# Studies of Protein Folding by NMR Spectroscopy

Stephen J. Eyles

A thesis submitted to the Board of the Faculty of Physical Sciences (Chemistry) of the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

St Catherine's College
Oxford

Trinity Term
1995

To my parents

# Abstract

## *Studies of Protein Folding by NMR Spectroscopy*

Stephen J. Eyles
St Catherine's College.

D. Phil. Thesis
Trinity Term, 1995

This thesis describes an investigation of the folding and stability of a series of derivatives of the proteins lysozyme and α-lactalbumin which lack one or more of their four native disulphide bridges. Removal of the disulphide bridge which links the N- and C-termini from hen lysozyme results in a three-disulphide derivative ($CM^{6,127}$-lysozyme). This has a profound effect on its stability against thermal denaturation, the $T_m$ for unfolding being reduced by 25°C at pH 3.8. Calorimetric measurements performed on this three-disulphide derivative indicate that this reduction in stability may be attributed entirely to an increase in the entropy difference between the native and denatured states.

Kinetic refolding studies of $CM^{6,127}$-lysozyme using stopped flow optical methods and hydrogen exchange pulse labelling in conjunction with NMR and electrospray ionisation mass spectrometry (ESI-MS) suggest that this reduced stability manifests itself primarily in the α-domain of the protein. A transient intermediate populated during refolding of the unmodified protein can no longer be detected during folding of the derivative resulting in highly cooperative folding under the conditions investigated.

The structure and stability of a three- and two-disulphide derivative of the homologous protein, α-lactalbumin have been investigated by NMR spectroscopy. The three-disulphide species, like its lysozyme counterpart, can adopt native structure but this is much more unstable than the intact protein. Removal of a second disulphide bridge, however, destabilises α-lactalbumin to the extent that the native state is no longer formed. Instead, in the presence of $Ca^{2+}$ and high concentrations of salt, a partially structured state is induced which has some elements of tertiary structure present.

Novel techniques of ESI-MS have been developed to study protein folding and stability using hydrogen exchange techniques. Applications to the investigation of cooperativity in protein folding, stability in native, partially folded and unfolded states, and the interactions of a partially folded protein with the chaperone GroEL are described.

# Table of Contents

## Chapter One

## Chapter Two

# Acknowledgements

First and foremost I would like to thank my supervisor, Dr Chris Dobson for taking me on as a student in the first place and for convincing me on several occasions that protein folding really was interesting in spite of occasional evidence to the contrary. Also to Dr Sheena Radford for her expert help and constant encouragement, with whom many hours were spent keeping equipment up and running, and for valuable discussions, some of which were about science.

Special thanks to the other people who have helped with the experimental aspects of this research. To Alan Cooper in Glasgow for DSC experiments, Michael Groβ who slaved over a cauldron of GroEL and who performed all of the chaperone binding experiments, and of course to Carol Robinson without whom mass spectrometry at such a high level would not be possible in Oxford. Also to Jonathan Boyd, Nick Soffe and Christina Redfield for maintaining the NMR machines and computers in pristine condition.

My thanks must also go to all members of the Dobson and Radford groups, past and present, most of whom drove me completely mad, but many who made the ICL and NCL a great place to work. Thanks for putting up with me. Special thanks to Andrew and of course Kim who, in spite of being American (!) have been great friends. To Caroline who has made endless cups of coffee, sneaked me in for quick meetings with the boss, and generally been totally wonderful throughout my time here. Kev, who I'm sure will never see this deserves thanks too, if only for telling me not to do a D.Phil. in the first place.

Finally to everyone else in Oxford who has made these past few years the best and worst in my life. I am indebted to you all.

# Publications

Publications arising from the work in this thesis:

Cooper A., Eyles S.J., Radford S.E. & Dobson C.M. (1992), "Thermodynamic consequences of the removal of a disulphide bridge from hen lysozyme.", *J. Mol. Biol.*, **225**, 939-943.

Eyles S.J., Radford S.E., Robinson C.V. & Dobson C.M. (1994), "Kinetic consequences of the removal of a disulphide bridge on the folding of hen lysozyme.", *Biochemistry*, **33**, 13038-13048.

Robinson C.V., Groß M., Eyles S.J., Ewbank J.J., Mayhew M., Hartl F.U., Dobson C.M. & Radford S.E. (1994), "Conformation of GroEL-bound α-lactalbumin probed by mass spectrometry.", *Nature*, **372**, 646-651.

Hooke S.D., Eyles S.J., Robinson C.V., Radford S.E. & Dobson C.M. (1995), "Cooperative elements in protein folding monitored by electrospray ionisation mass spectrometry.", *J. Amer. Chem. Soc.*, in press.

# Abbreviations used.

| | |
|---|---|
| $2SS_{cam}$ | bovine α-lactalbumin in which Cys6–Cys120 and Cys28–Cys111 are reduced and carboxyamidomethylated. |
| $3SS_{cam}$ | bovine α-lactalbumin in which Cys6–Cys120 is reduced and carboxyamidomethylated. |
| BLA | bovine α-lactalbumin |
| CD | circular dichroism |
| $CM^{6,127}$-lysozyme | hen lysozyme in which Cys6–Cys127 is reduced and carboxymethylated. |
| COSY | J-correlated spectroscopy |
| DSC | differential scanning calorimetry |
| ESI-MS | electrospray ionisation mass spectrometry |
| GuHCl | guanidine hydrochloride |
| GuDCl | guanidine deuterochloride |
| GPLA | guinea pig α-lactalbumin |
| HEWL | hen egg white lysozyme |
| HL | human milk lysozyme |
| JQL | Japanese quail egg lysozyme |
| LAPL | Lady Amherst pheasant egg lysozyme |
| MeUdiNAG | 4-methylumbelliferyl N,N′ di-N-acetylglucosamine |
| NMR | nuclear magnetic resonance |
| nOe | nuclear Overhauser enhancement |
| ppm | parts per million |
| $T_m$ | midpoint temperature of thermal denaturation |

# Chapter One

# Introduction

## 1.1 *The protein folding problem*

Proteins perform a huge variety of functions, from enzymes with catalytic roles in essential biological cycles, through to structural proteins, from antibodies to virus coatings, even proteins which assist in the folding of other proteins. In spite of this diversity of function, the code within the primary amino acid sequence which determines the structures of these proteins, and the properties which influence the folding and stability of proteins remain largely unknown. Levinthal (1968) postulated that the rate at which proteins folded excluded the possibility of a random search energy minimisation process. At least in the case of folding *in vivo* this may be partly explained with the knowledge that a large number of other factors affect the folding of a nascent polypeptide as it emerges from the ribosome. A group of helper proteins – chaperones – are known to play an important role in preventing aggregation of unfolded and partially folded proteins, assisting in the refolding process and catalysing processes such as disulphide bridge formation and proline isomerisation. To investigate the protein folding process within the cell, however, would obviously require an in depth knowledge of the interaction of all these factors. An obvious simplification, therefore, is to isolate the protein from its natural environment and to study folding *in vitro,* thus allowing the fundamental processes which affect the stability and folding properties of proteins to be elucidated.

Why do we need to understand protein folding? Apart from the obvious academic interest, the ability to determine *de novo* the structure of a protein solely from knowledge of its primary sequence would enable prediction of the structure of a protein from its primary sequence without the need for time consuming structure determination by X-ray crystallography or NMR . More importantly it would enable protein engineers to design protein sequences with specific known structures and hence to design specific substrates and binding sites, or to create enzymes with precisely placed catalytic residues. This could have wide ranging pharmaceutical possibilities, allowing the use of engineered proteins as highly specific and highly effective therapeutic drugs. Knowledge of the events which occur during protein folding and indeed misfolding may also be essential to the understanding of the action of a number of diseases. The causes of conditions such as cystic fibrosis, amyloidosis and Alzheimer's disease have been linked to the presence of misfolded proteins in the body (Thomas *et al.*, 1992; Pepys *et al.*, 1993).

The situation of *de novo* protein design is, however, still some way off. Exact understanding of the forces which influence docking of secondary structure elements and formation of tertiary interactions are required, since a large number of compensating energy terms contribute to the relatively low stability of a folded protein (Creighton, 1992b). Much work has gone into the study of protein databases in attempts to correlate primary sequences with structural motifs, and to allow knowledge based structure prediction (Chelvanayagam *et al.*, 1992; Rooman & Wodak, 1992). It is only possible so far, however, to design proteins with simple structural motifs, such as four helix bundles, but even these tend only to be partially structured (Betz *et al.*, 1993 and references therein). It is still impossible to predict accurately the preferred contact surfaces between regions of secondary structure to avoid incorrect juxtaposition of α-helices and sliding of β-strands (Quinn *et al.*, 1994). For this reason, much work is being concentrated on characterisation of the events which occur during protein folding by study of unfolded and partially folded intermediate states, and the kinetic events on the folding pathway of proteins, to investigate the interactions which are key to formation of native structure.

## *1.2 Techniques for studying protein folding.*

A variety of methods have been established for the study of protein folding. One of the first techniques developed for studying the sequence of events involved in the folding of a protein *in vitro* was to investigate the order of formation of disulphide bridges during the oxidative refolding of a reduced protein sequence (Creighton, 1974, 1975, 1986). By oxidatively refolding reduced bovine pancreatic trypsin inhibitor (BPTI) for varying periods of time, then quenching any further reaction and trapping all remaining free thiols with a reagent such as iodoacetic acid, it was possible to trap incompletely folded species lacking the full complement of native disulphide bridges. This enabled determination of the positions of those disulphide bridges which had formed and of the order of formation of the three native disulphide bonds on the folding pathway. This elegant technique has been used for a number of disulphide containing proteins to elucidate the oxidative folding pathway (Perraudin *et al.*, 1983). More recently, a complementary method has been developed which involves an acid quench mechanism to trap free thiols (Weissman & Kim, 1991). By this technique, folding may be arrested and the product mixture identified by HPLC, then by raising the pH once more, refolding continues, hence allowing the fate of specific intermediates to be investigated.

Alternative methods of studying protein folding may be loosely grouped into two, involving either kinetic studies to identify and characterise intermediates on the refolding pathway which may only exist transiently, or equilibrium methods which involve study of partially folded states populated under equilibrium conditions. These are described below.

### 1.2.1 Kinetic methods.

Advances in stopped flow technology have enabled optical techniques to be applied to the study of rapid events in protein folding. In the past, manual mixing experiments have allowed observation of the slowest steps in folding, but with improved optics and mixing precisely controlled electronically it is now possible to observe directly changes in optical properties of a refolding protein on the millisecond time scale. The mostly widely used optical technique is probably that of circular dichroism which is highly sensitive to small changes in secondary and tertiary interactions in proteins (Bolotina & Lugauskas, 1985; Woody, 1985; Johnson, 1990). For example the ellipticity in the far-UV region at 222 nm is characteristic of α-helix formation and hence may be used to monitor secondary structure formation. A recent study by CD of the kinetic refolding of 14 proteins demonstrated the formation of collapsed intermediates with substantial secondary structure formation at the earliest detectable time points (Kuwajima *et al.*, 1993). This is consistent with the findings of a number of other researchers (Ikeguchi *et al.*, 1986a; Chaffotte *et al.*, 1992; Elöve *et al.*, 1992; Radford *et al.*, 1992b) and lends weight to the hypothesis that this collapsed state is a common intermediate in the folding of globular proteins.(Ptitsyn *et al.*, 1990).

Signals in the near-UV CD region arise from the presence of fixed tertiary interactions (Goux & Hooker, 1980; Bolotina, 1987). Stopped flow techniques monitoring at these wavelengths can be used, therefore, to detect formation of native like tertiary structure later in the folding pathway (Radford *et al.*, 1992b). This may be combined with UV absorbance and intrinsic protein fluorescence techniques to measure the rates of burial of aromatic residues and formation of stable hydrophobic core regions (Itzhaki *et al.*, 1994). Binding of the hydrophobic dye ANS (1-anilinonaphthalenesulphonic acid) has been used as a fluorescent probe to measure the rate of burial of hydrophobic surface during folding (Ptitsyn *et al.*, 1990; Semisotnov *et al.*, 1991), and fluorescently labelled inhibitors may be employed to monitor formation of native substrate binding sites (Itzhaki *et al.*, 1994).

