol.*, **17**, 136.

Hermans, J. & Scheraga, H.A. (1961) *J. Am. Chem. Soc.*, **83**, 3283.

Hiraoka, Y., Segawa, T., Kuwajima, K., Sugai, S. & Murai, N. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1098.

Hiraoka, Y. & Sugai, S. (1984) *Int. J. Peptide Protein Res.*, **24**, 874.

Hiraoka, Y. & Sugai, S. (1985) *Int. J. Peptide Protein Res.*, **26**, 252.

Hohn, T. (1969) *Eur. J. Biochem.*, **8**, 552.

Hollecker, M. & Creighton, T.E. (1982) *Biochim. Biophys. Acta*, **701**, 395.

Howarth, O.W. (1975) *J. Chem. Soc. Faraday Trans. 1*, **71**, 2303.

Hung, P.P., Ling, C.M. & Overby, L.R. (1969) *Science*, **166**, 1638.

Ikeguchi, M., Kuwajima, K. & Sugai, S. (1986) *J. Biochem.*, **99**, 1191.

Ikeguchi, M., Sugai, S., Fujino, M., Sugawara, T. & Kuwajima, K. (1992) *Biochemistry*, **31**, 12695.

Iyer, K.S. & Klee, W.A. (1973) *J. Biol. Chem.*, **248**, 707.

Johnson, C.M., Cooper, A. & Stockley, P.G. (1992) *Biochemistry*, **31**, 9717.

Kalinichenko, L.P. & Lobyshev, V.P. (1976) *Stud. Biophys.*, **58**, 235.

Karplus, M., Ichiye, T. & Pettitt, B.M. (1987) *Biophys. J.*, **52**, 1083.

Kauzmann, W. (1959) *Adv. Protein Chem.*, **14**, 1.

Kidera, A. & Go, N. (1990) *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 3718.

Kidera, A. & Go, N. (1992) *J. Mol. Biol.*, **225**, 457.

Kidera, A., Inaka, K., Matsushima, M. & Go, N. (1992) *J. Mol. Biol.*, **225**, 477.

Kidera, A., Inaka, K., Matsushima, M. & Go, N. (1994) *Protein Sci.*, **3**, 92.

Kiefhaber, T., Labhardt, A.M. & Baldwin, R.L. (1995) *Nature*, **375**, 513.

Klotz, I.M. & Franzen, J.S. (1962) *J. Am. Chem. Soc.*, **84**, 3461.

Kneale, G.G., Freeman, R. & Marvin, D.A. (1982) *J. Mol. Biol.*, **156**, 279.

Kneale, G.G. & Marvin, D.A. (1982) *Virology*, **116**, 53.

Kneale, G.G. (1992) *Curr. Opin. Struct. Biol.*, **2**, 124.

Kossiakoff, J.A., Sintchak, M.D., J.H., Shpungin, J. & Presta, L.G. (1992) *Proteins: Struct Funct Genet.*, **12**, 223.

Kresheck, G.C., Schneider, H. & Scheraga, H.A. (1965) *J. Phys. Chem.*, **69**, 3132.

Kuroki, R., Inaka, K., Taniyama, Y., Kidokoro, S., Matsushima, M., Kikuchi, M. & Yutani, K. (1992a) *Biochemistry*, **31**, 8323.

Kuroki, R., Nitta, K. & yutani, K. (1992b) *J. Biol. Chem.*, **267**, 24297.

Kuroki, R., Kawakita, S., Nakamura, H. & yutani, K. (1992c) *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6803.

Kuwajima, K., Nitta, K., Yoneyama, M. & Sugai, S. (1976) *J. Mol. Biol.*, **106**, 359.

Kuwajima, K. (1977) *J. Mol. Biol.*, **114**, 241.

Kuwajima, K., hiraoka, Y., Ikeguchi, M. & Sugai, S. (1985) *Biochemistry*, **24**, 874.

Kuwajima, K. (1989) *Proteins: Struct. Funct. Genet.*, **6**, 87.

Kuwajima, K., Ikeguchi, M., Sugawara, T., Hiraoka, Y. & Sugai, S. (1990) *Biochemistry*, **29**, 8240.