All of these optical techniques can shed a great deal of light on the protein folding process and enable the extraction of order parameters, but they can only yield limited information as to the events occurring on a submolecular level, such as which particular regions of structure are forming at each stage of the folding process. One of the most significant recent advances in the protein folding field has been pioneering work using a technique of hydrogen exchange pulse labelling (Roder *et al.*, 1988; Udgaonkar & Baldwin, 1988) . The principle of this method is to identify regions of structure in the folding protein by virtue of the protection of labile amide hydrogens against isotope exchange with solvent. This may occur as a result of formation of hydrogen bonds in secondary structure, burial of amides in regions of the structure inaccessible to solvent, or formation of longer range hydrogen bonds in tertiary structure.

This technique is limited by the requirement for a fully assigned NMR spectrum of the protein and amides which are highly protected against exchange in the native state, and hence has only been applied to a relatively small number of proteins to date (Baldwin, 1993; Woodward, 1994 and references therein). Nevertheless, the kinetic refolding pathways of a number of small globular proteins have been elucidated, for instance cytochrome *c* (Roder *et al.*, 1988; Jeng & Englander, 1991), ribonuclease A (Udgaonkar & Baldwin, 1988), barnase (Bycroft *et al.*, 1990), *c*-type lysozymes (Radford *et al.*, 1992b; Dobson *et al.*, 1994; Hooke *et al.*, 1994) and interleukin-1β (Varley *et al.*, 1993).

A further recent advance in the use of hydrogen exchange pulse labelling is the use of electrospray ionisation mass spectrometry (ESI-MS) to measure the extent of protection against hydrogen exchange (Miranker *et al.*, 1993). Unlike NMR this technique has not yet yielded residue specific information, but can distinguish refolding populations by differences in mass as a result of the number of deuterium atoms incorporated. ESI-MS may be applied to much larger proteins which are inaccessible to NMR methods and a number of developments are currently underway to further advance the use of this technique, as described in this thesis.

### 1.2.2 Equilibrium methods

An alternative means of investigation is to study the structures of intermediate species formed on the folding pathway. This is often difficult since protein folding is generally a highly cooperative event (Tanford, 1970; Privalov & Khechinashvili, 1974; Pfeil & Privalov, 1976; Privalov, 1979), and stable intermediate species tend not to accumulate in appreciable amounts. There are some proteins, however, which exhibit an intermediate partially folded structure, termed a molten globule (Ohgushi & Wada, 1983), under mild denaturing conditions, at high or low pH or at elevated temperature. These species are stable at equilibrium and may be characterised by optical techniques as well as by NMR (Dobson, 1992). By the very nature of these states being only partially folded, often lacking any fixed tertiary interactions and with only a subset of the full complement of secondary structure, they are very difficult to characterise structurally in detail. Dynamic averaging and sample heterogeneity preclude any studies of these species by X-ray crystallography, but some success has been gained in the study of these states by NMR using indirect methods such as magnetisation transfer and hydrogen exchange, and equilibrium studies by CD and NMR. These methods have been applied to the molten globule states of several proteins including $\alpha$-lactalbumin (Baum *et al.*, 1989; Kuwajima, 1989; Alexandrescu *et al.*, 1993; Chyan *et al.*, 1993), cytochrome *c* (Jeng *et al.*, 1990) and apomyoglobin (Hughson *et al.*, 1990).

Advances in NMR technology have allowed characterisation of a variety of unfolded and partially folded states at equilibrium, particularly using multidimensional heteronuclear methods. With the ability to enrich proteins with $^{15}$N and $^{13}$C isotopes, it has become possible even to assign NMR spectra of unfolded states, whose resonances are generally highly degenerate in the $^{1}$H dimenson (Wüthrich, 1994). This approach has been applied to the identification of regions of residual structure in denatured states in urea of the 434-repressor protein (Neri *et al.*, 1992a) and FK506 binding protein (Logan *et al.*, 1994). More structured regions have been identified by heteronuclear NMR in mutants of staphylococcal nuclease (Shortle & Abeygunawardana, 1993; Alexandrescu *et al.*, 1994), and at the other extreme, local unfolding of structure in an otherwise native state has been observed in interleukin-4 (Redfield *et al.*, 1994) and *apo* cytochrome $b_{562}$ (Feng *et al.*, 1994). An alternative method of inducing formation of a partially folded state is to dissolve proteins in organic cosolvents, the most popular of which is trifluoroethanol (TFE). Such states have been observed to have certain characteristics of molten globule states, namely extensive secondary structure, but in the absence of any fixed tertiary interactions. Partially folded states which have been characterised in this manner include lysozyme (Buck *et al.*, 1993) and α-lactalbumin in TFE (Alexandrescu *et al.*, 1994), ubiquitin in aqueous methanol (Harding *et al.*, 1991), and monellin in ethanol (Fan *et al.*, 1993).

Studies of this type can yield only limited information as to the factors which influence the stability of these partially folded states. One approach which has been used successfully is to dissect complete proteins into smaller parts, using peptides to mimic regions of structure in the intact protein. In the simplest cases these may be peptide analogues of α-helical regions (Goodman & Kim, 1989; Bolin, 1994), β-sheets (Dyson *et al.*, 1992; Yang *et al.*, 1994) and small hydrophobic clusters (Neri *et al.*, 1992b; Smith *et al.*, 1994) detected in unfolded states in order to determine the intrinsic stability of these structures in the absence of the complete protein sequence. Peptides corresponding to larger sections of the structure of myoglobin (Shin *et al.*, 1993a; Shin *et al.*, 1993b; Waltho *et al.*, 1993) and BPTI (Oas & Kim, 1988; Staley & Kim, 1990) and indeed whole subdomains of proteins (Peng & Kim, 1994) have enabled much information to be gained about the stabilising features of protein folding intermediates.

## *1.3 Disulphide bridges*

Besides the peptide backbone, another type of covalent bond is commonly present in many proteins, namely that linking two cysteine residues together to form disulphide bridges. These disulphide bonds can contribute significantly to the stability of proteins, and their effect on the stability and folding properties of proteins has been investigated for a number of proteins The ability to introduce a disulphide bridge into a protein and hence increase its stability clearly would have far reaching consequences, particularly in pharmaceutical applications since a more stable protein should last longer and may be more potent (Creighton, 1988). Typically a disulphide bridge may contribute 8–20 kJ mol$^{-1}$ stability to the folded state of a protein.

The factors which affect the contribution of a disulphide bridge to protein stability are many and varied. The bond is highly directional with strict stereochemical requirements (Steudel, 1975) so any deviation from optimal bond angles and lengths will affect the stability of the protein; natural disulphide bonds obey these requirements almost invariably (Thornton, 1981). Effects on stability also depend on the exact positioning of the disulphide bond within the tertiary structure of the protein. Thus a bridge which is buried within the structure may improve hydrophobic packing, whereas one located nearer to the protein surface may have a primarily entropic effect on the stability by reducing the configurational entropy of the unfolded state. This has recently been reviewed in some depth (Creighton, 1988; Pace, 1990; Doig & Williams, 1991; Zhou *et al.*, 1993; Zhang *et al.*, 1994).

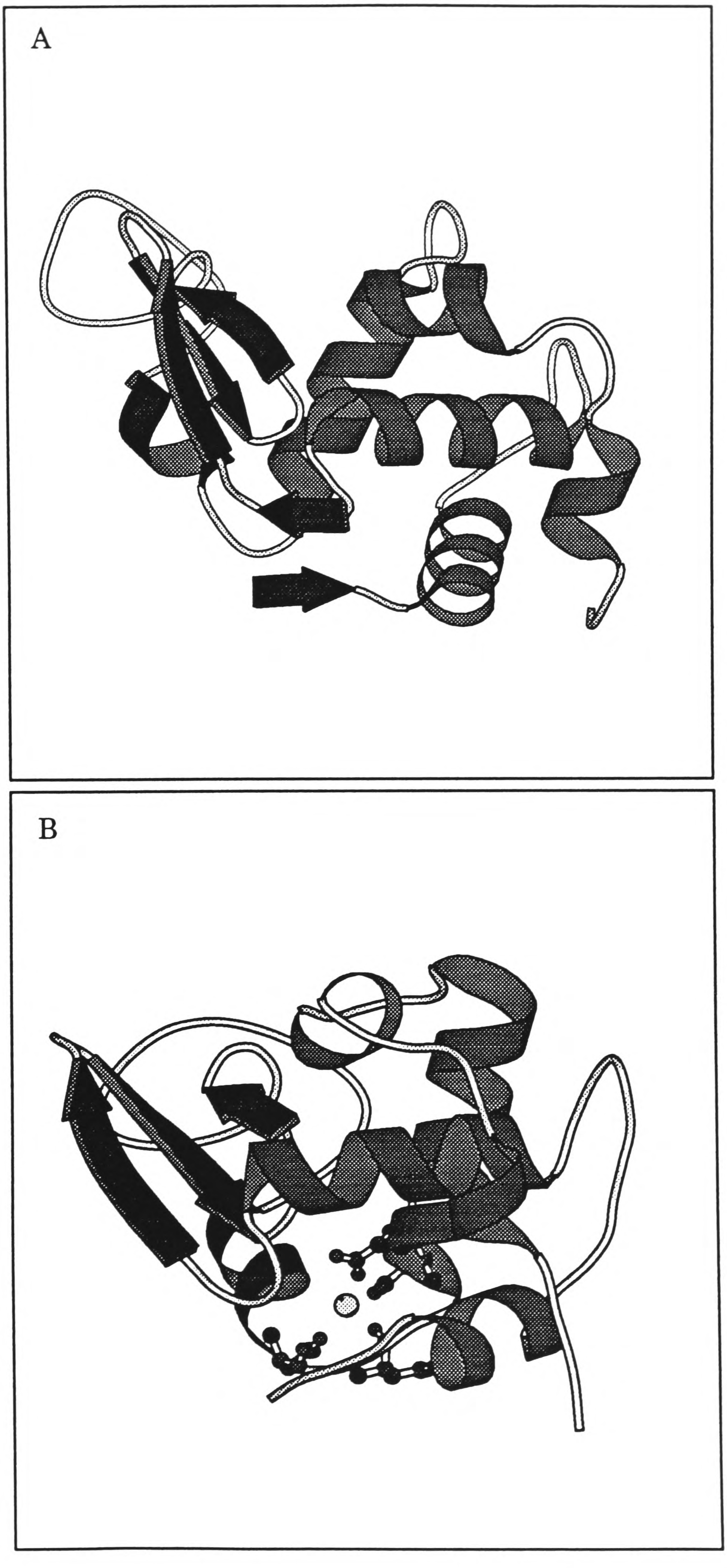
Experiments have been performed to investigate the effect on the stability of proteins both of removing disulphide bridges and introducing them into proteins. Examples of the former include ribonuclease $T_1$ (Pace *et al.*, 1988), hen lysozyme (Denton & Scheraga, 1991; Radford *et al.*, 1991; Cooper *et al.*, 1992), human lysozyme (Kuroki *et al.*, 1992). As one might expect, all of these studies concluded that removal of disulphide bridges resulted in a reduced stability of the native state. Introduction of disulphides into proteins by genetic engineering has often been found to stabilise proteins (Creighton, 1988; Matsumura *et al.*, 1989; Perry & Wetzel, 1991; Clarke & Fersht, 1993), although in some cases no increase in stability was detected, and in some cases the the protein was in fact destabilised (Wells & Powers, 1986; Wetzel, 1987; Wetzel *et al.*, 1988). Similarly, altering the number of disulphides has been shown to alter the kinetic refolding of proteins, often in an unpredictable manner (Denton *et al.*, 1994; Mücke & Schmid, 1994a,b). This once more demonstrates our lack of understanding of some of the fundamental principles of protein folding and the need to investigate further.

## *1.4 Lysozymes and α-lactalbumins.*

This thesis concentrates on the folding properties of a series of small globular proteins, namely lysozymes and α-lactalbumins. These are a highly homologous group of proteins both in sequence and in structure (Blake *et al.*, 1965; Acharya *et al.*, 1991). They have two structural lobes, the α-domain composed of four α-helices, and the β-domain consisting mainly of β-sheet structure (although the β-structure of α-lactalbumin is somewhat less well defined than in lysozyme). Figure 1.1 compares the structures of hen egg white lysozyme and bovine α-lactalbumin, and indicates the location of regions of secondary structure. The cleft formed between these two structural lobes forms the active site cleft in lysozymes, catalysing hydrolysis of β-1,4-glycosidic linkages in bacterial cell walls. α-Lactalbumin is involved in the lactose synthesis reaction, interacting with the enzyme galactosyltransferase. Here the sugar-binding cleft is partially blocked by a tyrosine side chain, and the protein has a high affinity calcium binding loop involving the four aspartate residues Asp82, Asp84, Asp87 and Asp88 (Hill & Brew, 1975; Acharya *et al.*, 1991).

---

**Figure 1.1.** Schematic representations of the three dimensional structures of (A) lysozyme and (B) α-lactalbumin showing the $Ca^{2+}$ binding site. This diagram was produced using the program MolScript (Kraulis, 1991).


A
B

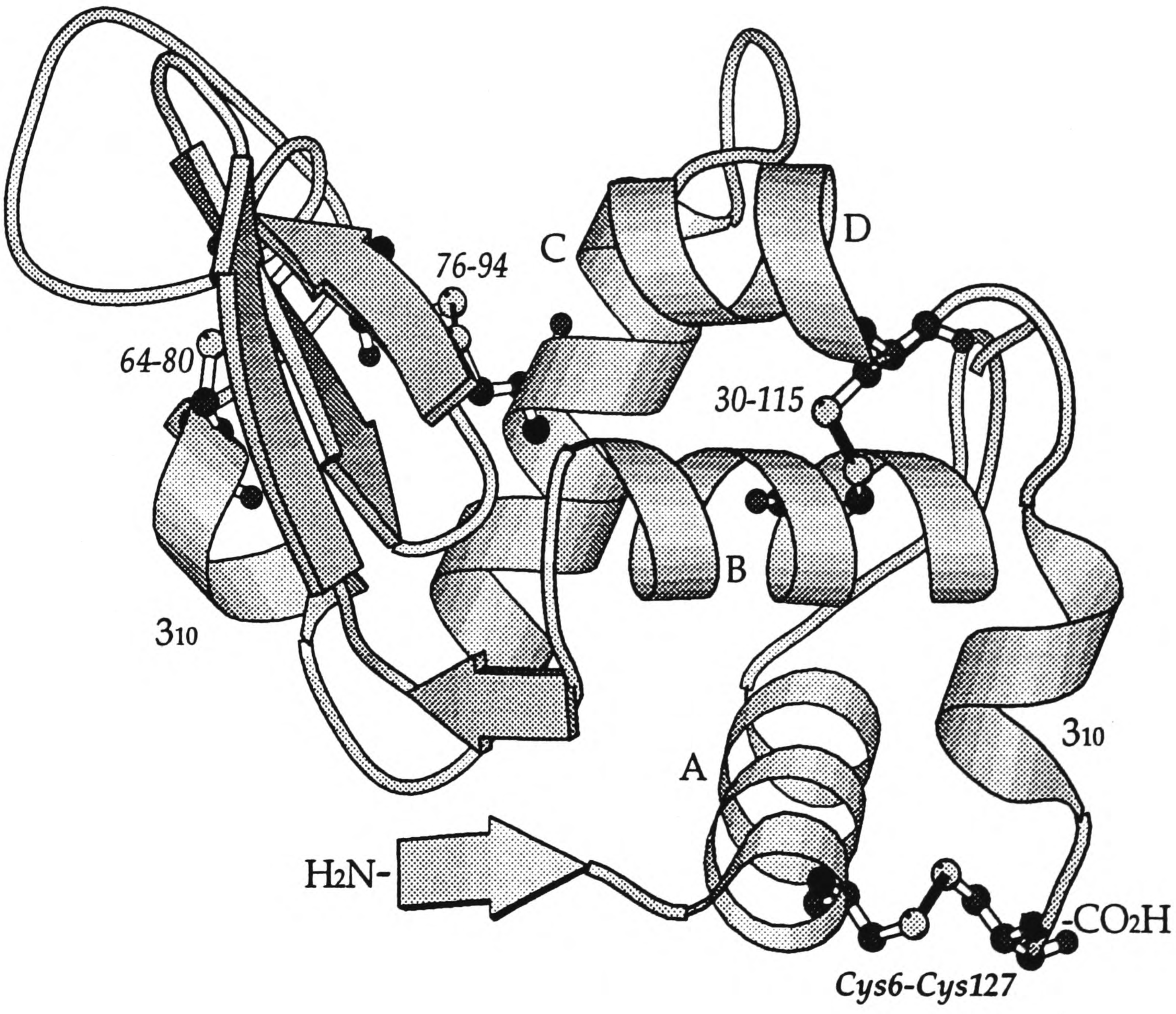


**Figure 1.2.** Schematic diagram of the structure hen lysozyme, showing the positions of the secondary structure elements and the four native disulphide bridges. This diagram was created using the program MolScript (Kraulis, 1991)

These proteins all contain four homologous disulphide bridges which stabilise the folded structure. In hen lysozyme, these disulphides are Cys6–Cys127 and Cys30–Cys115, located in the α-domain; Cys64–Cys80 in the β-domain and Cys76–Cys94 crosslinking the long loop region to the α-domain. The locations of the four native disulphide bridges in hen lysozyme are shown in Figure 1.2.

In spite of the homology in structure and sequence of lysozymes and α-lactalbumins, the folding properties of these two classes of protein show significant differences from one another. Under equilibrium conditions, whereas lysozymes appear in general to obey classical two-state behaviour very closely (Tanford, 1970; Khechinashvili *et al.*, 1973), α-lactalbumin unfolds *via* a partially folded molten globule state (Kuwajima *et al.*, 1976). In kinetic refolding experiments, a similar such intermediate has been transiently detected not only for α-lactalbumin but also for lysozyme (Ikeguchi *et al.*, 1986a), suggesting that their kinetic refolding pathways are similar. Much investigation has taken place to determine the origins of these differences in folding properties; the $Ca^{2+}$ binding affinity of α-lactalbumin appears to play some part (Ikeguchi *et al.*, 1986b), although a more important factor may be the higher incidence of hydrophobic residues as compared to the lysozymes, which may stabilise a molten globule state at equilibrium. Recent results have indicated the existence of partially folded states for some *c*-type lysozymes, suggesting that even this division between lysozymes and α-lactalbumins is not so clear (Morozova *et al.*, 1991; Haezebrouck *et al.*, 1995).

## 1.5 *Outline of this thesis.*

This thesis focuses primarily on the detailed investigation of the stability and folding properties of a derivative of hen lysozyme, whose thermal stability is reduced relative to the wild type protein by the removal of one of the four native disulphide bridges. These characteristics are then compared to similar derivatives of the homologous protein, bovine α-lactalbumin, and to wild type lysozyme destabilised by the addition of chemical denaturant. General methods employed are described in Chapter Two. More specific details of experimentation are described in the remaining chapters.

Chapter Three of this thesis presents the results of stability studies on a three-disulphide derivative hen lysozyme in which a single disulphide bridge has been selectively reduced and chemically blocked ($CM^{6,127}$-lysozyme). This disulphide bridge occurs in the α-domain of the protein and links the N- and C- termini. Global thermal stability was investigated over a range of pH values by differential scanning calorimetry. The stability of local structural regions was studied by measurement of site specific protection of backbone amides against hydrogen exchange, monitored by 2D NMR methods.

A detailed investigation of the kinetic refolding of $CM^{6,127}$-lysozyme is described in Chapter Four. The data presented include kinetic measurements by stopped flow circular dichroism and fluorescence, and hydrogen exchange pulse labelling in conjunction with NMR and ESI-MS. The folding of this protein is compared with that of unmodified lysozyme to determine the effect of removal of this disulphide bridge on the refolding characteristics of the protein.

Chapter Five describes studies by NMR of derivatives of bovine α-lactalbumin. As discussed in this chapter, lysozyme and α-lactalbumin are homologous, both in structure and sequence, and have homologous disulphide pairings. Two derivatives have been studied, the first being a three-disulphide protein which, as in the case of CM$^{6,127}$-lysozyme, lacks the disulphide linking the N- and C- terminal regions. The second protein studied is a two-disulphide species, lacking this and a second disulphide bridge, which is also located in the α-domain. The influence of these disulphide bonds on the structure and stability of the native and molten globule states is described.

Details of novel applications of ESI-MS to the study of protein folding are described in Chapter Six. This section presents data from experiments performed using hydrogen exchange pulse labelling coupled with ESI-MS to study cooperative elements in protein folding, and to investigate small differences in the refolding properties of variant lysozymes. Results of studies are presented to demonstrate that the standard ESI-MS technique may be modified to monitor hydrogen exchange in native, molten globule and denatured states of proteins, and also to investigate the nature of the interaction of a partially structured polypeptide with the chaperone protein GroEL.

Finally in Appendix C preliminary results of experiments to investigate the effect on the kinetic refolding of an alternative method of destabilising a protein are described. The effects of addition of various concentrations of the chemical denaturant guanidinium chloride to lysozyme during refolding on the kinetics and mechanism of folding are presented, and this is related to refolding studies on the three-disulphide protein as described in earlier chapters.

# Chapter Two

# Materials and Methods

## 2.1 *Reagents*

All chemicals were of Analar grade or above, and obtained from Sigma Chemical Co. Ltd. unless otherwise stated.

## 2.2 *Preparation and characterisation of CM$^{6,127}$-lysozyme.*

CM$^{6,127}$-lysozyme was prepared and extensively purified by a modification of the method described by Radford *et al.* (1991). Hen lysozyme (20 mg/ml) was incubated with constant mixing at 15°C with 5 mM dithiothreitol in 100 mM Tris-acetate buffer at pH 7.8. The instant turbidity induced by the formation of insoluble aggregates was monitored by optical density at 600 nm. When this turbidity had increased 40-fold relative to the starting point, the reaction was quenched by the addition of 25 mM iodoacetic acid and incubated for 15 min at pH 8.0. Further reactions, such as carboxymethylation of His15 (Yamada *et al.*, 1984), or oxidations caused by the presence of free iodine (Imoto & Rupley, 1973), were avoided by quenching to pH 3 with HCl, followed by extensive dialysis of the protein at 4°C against 1 mM HCl. The protein was then lyophilised. The lyophilised protein was suspended in a minimum volume of $H_2O$ and the insoluble fraction was removed by centrifugation (3000×*g* for 15 min at 4°C). CM$^{6,127}$-lysozyme was then purified by ion-exchange chromatography on a

BioRex-70 column (BioRad Laboratories) column dimensions 62 x 5 cm, and eluted with 150 mM sodium phosphate buffer, pH 7.0. Unmodified lysozyme which had bound to the column was then eluted using the same buffer containing 1M NaCl. Purified $CM^{6,127}$-lysozyme was dialysed extensively at 4°C against 0.16 mM HCl (pH 3.8) and then lyophilised. Extreme care was taken to ensure the highest purity of the $CM^{6,127}$-lysozyme used in refolding studies, since any contaminant of unmodified lysozyme could mask refolding of the derivative in the optical measurements. Purity was confirmed by acid urea polyacrylamide gel electrophoresis, amino-acid analysis (Moore *et al.*, 1968; Heinrikson & Meredith, 1984) and ESI-MS, and was found to be >98% pure. The total average yield of pure $CM^{6,127}$-lysozyme was ≤10% of the starting material.