Lala, A.K. & Kaul, P. (1992) *J. Biol. Chem.*, **267**, 19914.

Lakowicz, J.R. (1983) *Principles of Fluorescence Spectroscopy*. Plenum Press, New York.

Lehrer, S.S. (1971) *Biochemistry*, **10**, 3254.

Luisi, B.F. & Sigler, P.B. (1990) *Biochim. Biophys Acta*, **1048**, 113.

Maeda, K., Kneale, G.G., Tsugita, A., Short, N.J., Perham, R.N., Hill, D.F. & Peterson, G.B. (1982) *EMBO J.*, **1**, 255.

Manly, S.P., Matthews, K.S. & Sturtevant, J.M. (1985) *Biochemistry*, **24**, 3842.

Mao, C., Carlson, N.G. & Little, J.W. (1994) *J. Mol. Biol.*, **235**, 532.

McAuley-Hecht, K.E. (1993) *Thermodynamics of Biomolecular Recognition*. Ph.D. thesis, University of Glasgow.

McCammon, J.A. & Harvey, S.C. (1987) *"Dynamics of Proteins and Nucleic Acids"*. Cambridge Univ. Press.

McKenzie, H.A. & White, F.H. (1991) *Adv. Protein Chem.*, **41**, 173.

Meot-Nir, M. & Sieck, L.W. (1986) *J. Am. Chem. Soc.*, **108**, 7444.

Michel, B. & Zinder, N.D. (1989) *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 4002.

Min-Jou, W., Haegeman, G., Ysebaert, M. & Fiers, W. (1972) *Nature*, **237**, 82.

Mirsky, A.E. & Pauling, L. (1936) *Proc. Natl. Acad. Sci. U.S.A.*, **22**, 439.

Murakami, K., Andree, P.J. & Berliner, L.J. (1982) *Biochemistry*, **22**, 5061.

Nemethy, G. & Scheraga, H.A. (1964) *J. Phys. Chem.*, **41**, 680.

Nitta, K., Tsuge, H. & Iwamoto, H. (1993) *Int. J. Peptide Protein Res.*, **41**, 118.

North, A.K., Smith, M.C.M. & Baumberg, S. (1989) *Gene*, **80**, 29.

Pabo, C.O. & Sauer, R.T. (1984) *Annu. Rev. Biochem.*, **53**, 293.

Pabo, C.O. & Sauer, R.T. (1992) *Annu. Rev. Biochem.*, **61**, 1053.

Pace, C.N., Grimsley, G.R., Thomson, J.A. & Barnett, B.J. (1988) *J. Biol. Chem.*, **263**, 11820.

Pauling, L. (1939) *The Nature of the Chemical Bond*. Ithaca, NY: Cornell Univ. Press.

Permyakov, E.A., Yarmolenko, V.V., Kalinichenko, L.P., Morozova, L.A. & Burstein, E.A. (1981) *Biochem. Biophys. Res. Commun.*, **100**, 191.

Persson, K.M. & Gekas, V. (1994) *Process Biochem.*, **29**, 89.

Pfeil, W. & Privalov, P.L. (1976) *Biophys. Chem.*, **4**, 41.

Pfeil, W. & Sadowski, M.L. (1985) *Stud. Biophys.*, **109**, 163.

Phillips, S.E.V., Manfield, I., Parsons, I., Davidson, B., Rafferty, J.B., Somers, W.S., Margarita, D., Cohen, G.N., saint-Girons, I. & Stockley, P.G. (1989) *Nature*, **341**, 711.

Pickett, G.G. & Peabody, D.S. (1993) *Nucleic Acids Res.*, **21**, 4621.

Plyte, S.E. & Kneale, G.G. (1991) *Protein Eng.*, **4**, 553.

Plyte, S.E. & Kneale, G.G. (1993) *Biochemistry*, **32**, 3623.

Poland, D.C. & Scheraga, H.A. (1965) *Biopolymers*, **3**, 379.

Pratt, D., Laws, P., & Griffiths, J. (1974) *J. Mol. Biol.*, **82**, 425.