## 2.3 *Differential scanning calorimetry.*

Hen lysozyme and $CM^{6,127}$-lysozyme were prepared and/or purified as described above. Lyophilised protein samples for DSC were taken up in buffer and degassed briefly at room temperature prior to loading. Buffers used were 50 mM glycine/HCl (pH 1.5–4), 50 mM sodium acetate (pH 4.75), or 0.1 M HCl as appropriate. Neither dialysis against large volumes of buffer nor centrifugation of samples prior to calorimetry had any effect on the observations. Sample pH was confirmed at room temperature to be identical both before and after DSC, and protein concentrations (0.7–4 mg/ml) were determined by UV spectrophotometry assuming $\varepsilon_{280}{}^{1\ \mathrm{mg/ml}} = 2.65$ and 14.3 kDa molecular weight for both proteins. Samples were scanned at 60°C $hr^{-1}$ using a Microcal MC-2D instrument and data were analysed, after buffer baseline subtraction, in terms of single cooperative transitions using standard procedures. Re-scans of both modified and unmodified samples after thermal unfolding and re-cooling in the calorimeter showed that the transitions were reversible (>90% reversibility in most cases), and none of the samples showed any turbidity after the experiment indicating in accord with Sophianopoulos & van Holde (1964) that aggregation was insignificant under the conditions used.

## 2.4 *Pulsed amide hydrogen exchange labelling*

Quenched flow experiments were carried out using a Bio-Logic QFM-5 module, using a modification of the protocol described previously (Radford *et al.*, 1992b). A schematic view of the QFM-5 is shown in Figure 2.1. Denatured protein (S1) was mixed rapidly with sodium acetate buffer, pH 4.5 (S2), and allowed to refold for a time controlled by the length of delay loop L1. After this time, the protein was hydrogen exchange labelled by mixing with sodium borate buffered to pH 9.5 for 8.3 ms (controlled by L2). Finally, the refolding protein was quenched to pH 3.8 and the sample collected, after discarding waste from the lines (S5).

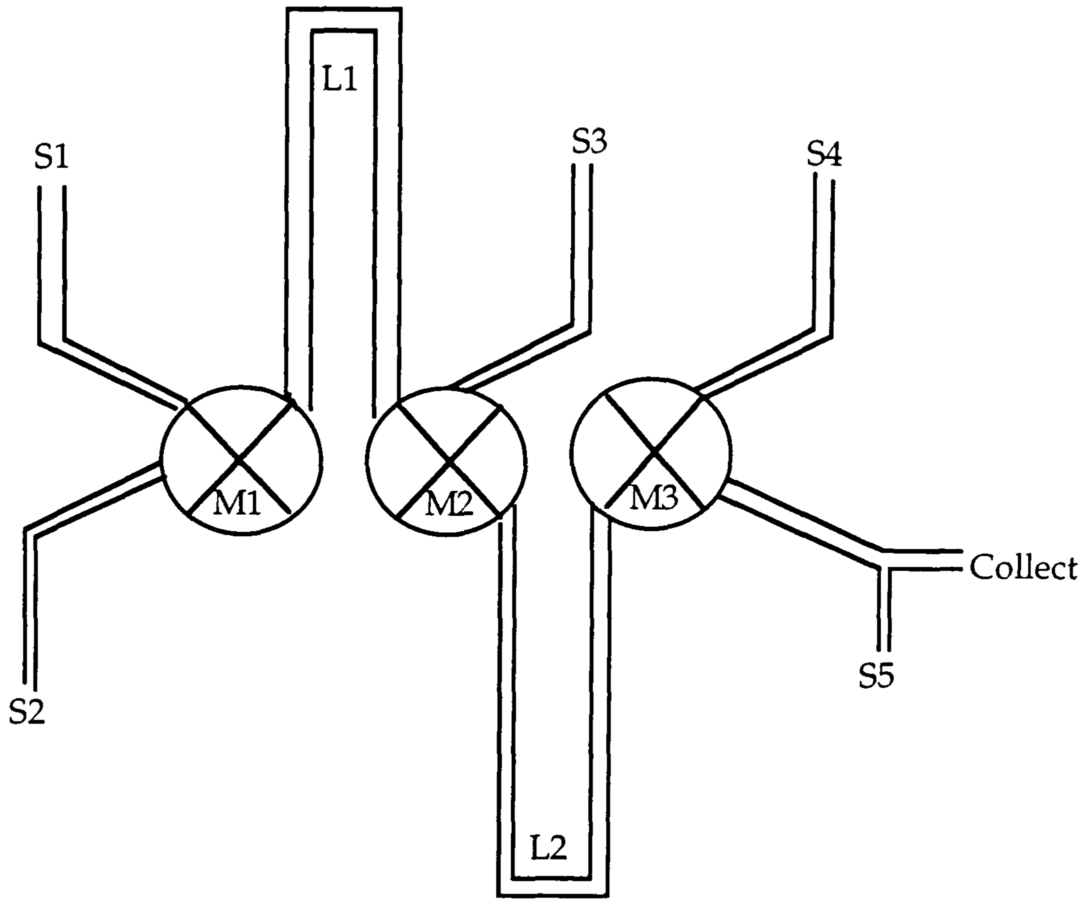


**Figure 2.1**. Schematic representation of the Biologic QFM-5 quenched flow apparatus. S1-5=syringes, M1-3=mixers, L1-2=delay loops.

Deuterated $CM^{6,127}$-lysozyme, in which all of the labile NH hydrogens were exchanged for deuterium, was prepared by incubation in $D_2O$ at pH 3.8, 60°C for 10 min, followed by lyophilisation. This process was repeated three times to ensure complete deuteration of all exchangeable amides. $CM^{6,127}$-lysozyme was denatured at 20 mg/ml in 3M GuDCl. Refolding was initiated by 11-fold dilution into 20 mM sodium acetate buffered to pH 4.5. After various refolding times between 3.5 ms and 5 s, all unprotected amide deuterons were exchanged for protons by the addition of a volume of 20 mM borate buffer (pH 9.5) 5 times that of the starting volume. After a labelling time of 8.4 ms, exchange was quenched by reducing the pH to 4.0 by a further dilution with 0.5 M acetic acid. The protein was then allowed to refold completely under these conditions where further hydrogen exchange was negligible. Samples were concentrated using Centriprep-10 and Centricon-10 micro-concentrators (Amicon, Inc.). For NMR experiments, samples were buffer-exchanged into 40 mM deuterated sodium acetate at pH 3.8 to minimise the need for solvent suppression. The final concentration of samples for NMR experiments was approximately 2 mM.

As a control, a sample of protein was heated to 80°C for 15 min in the same final buffer and isotope concentrations as in the labelling pulse to ensure complete labelling of all amides with correct isotopic distribution. The sample was buffer exchanged in the same way as previously described. This sample was assumed to have 100% proton occupancy and was used to normalise the peak intensities for each amide. A further control was carried out by using an initial protein concentration of 2 mg/ml to ensure no dependence of refolding on the protein concentration. Several time points were measured at each concentration and were confirmed to be identical, within error, at the two concentrations examined. For pulse intensity experiments, the pH of the borate buffer was adjusted to give the desired labelling pH between pH 9.0 and 10.5, and the concentration of acetic acid was varied accordingly to ensure efficient quenching to pH 4.

## 2.5 *NMR spectroscopy*

Phase-sensitive J-correlated spectroscopy (Aue *et al.*, 1976; Bax & Freeman, 1981) was performed at 500 MHz on a GE/Nicolet spectrometer of the Oxford Centre for Molecular Sciences, at a temperature of 20°C. This temperature was chosen to minimise amide hydrogen exchange during acquisition while maximising resolution of cross peaks. Data sets consisting of 256 complex $t_1$ increments of 32 transients were collected, using a sweep width of 7042 Hz and a relaxation delay of 700 ms. Data were processed and peak-picked using the FTNMR and FELIX programs (Hare Research Inc.) on a SUN workstation. Resolution enhancement was by trapezoidal multiplication and double-exponential multiplication in $t_2$, and in $t_1$ by trapezoidal multiplication alone. After zero-filling, the final digital resolution was 1.7 Hz/point in both dimensions.

$\alpha$CH–NH cross peak intensities were measured as the sum of absolute peak heights of the four phase sensitive components of each peak, and intensities in each spectrum were normalised to two non-exchangeable aromatic cross peaks (Tyr23 and Tyr53 $C_\delta H$–$C_\varepsilon H$). Individual cross peak intensities were then normalised to the values obtained from the zero time point sample. Protection profiles were fitted to the sum of two exponential functions by non-linear regressional analysis, using the program Kaleidagraph (Abelbeck Software, Inc.).

## 2.6 *Electrospray ionisation mass spectrometry*

For mass spectrometry experiments, samples from pulse labelling experiments were exhaustively desalted by washing on Centricon-10 membranes with Milli-Q water adjusted to pH 3.8 with HPLC grade formic acid. Experiments were performed using a modification of the method of Miranker *et al.* (1993), using a lower temperature and a less stringent solvent system to minimise unfolding and hydrogen exchange of this less stable protein. All mass spectra were acquired on a platform single analyser mass spectrometer equipped with an electrospray interface (VG Biotech, Fisons Instruments). The sample interface was equilibrated overnight with water/formic acid at 4°C, pH 3.8. The solvent delivery system (Michrom BioResources, Inc.) and the nebuliser gas were immersed in an ice/salt bath. Protein samples at 20 pmoles/µl in water/formic acid, pH 3.8, were introduced into the ion source at 10 µl/min using the same buffer as the mobile phase. The instrument was scanned from 1260-1480 m/z and calibrated against protonated wild-type hen lysozyme. Under these conditions, no hydrogen exchange was detected during the acquisition time of the spectrum (~2 min).

## 2.7 *Circular dichroism*

To monitor equilibrium unfolding of $CM^{6,127}$-lysozyme, samples were dissolved to a protein concentration of 0.2 mg $ml^{-1}$ in unbuffered GuHCl solutions, adjusted to pH 3.8 with small amounts of HCl or NaOH. Ellipticity at 222 nm was measured on a Jasco J-720 spectropolarimeter in 1 mm cells. Stopped flow experiments were carried out using a Jasco J-720 spectropolarimeter equipped with a Bio-Logic SFM-3 stopped-flow module. $CM^{6,127}$-lysozyme at 20 mg/ml for near-UV CD, or 2 mg/ml for far-UV CD was dissolved in 3M GuHCl. Refolding was initiated by an 11-fold dilution into 20 mM sodium acetate buffered to pH 4.5, thus giving refolding conditions identical to the pulse labelling experiments. Refolding was monitored over a period of 3 s at either 289 nm (near-UV) or 225 nm (far-UV), using a 2 ms step resolution, averaging data points over 4 ms.

## 2.8 *Fluorescence*

Experiments were carried out using identical refolding conditions to the near-UV CD measurements. An Applied Photophysics SX17MV stopped-flow instrument, equipped with 2.5 ml and 250 μl syringes, to allow a 10:1 (buffer : protein) mix, was used with an excitation wavelength of 285 nm. Total fluorescence emission above 320 nm was measured using a high pass filter. Refolding was monitored for 2s with a 1 ms step resolution.

Formation of the active site was monitored by binding of the fluorescently labelled inhibitor 4-methyl umbelliferyl N,N'-di-N-acetylglucosamine (MeU-diNAG), which was added to a concentration of 6 μM into the refolding buffer. Total fluorescence above 335 nm was monitored, with excitation at 330 nm.

## 2.9 *NMR studies of α-lactalbumin derivatives*

Bovine α-lactalbumin (Type I) was obtained from the Sigma Chemical Co. Ltd., desalted by extensive dialysis at pH 2.0, lyophilised and stored at -20°C. Guinea pig α-lactalbumin was purified from milk whey obtained from Porcellus Animal Breeding, Sussex, as described by Chyan *et al.* (1993). Samples of a three-disulphide derivative of bovine α-lactalbumin, in which the Cys6–Cys120 disulphide bridge had been selectively reduced and carboxyamidomethylated ($3SS_{cam}$), and a two-disulphide derivative in which this and the Cys28–Cys111 disulphide were similarly treated ($2SS_{cam}$), prepared as described in Ewbank & Creighton (1993b), were a generous gift from J.J. Ewbank (EMBL, Heidelberg). Samples of the *apo* protein were prepared by dissolving lyophilised powder in 0.1 mM HCl and adjusting to pH 2.0 as necessary with small amounts of HCl or NaOH. The high pH forms of the proteins were prepared by dissolving in 0.1 M $NH_3$ solution and lowering the pH with HCl. This avoids any irreversible precipitation on passing through the isoelectric point (~pH 4 ) of the protein. Addition of $CaCl_2$ to a concentration of 1 mM was sufficient to ensure formation of the *holo* protein where required.

Samples for NMR were 2 mM for unmodified α-lactalbumin and $3SS_{cam}$; those of $2SS_{cam}$ were somewhat less concentrated (<1 mM) due to low solubility of the protein. $^1H$ NMR experiments were performed at 600 MHz on a Bruker AM600 spectrometer. COSY, DQF-COSY and NOESY experiments (Aue *et al.*, 1976; Jeener *et al.*, 1979; Bax & Freeman, 1981; Macura *et al.*, 1981; Rance *et al.*, 1983) were acquired using time proportional phase incrementation (Marion & Wüthrich, 1983). 512 $t_1$ increments of 48 transients each were acquired over a sweep width of 8300 Hz in each dimension. Spectra were processed in $t_2$ by trapezoidal multiplication and double-exponential multiplication and in $t_1$ by trapezoidal multiplication only, using the program FELIX 2.3 (Biosym Inc.). After zero filling, final resolution was 4.0 Hz/point in each dimension.

# Chapter Three

# Thermodynamic Consequences of the Removal of a Disulphide Bridge from Hen Lysozyme

## 3.1 *Introduction*

Many proteins rely on disulphide bridges for the integrity of their native three-dimensional structures and removal of one or more of these results in a reduction of the stability of the native relative to the denatured state (Creighton, 1974; White, 1982; Taniyama *et al.*, 1988; Taniyama *et al.*, 1990; Inaka *et al.*, 1991; Betz, 1993; Zhang *et al.*, 1994). Consequently, there has been a great deal of interest in the significance of disulphides for the stability of folded proteins. Protein engineering techniques have been employed recently in attempts to increase the overall stability of proteins by the introduction of non-native disulphide linkages into the structure. Although in some cases this has increased the stability of the protein relative to the wild type form (Creighton, 1988; Matsumura *et al.*, 1989), this has not always been true (Wells & Powers, 1986; Wetzel, 1987; Wetzel *et al.*, 1988). Despite these findings, few detailed thermodynamic analyses of the effect of disulphide bonds on protein stability have been carried out. Most studies have suggested that disulphide linkages confer stability primarily by limiting the conformational entropy of the unfolded state and hence destabilising this relative to the native state (Anfinsen &

Scheraga, 1975; Creighton, 1988). By contrast, however, a recent report (Doig & Williams, 1991) suggests that disulphides actually destabilise the native state entropically by reducing exposure of hydrophobic residues to solvent in the denatured state, and that their role in stabilising proteins is predominantly enthalpic as a consequence of less favourable hydrogen bonding networks in the cross-linked denatured state relative to that in the absence of disulphides.

In order to test directly the role of disulphides in protein stability we have examined the thermodynamics of unfolding of a three-disulphide derivative of hen lysozyme (CM$^{6,127}$-lysozyme) in which the disulphide bridge linking residues Cys-6 and Cys-127 has been specifically reduced and carboxymethylated. NMR studies of this protein have shown that it retains the same secondary and tertiary structure as the original four-disulphide enzyme, but that its thermal stability is 25°C lower at pH 3.8 than that of the wild type protein (Radford *et al.*, 1991). In this chapter, results are described of detailed DSC experiments, performed under conditions of pH where the $T_m$ of CM$^{6,127}$-lysozyme varies between 26 and 53°C, compared with data for the wild type protein obtained under similar conditions.

An alternative method of measuring protein stability is to consider the protection of labile amide hydrogens against exchange in the native state. Exchange at amide sites which are exposed to solvent can occur trivially, but hydrogen exchange which occurs at sites buried within the protein structure can give useful information as to local stability of regions of protein structure, as described later in this chapter (§3.3). This technique is complementary to the calorimetric approach, which can accurately measure global thermal stability of proteins.

## 3.2 *DSC studies on CM$^{6,127}$-lysozyme.*

Representative heat capacity data for thermal unfolding of unmodified and CM$^{6,127}$-lysozymes are given in Figure 3.1. Over the pH range studied (pH 1-5) both proteins exhibit DSC thermograms typical of cooperative unfolding of small globular proteins, with a significant increase in excess heat capacity ($\Delta C_p$) for the unfolded polypeptide with respect to the folded protein. Thermal stability is quite sensitive to pH in this region for both proteins (Figure 3.2), with CM$^{6,127}$-lysozyme being 25-30°C less stable, depending on pH, than its unmodified counterpart under the same conditions, in agreement with $^1$H NMR data previously reported (Radford *et al.*, 1991). Despite this large difference in stability there is no indication of any major changes in $\Delta C_p$ between the two lysozyme species. Heat capacity data, in both cases, consistently fit to a single transition model with a $\Delta C_p$ of about 6300 J K$^{-1}$ mol$^{-1}$, in agreement with previous observations for unmodified lysozyme (Privalov & Khechinashvili, 1974; Privalov & Gill, 1988). Determinations of $\Delta C_p$ from such traces are, however, subject to potential error due to uncertainties in baseline extrapolations, and also rely on the assumption that $\Delta C_p$ is itself temperature independent. Detailed studies (A. Cooper, unpublished results; see also Privalov & Gill (1988)) indicate that this is not necessarily correct, but that a more reliable comparison is given by plotting calorimetric enthalpies of unfolding *versus* $T_m$ (Figure 3.3).

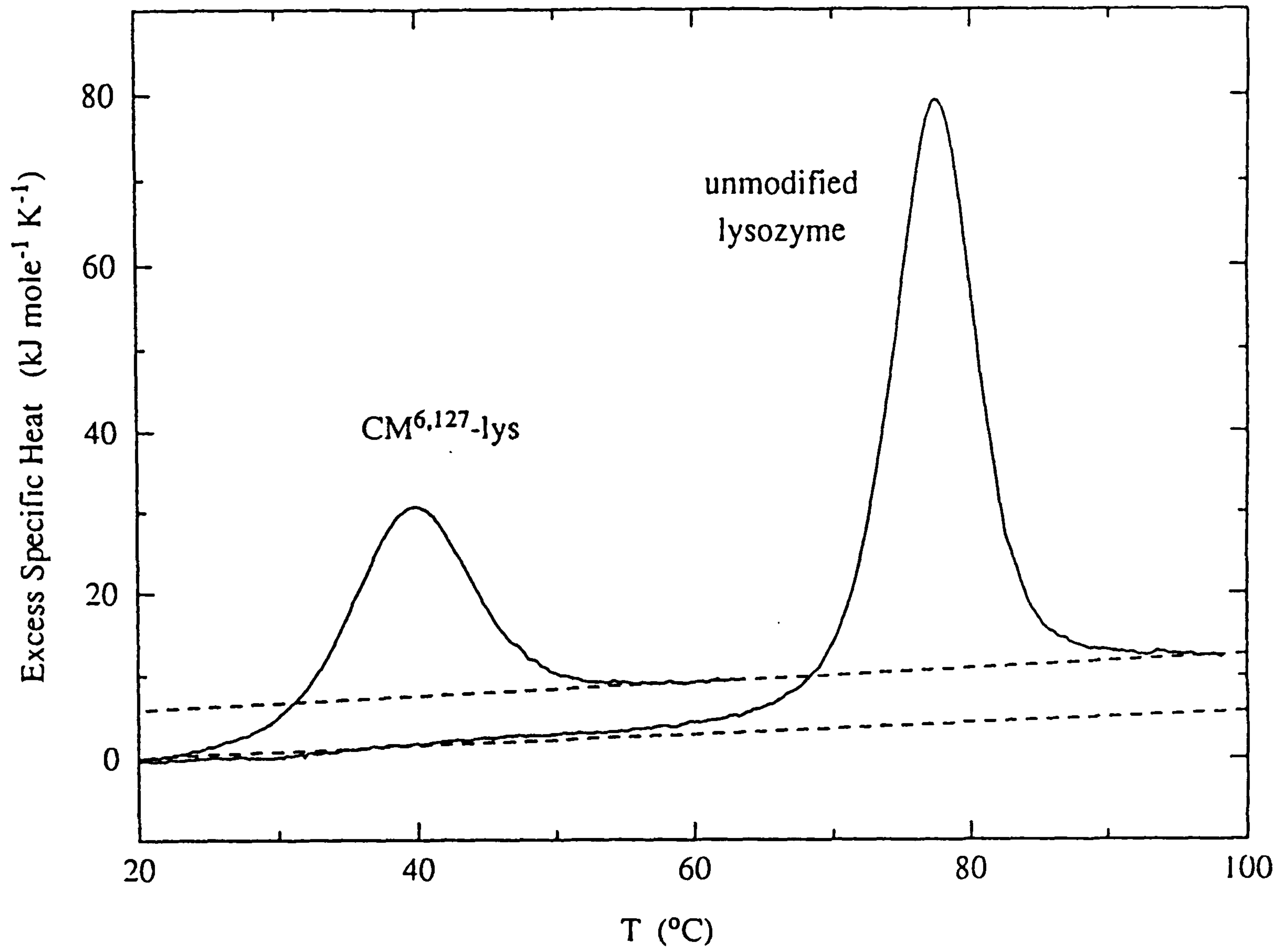


**Figure 3.1.** Normalised excess heat capacity data for DSC transitions of CM$^{6,127}$-lysozyme (pH 2.96, 0.201 mM) and unmodified hen egg-white lysozyme (pH 3.8, 0.215 mM). The dashed lines indicate pre- and post-transition baselines with $\Delta C_p$ = 6300 J K$^{-1}$ mol$^{-1}$.