Privalov, P.L. (1974) *FEBS Lett.*, **40**, supplement, S140.

Privalov, P.L. & Potekhin, S.A. (1986) *Meth. Enzymol.*, **131**, 4.

Privalov, P.L. & Gill, S.J. (1988) *Adv. Protein Chem.*, **39**, 191.

Ptitsyn, O.B., (1992) in *Protein Folding*, Ed Creighton, T.E., Freeman, New York, pp 243-300.

Radford, S.E., Woolfson, D.N., Martin, S.R., Lowe, G. & Dobson, C.M. (1991) *Biochem. J.*, **273**, 211.

Rose, G.D. & Wolfenden, R. (1993) *Annu. Rev. Biophys. Biomol. Struct.*, **22**, 381.

Sauer, R.T., Jordan, S.R. & Pabo, C.O. (1990) *Adv. Protein Chem.*, **40**, 1.

Schellman, J.A. (1955) *C. R. Trav. Lab. Carsburg Ser. Chim.*, **29**, 223.

Schmidt, D.V. & Ebner, K.E. (1971) *Biochim. Biophys Acta*, **243**, 273.

Segawa, T. & Sugai, S. (1983) *J. Biochem.*, **93**, 1321.

Shakhnovich, E.I. & Finklelstein, A.V., (1989) *Biopolymers*, **28**, 1667.

Shechter, Y., Patchornik, A. & Burstein, Y. (1973) *Biochemistry*, **12**, 3407.

Shimizu, A., Ikeguchi, M. & Sugai, S. (1993) *Biochemistry*, **32**, 13198.

Shirley, B.A., Stanssens, P., Hahn, U. & Pace, C.N. (1992) *Biochemistry*, **31**, 725.

Smith E.B. (1992) *Basic Chemical Thermodynamics (Fourth Edition).* Oxford University Press, Oxford.

Somers. W.S. & Phillips, S.E.V. (1992) *Nature*, **346**, 586.

States, D.J., Dobson, C.M., Karplus, M. & Creighton, T.E. (1984) *J. Mol. Biol.*, **174**, 411.

States, D.J., Creighton, T.E., Dobson, C.M. & Karplus, M. (1987) *J. Mol. Biol.*, **195**, 731.

Stonehouse, N.J. & Stockley, P.G. (1993) *FEBS Lett.*, **334**, 355.

Sturtevant, J.M. (1974) *Ann. Rev. Biophys. Bioeng.*, **3**, 35.

Sturtevant, J.M. (1977) *Biochemistry*, **74**, 2236.

Sturtevant, J.M. (1987) *Ann. Rev. Phys. Chem.*, **38**, 463.

Sugiyama, T. & Nakada, D. (1967) *Biochemistry*, **57**, 1744.

Sun, M. & Song, P.S. (1977) *Photochem. Photobiol.*, **25**, 3.

Tanford, C. (1968) *Adv. Protein Chem.*, **23**, 121.

Teale, F.W.J. & Badley, R.A. (1970) *Biochem. J.*, **116**, 341.

Traube, J. (1891) *Liebigs Ann. Chem.*, **265**, 27.

Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1990) *Nature*, **345**, 36.

Valegård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. & Liljas, L. (1994) *Nature*, **371**, 623.

Vandenberghe, A., Min-Jou, W. & Fiers, W. (1975) *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 2559.

Wilson, G., Ford, S.J., Cooper, A., Hecht, L., Wen, Z.Q. & Barron, L.D. (1995) *J. Mol. Biol.*, **254**, 747.

Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984) *J. Am. Chem. Soc.*, **106**, 765.

Xie, D., Bhakuni, V. & Freire, E. (1991) *Biochemistry*, **30**, 10673.

Xie, D., Bhakuni, V. & Freire, E. (1993) *J. Mol. Biol.*, **232**, 5.

Yang, A.-S. & Honig, B. (1993) *J. Mol. Biol.*, **231**, 459.

Yutani, K., Ogasahara, K. & Kuwajima, K. (1992) *J. Mol. Biol.*, **228**, 347.