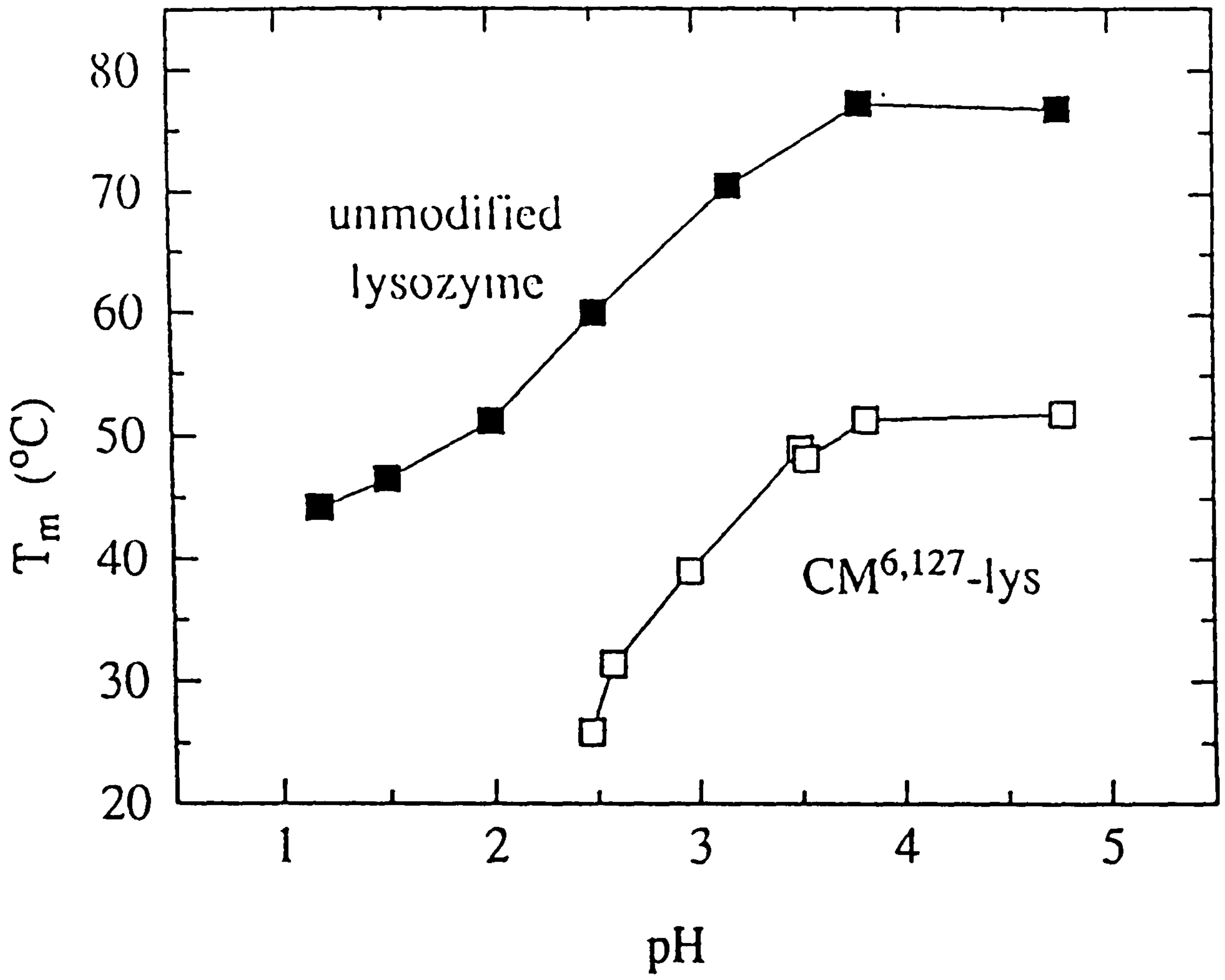


**Figure 3.2.** pH dependence of the mid-point temperatures of thermal unfolding transitions ($T_m$) of unmodified (■ ) and CM$^{6,127}$- (□ ) lysozymes.

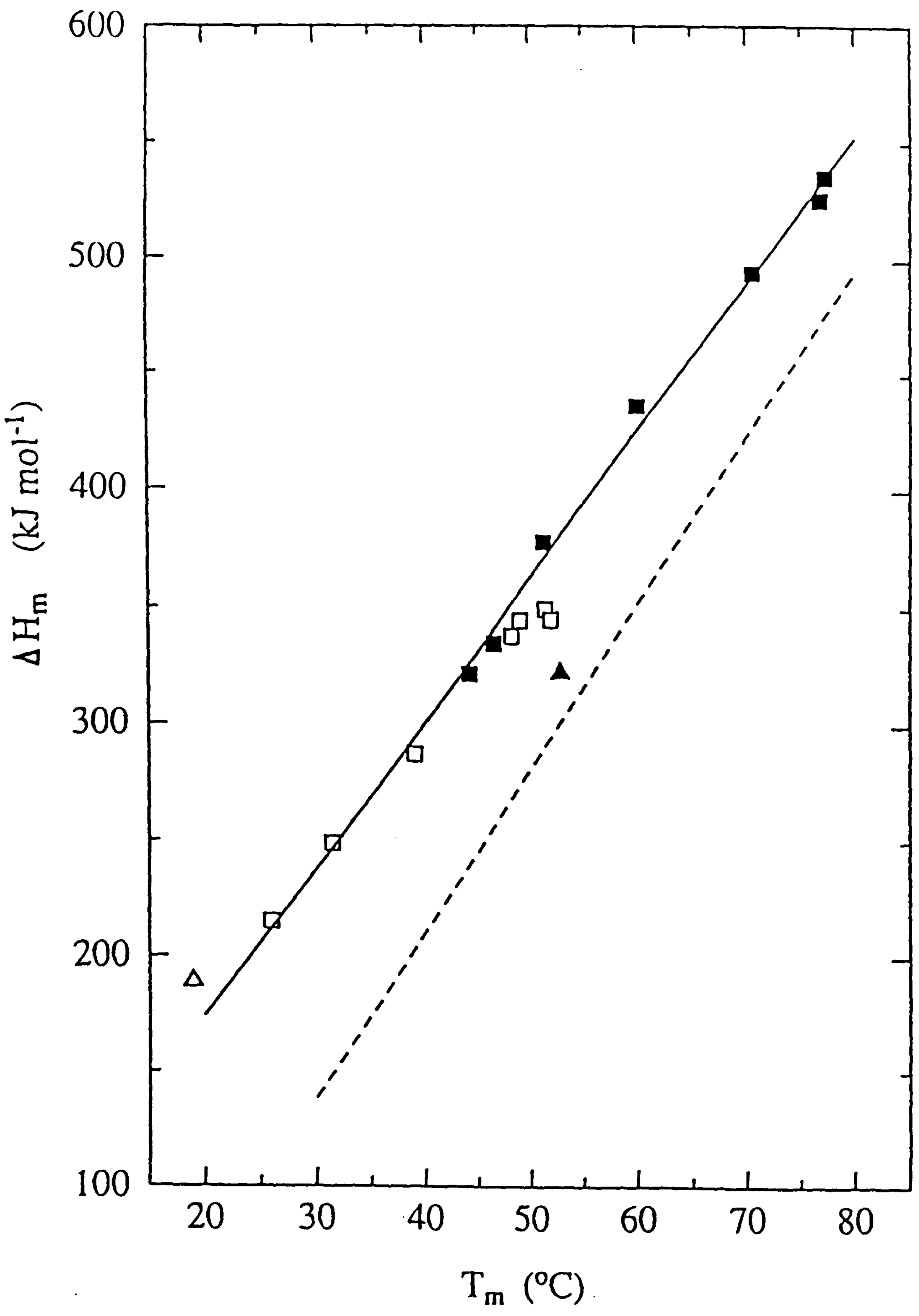


**Figure 3.3.** Calorimetric enthalpies of unfolding at the mid point of transition ($\Delta H_m$) *versus* $T_m$ for unmodified (■ ) and CM$^{6,127}$- (□ ) lysozymes. The solid line is a linear least squares fit to the data: $\Delta H_m = 48.6 + 6.30\ T_m$ / kJ mol$^{-1}$. The dashed line indicates the predicted temperature dependence of ΔH, using the equations suggested by (Doig & Williams, 1991). For comparison, the values obtained for $\Delta H(T_m)$ of unmodified (▲) and CM$^{6,127}$-lysozyme(△) spectrophotometrically by (Denton & Scheraga, 1991) are marked.

It is clear that the data for $CM^{6,127}$-lysozyme are contiguous with the unmodified protein and are consistent with a common $\Delta C_p$, from linear regression, of 6300 (±210) J $K^{-1}$ $mol^{-1}$. (Least squares analysis of the $CM^{6,127}$-lysozyme data alone indicates that the $\Delta C_p$ for the modified protein may be slightly smaller at 5200 (±210) J $K^{-1}$ $mol^{-1}$, though this may be modelled satisfactorily by a slight curvature of the $\Delta H_m$ *versus* $T_m$ plot.)

Using $\Delta C_p$ values we can transform the calorimetric data to give normalised thermodynamic parameters ($\Delta H_{298}$, $\Delta S_{298}$ and $\Delta G_{298}$) for unfolding of each protein at a common reference temperature (25°C). For unmodified lysozyme the mean $\Delta H_{298}$ is 210 (±8) kJ $mol^{-1}$, compared with a value of 209 (±6) kJ $mol^{-1}$ for $CM^{6,127}$-lysozyme. Neither value shows any significant variation with pH over the observed range. These data agree well with the values reported previously (Privalov & Khechinashvili, 1974; Privalov & Gill, 1988) for unmodified lysozyme, and by Denton & Scheraga (1991) for $CM^{6,127}$-lysozyme (determined by non-calorimetric methods at pH 2, 19°C). The $\Delta H_{298}$ values are also similar to those obtained by $^1H$ NMR for the wild type and three-disulphide proteins (Dobson & Evans, 1984; Radford *et al.*, 1991). More significantly, the close similarity of the normalised enthalpies of unfolding of the two proteins suggests that the reduced stability of three-disulphide lysozyme arises mainly from entropic effects. This is confirmed by considering the pH dependence of the normalised entropies and free energies of unfolding (Figure 3.4). Although $\Delta S_{298}$ varies significantly with pH (due to the effect of pH on the stabilities of both proteins), the entropy of unfolding of $CM^{6,127}$-lysozyme at 25°C is higher than that of the wild type protein by some

90 J $K^{-1}$ $mol^{-1}$. This is consistent with the anticipated increase in conformational entropy in the unfolded polypeptide resulting from removal of a single disulphide crosslink (Schellman, 1955; Flory, 1956; Poland & Scheraga, 1965; Creighton, 1988; Pace *et al.*, 1988). Various theoretical methods for estimating the magnitude of this effect, based on the size of the polypeptide loop, but ignoring the effect of other overlapping disulphides, have been proposed. For the 122 residue loop involved here, $\Delta\Delta S$ estimates range from about 70 to 95 J $K^{-1}$ $mol^{-1}$ (Pace *et al.*, 1988). The observed $\Delta\Delta S_{298}$ (Figure 3.4) is consistent with these estimates.

Removal of the disulphide bond in $CM^{6,127}$-lysozyme is accompanied by the introduction of two carboxymethyl groups into the protein. The possibility that the thermodynamic parameters reported here are significantly affected by electrostatic effects associated with these groups can be ruled out by examining the pH dependence of the free energies of unfolding ($\Delta\Delta G_{298}$). Linked-function analysis (Wyman, 1964; Pfeil & Privalov, 1976) shows that the variation in free energy of any process due to pH change is associated with the number of $H^+$ ions ($\Delta v$) taken up in the process; in the present case $d(\Delta G_{298})/d(pH)=2.303RT\Delta v$. Plots of the unfolding free energies (25°C) *versus* pH (Figure 3.4) are quite parallel for the two proteins, with a maximal slope equivalent to $\Delta v = 3$ in both cases, in agreement with previous studies of unmodified lysozyme (Pfeil & Privalov, 1976).

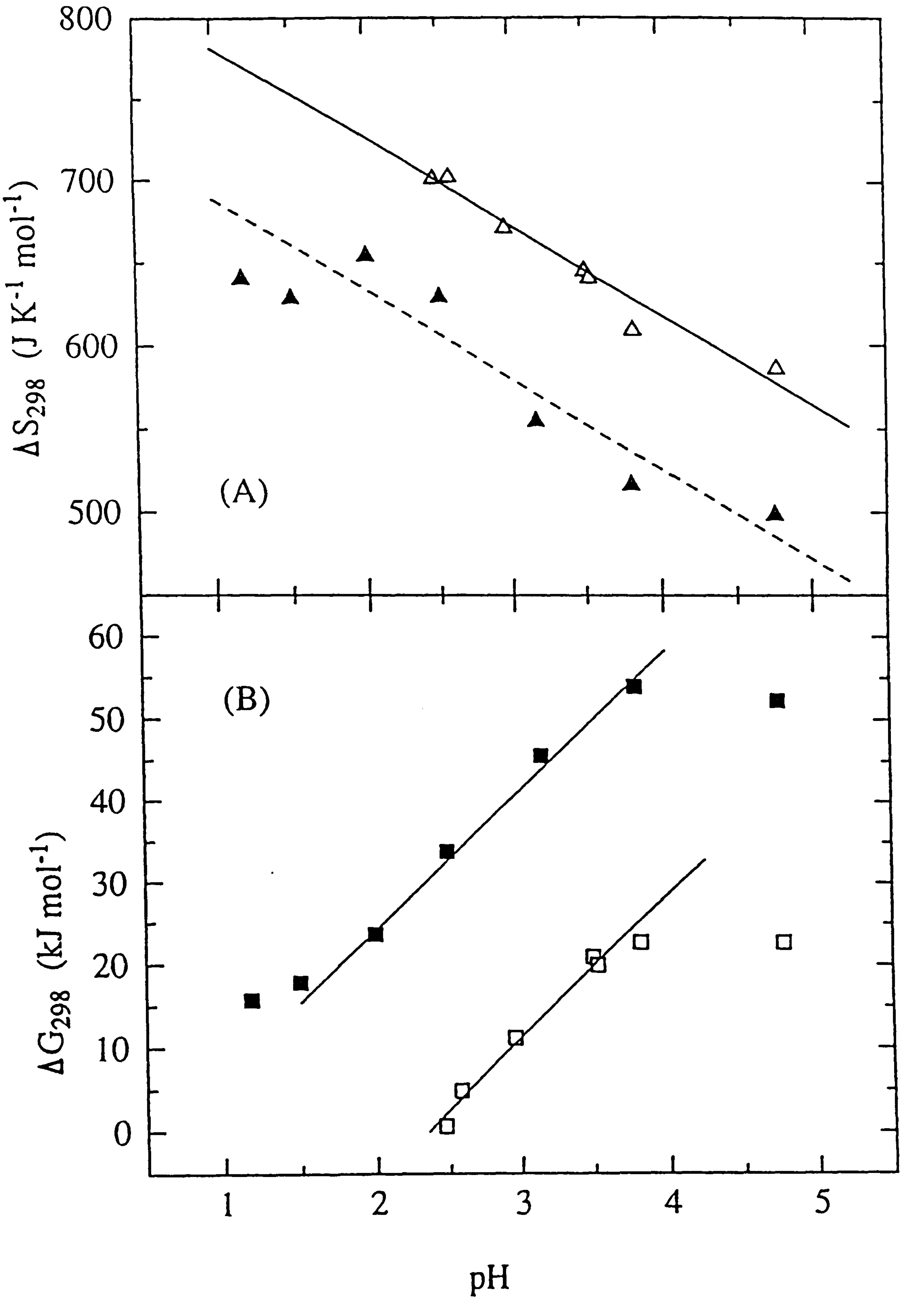


**Figure 3.4.** (A): Entropies of unfolding ($\Delta S_{298}$) *versus* pH for unmodified (▲) and CM$^{6,127}$-(△) lysozyme. (B): Free energies of unfolding ($\Delta G_{298}$) at 25°C *versus* pH for unmodified (■) and CM$^{6,127}$- (□) lysozyme. The straight lines are drawn with a slope equivalent to the uptake ($\Delta v$) of 3.0 $H^+$ ions per molecule during unfolding

Thus, although the two carboxymethyl groups (estimated $pK_a$ ca. 3.7 (Jencks & Regenstein, 1975)) are expected to titrate over this pH range, this result indicates that their $pK_a$ values are not significantly different between the folded and unfolded state (presumably because of their exposed surface location) and show no thermodynamic linkage to the unfolding process. The effect of pH on the thermal stability of the proteins must be due to other ionisable groups which are common to both modified and unmodified forms. The low heat of ionisation of carboxylic acid groups (Christensen *et al.*, 1976) indicates that protonation of aspartate or glutamate side chains is predominantly an entropic process so that, in agreement with our observations, we should anticipate very little variation in $\Delta H_{298}$ with pH in the acidic range.

It has been suggested (Denton & Scheraga, 1991) that there may be some small enthalpic contribution to the destabilisation of the native state of $CM^{6,127}$-lysozyme. NMR data (Radford *et al.*, 1991) show that the loss of this disulphide linkage may alter the relative orientation of the N- and C- termini to some, presumably small, extent, as indicated by the loss of a number of nOe's between the side chain of Leu-129 and the N-terminal helix. That the difference in structure around this region is small is confirmed by the X-ray crystal structure (Hill *et al.*, 1993). The present data show that any energy change associated with this effect is not large. The results of this study are, therefore, completely consistent with a reduction in the stability of $CM^{6,127}$-lysozyme being attributable simply to an increase in the conformational entropy of the unfolded state. In accord with this, stabilisation of hen lysozyme by introduction of an ester linkage between residues Glu-35 and Trp-108 (Imoto & Rupley, 1973) gave no difference in $\Delta C_p$ of unfolding from the unmodified protein (Johnson *et al.*, 1978), indicating a wholly entropic contribution to stability.

It is of interest to compare these findings with the conclusions from the recent correlation reported for a series of six proteins between $\Delta C_p$ and the number of disulphide bridges in the protein (Doig & Williams, 1991). Based on this correlation, a $\Delta C_p$ value of some 7100 J $K^{-1}$ $mol^{-1}$ would be predicted for $CM^{6,127}$-lysozyme, giving a variation of $\Delta H$ with temperature as indicated in Figure 3.3, together with significant reductions in $\Delta H_{298}$ and $\Delta S_{298}$. Our data do not support this conclusion, but suggest that some property of the proteins, other than the number of disulphide bonds, contributes to the correlation observed by Doig & Williams (1991). Large heat capacity changes have customarily been related to contributions to the free energy of folding from the hydrophobic effect, associated with the difference in exposure of hydrophobic groups in the native and denatured states of proteins. The correlation observed by Doig & Williams could therefore indicate that protein structures may utilise disulphide bridges for stability if $\Delta C_p$ per residue is small, for example as a consequence of a small hydrophobic core (P.A. Evans, personal communication). In a more general sense, incorporation of disulphide bonds may, through their effects on the entropy of the denatured state, promote folding of a protein to a state that is functionally favourable, but does not otherwise have sufficient intrinsic stability. Further studies of the type described here, and the use of mutants, for a wider range of proteins are necessary to establish general conclusions of this kind.

## 3.3 *Stability of proteins monitored by hydrogen exchange.*

Whilst calorimetry is an extremely accurate method of measuring global stabilities of proteins during thermal unfolding, much information can be gained about the stability and folding of proteins from their dynamic properties in solution (McCammon & Karplus, 1983; Williams, 1989). Hydrogen exchange at labile sites in a protein structure has been identified as one such method of measuring these dynamic properties (Linderstrøm-Lang, 1955). In order for exchange to occur of labile hydrogens which are completely or partially buried within a protein structure, some fluctuation in the structure must occur. This has led to the development of a number of models for the exchange process, from local unfolding (Englander *et al.*, 1980) to solvent penetration into mobile defects (Lumry & Rosenberg, 1975), which have been reviewed in great depth (Woodward *et al.*, 1982; Englander & Kallenbach, 1984). If the NH–αCH resonances in the $^1$H NMR spectrum of a protein are fully assigned then detailed information may be obtained on a residue specific basis as to the hydrogen exchange events occurring, and to the regions of protein structure which are protected against exchange. A number of small proteins have been studied using these methods, over a range of pH and temperature, including BPTI (Wagner & Wüthrich, 1982), cytochrome *c* (Wand *et al.*, 1986) and hen lysozyme (Wedin *et al.*, 1982; Delepierre *et al.*, 1987; Pedersen *et al.*, 1991), and have shown that hydrogen exchange in the native state of proteins may be influenced by a large number of factors involving side chain motions, local and global unfolding events. This approach may be particularly usefully employed for direct comparison of similar proteins, since it can pinpoint small differences in the behaviour of each protein. The effect of introducing disulphide bridges into barnase on the dynamics and stability of the protein has been investigated (Clarke & Fersht, 1993; Clarke *et al.*, 1994). In the following section, hydrogen exchange experiments are described which compare the native state of hen lysozyme with that of the three-disulphide derivative CM$^{6,127}$-lysozyme.

## *3.4 Hydrogen exchange in native CM$^{6,127}$-lysozyme.*

The $^1$H NMR spectra of hen lysozyme and CM$^{6,127}$-lysozyme have been completely assigned at 35°C, pH 3.8 (Redfield & Dobson, 1988; Radford *et al.*, 1991). Both proteins are native under these conditions, with virtually identical structures, as judged by NMR and X-ray crystallography (Radford *et al.*, 1991; Hill *et al.*, 1993), and by optical methods (Denton & Scheraga, 1991); indeed the three-disulphide derivative largely retains its enzymatic activity (Radford *et al.*, 1991). At this temperature, however, although CM$^{6,127}$-lysozyme is still folded, the majority of backbone amides exchange, some to a substantial degree, with bulk solvent during the 4 hr acquisition time of a COSY experiment, whereas amides in the unmodified protein are stable against exchange for weeks under these conditions (Pedersen *et al.*, 1991). One possible explanation for this is that CM$^{6,127}$-lysozyme has a lower $T_m$ (53°C at pH 3.8) than that of the unmodified protein (77°C, pH 3.8 (Wedin *et al.*, 1982)), thus allowing increased exchange *via* an unfolding mechanism (Kim & Woodward, 1993). In accord with this, at a lower temperature of 20°C, hydrogen exchange is significantly retarded in native CM$^{6,127}$-lysozyme, allowing exchange to be monitored readily by 2D NMR. Under these conditions, exchange in the native protein is insignificant during spectral acquisition whilst the increased line widths and cross peak overlap at the lower temperature do not reduce substantially the number of amides which could be followed. Spectral assignment at 20°C was readily possible by direct comparison with spectra at 35°C since all changes in αCH chemical shifts were relatively small (≤ 0.1 ppm). Assignments were confirmed by means of NOESY experiments.

The rates of exchange in the native state of $CM^{6,127}$-lysozyme at 20°C were monitored for 65 amides. Protection factors, defined as the ratio of the intrinsic rate of hydrogen exchange in an unstructured polypeptide to the measured hydrogen exchange rate in the protein under identical conditions, were calculated using the latest correction factors for neighbouring side chain effects (Bai *et al.*, 1993). The exchange rate constants and calculated protection factors are presented in Table 3.1. Figure 3.5 shows the pattern of protection for $CM^{6,127}$-lysozyme (pH 3.8, 20°C), and alongside it the corresponding data obtained previously for the unmodified protein (pH 7.5, 30°C) (Radford *et al.*, 1992a) recalculated using the new correction factors of Bai *et al.*, (1993). The overall pattern of protection in $CM^{6,127}$-lysozyme is similar to that of the intact protein in that those amides which are most highly protected from exchange lie in regions of well-defined secondary structure, confirming the structural integrity of $CM^{6,127}$-lysozyme. The calculated protection factors are, however, generally lower in the case of the three-disulphide derivative. The most noticeable difference is that protection from exchange is, on average, approximately 100-fold lower in the N- and C- terminal regions of $CM^{6,127}$-lysozyme, near to the site of modification. The central strand of triple-stranded β-sheet is also somewhat less protected in $CM^{6,127}$-lysozyme, whilst in other regions the differences are relatively small.

| Residue | Measured rate ($s^{-1}$) | $\log_{10}$ (protection factor) |
|---|---|---|
| Phe 3 | 1.28e-02 | 2.57 |
| Leu 8 | 2.19e-03 | 3.52 |
| Ala 9 | 5.36e-04 | 4.12 |
| Ala 10 | 9.61e-04 | 4.08 |
| Ala 11 | 1.99e-04 | 4.76 |
| Met 12 | 7.99e-05 | 5.14 |
| Lys 13 | 1.54e-03 | 3.93 |
| Arg 14 | 1.24e-01 | 2.16 |
| His 15 | 6.88e-02 | 3.23 |
| Leu 17 | 6.84e-03 | 2.81 |
| Tyr 23 | 5.58e-03 | 3.21 |
| Asn 27 | 5.00e-03 | 4.01 |
| Trp 28 | 5.98e-07 | 7.19 |
| Val 29 | 3.21e-04 | 3.74 |
| Cys 30 | 1.27e-03 | 4.36 |
| Ala 31 | 1.44e-06 | 7.35 |
| Ala 32 | 5.98e-07 | 7.28 |
| Lys 33 | 4.03e-06 | 6.41 |
| Phe 34 | 3.43e-07 | 7.40 |
| Glu 35 | 1.10e-02 | 3.31 |
| Ser 36 | 4.00e-03 | 4.21 |
| Asn 37 | 5.76e-03 | 4.08 |
| Phe 38 | 9.47e-04 | 4.16 |
| Asn 39 | 5.11e-04 | 4.89 |
| Thr 40 | 3.60e-03 | 3.75 |
| Gln 41 | 1.38e-02 | 3.17 |
| Ala 42 | 4.97e-04 | 4.56 |
| Asn 44 | 2.28e-03 | 4.38 |
| Asn 46 | 2.52e-02 | 3.36 |
| Ser 50 | 8.78e-03 | 3.65 |
| Thr 51 | 6.12e-03 | 3.49 |
| Asp 52 | 2.98e-03 | 4.47 |
| Tyr 53 | 2.30e-03 | 4.02 |
| Gly 54 | 7.09e-03 | 3.52 |

| Residue | Measured rate ($s^{-1}$) | $\log_{10}$ (protection factor) |
|---|---|---|
| Ile 55 | 1.30e-03 | 3.38 |
| Leu 56 | 3.34e-03 | 2.72 |
| Gln 57 | 4.97e-07 | 7.21 |
| Ile 58 | 6.98e-07 | 6.68 |
| Asn 59 | 1.53e-02 | 3.13 |
| Ser 60 | 2.49e-03 | 4.35 |
| Arg 61 | 1.43e-03 | 4.28 |
| Trp 63 | 3.96e-07 | 8.00 |
| Cys 64 | 9.47e-07 | 8.00 |
| Asn 65 | 1.60e-04 | 5.80 |
| Leu 75 | 7.20e-04 | 3.94 |
| Cys 76 | 9.47e-05 | 5.42 |
| Ile 78 | 2.00e-06 | 6.34 |
| Cys 80 | 1.70e-06 | 7.13 |
| Ala 82 | 9.94e-03 | 3.36 |
| Leu 84 | 4.61e-03 | 2.61 |
| Ser 85 | 6.41e-02 | 2.40 |
| Asn 93 | 5.58e-03 | 3.66 |
| Cys 94 | 3.48e-03 | 4.38 |
| Lys 96 | 5.98e-07 | 7.24 |
| Lys 97 | 2.00e-07 | 7.83 |
| Ile 98 | 2.00e-06 | 6.14 |
| Val 99 | 4.97e-07 | 6.43 |
| Ser 100 | 6.98e-02 | 2.44 |
| Trp 108 | 7.02e-03 | 2.80 |
| Trp 111 | 3.16e-03 | 3.15 |
| Arg 112 | 5.44e-03 | 3.29 |
| Arg 114 | 7.31e-02 | 2.59 |
| Cys 115 | 8.64e-03 | 3.89 |
| Ile 124 | 8.39e-03 | 2.29 |
| Arg 125 | 6.41e-01 | 1.10 |

**Table 3.1.** Hydrogen exchange rates and corresponding protection factors measured for exchange of CM$^{6,127}$-lysozyme, pH 3.8, 20°C.

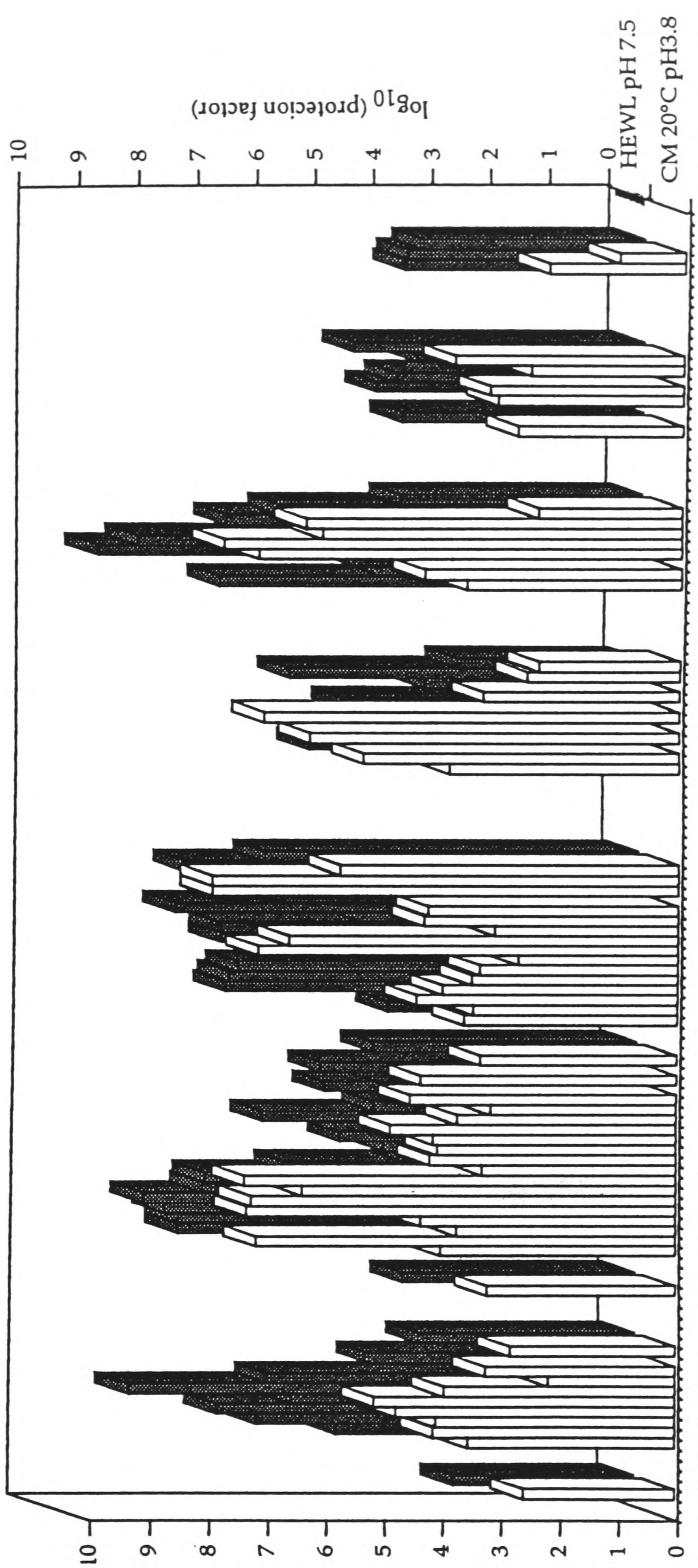


**Figure 3.5.** Amide hydrogen exchange protection in the native state of CM$^{6,127}$-lysozyme, pH 3.8, 20°C (shown in white), and unmodified hen lysozyme, pH 7.5, 30°C (shown in black). $\log_{10}$(protection factor) is plotted *versus* the sequence for the two proteins.

The amides monitored in CM$^{6,127}$-lysozyme can be divided into two distinct classes, namely those with protection factors $\leq 10^5$, and those with protection factors exceeding $10^5$. The more highly protected class consists of 17 residues, the majority of which are clustered in a region far from the site of modification: 9 residues in the B- and C- helices of the protein, 2 amides in the third strand of the triple-stranded β-sheet and 6 in the long loop region as indicated in Figure 3.6. These residues are quite close in the tertiary structure in a region containing two disulphide bridge crosslinks (64–80 and 76–94), and there is good correspondence with the most highly protected amides in wild-type lysozyme, which have protection factors $\sim 10^8$ (Pedersen *et al.*, 1991). By contrast, in the unmodified protein a large number of other amides, found throughout the entire protein sequence have protection factors exceeding $10^5$. A calculation using the methods of Lee & Richards (1971) indicates very little difference in solvent accessibility between the native states of the unmodified protein and its three-disulphide derivative. This suggests that a stable hydrogen exchange core persists within the structure of CM$^{6,127}$-lysozyme, of slightly lower stability than the unmodified protein, whereas the remainder of the molecule is less stable against hydrogen exchange, presumably due to increased frequency of local unfolding events following removal of the Cys6–Cys127 disulphide bridge. Nevertheless, protection factors in the three-disulphide protein are sufficiently high to enable pulsed hydrogen exchange labelling experiments measured by NMR to be performed at 20°C (see Chapter Four).

**Figure 3.6**. Schematic diagram of the native structure of hen lysozyme. The Cys6–Cys127 disulphide bridge is shown in white. The amino acid residues whose backbone amides are most protected against hydrogen exchange in the native state of CM$^{6,127}$-lysozyme at pH 3.8, 20°C, are indicated in grey, with the corresponding amide nitrogen represented in black. This figure was drawn using the program MolScript (Kraulis, 1991).

The mechanism by which amide hydrogens exchange with bulk solvent is complex (Englander & Kallenbach, 1984), but in the case of the native state of CM$^{6,127}$-lysozyme we clearly observe two classes of amides; a majority which have protection factors substantially lower than in intact lysozyme, and a small subset of 17 amides which remain protected to an extent comparable with that of the unmodified protein, and hence which are apparently relatively unaffected by the removal of the Cys6–Cys127 disulphide bridge. Whilst the most highly protected amides in CM$^{6,127}$-lysozyme correlate well with those most slowly exchanging in the unmodified protein, the slow exchange core is even more pronounced in the three-disulphide derivative. This may arise in part from the fact that the protection factors of the most protected amides in wild-type lysozyme are underestimates based on the observation that no exchange occurred in these amides over the time scale (70 days) of the experiment (Radford *et al.*, 1992a), hence over-emphasising the degree of protection in the core of CM$^{6,127}$-lysozyme. Nonetheless, the results show clearly that more frequent local unfolding events are occurring in CM$^{6,127}$-lysozyme than for its unmodified counterpart, especially in the regions close to the modification, under the conditions employed. It has been suggested that the most highly protected amides in the native state of a protein represent the earliest structural region to form during refolding (Woodward, 1994). This theory would predict therefore, that the B- and C-helices and the third strand of the main β-sheet in CM$^{6,127}$-lysozyme, should be the earliest regions to fold. The results described in Chapter Four, however, do not support this prediction; indeed residues Trp63, Cys64 and Ile78 are highly protected from exchange in the native protein but are in fact the last amides to attain full protection from exchange during folding. If this region is the earliest structure to form during folding therefore, it cannot be of sufficient stability to protect against hydrogen exchange.

# Chapter Four

# Kinetic Consequences of the Removal of a Disulphide Bridge on the Folding of Hen Lysozyme.

## 4.1 *Introduction*

Intensive efforts have been made over a number of years to elucidate the folding pathways of proteins. Recent advances in instrumentation have allowed a number of complementary techniques to be applied to mechanistic and kinetic studies of the folding of small globular proteins (Evans & Radford, 1994). One method is to trap intermediates as they are formed, for instance by blocking cysteine residues during oxidative refolding of disulphide-containing proteins (Creighton, 1992a; Weissman & Kim, 1992). However, due to the transient nature of many intermediate species involved in the formation of non-covalent secondary and tertiary interactions, alternative methods of detection are often necessary. Elegant experiments using pulsed hydrogen exchange labelling methods in conjunction with NMR have enabled detailed structural characterisation of such transiently formed species (Roder *et al.*, 1988; Udgaonkar & Baldwin, 1988; Radford *et al.*, 1992b). More recently, the use of electrospray ionisation mass spectrometry (ESI-MS) has been used to give information on populations of species present during the refolding of hen lysozyme (Miranker *et al.*, 1993). In conjunction with these techniques, rapid mixing optical methods have enabled information to be gained about global folding characteristics (Ikeguchi *et al.*, 1986a; Ptitsyn *et al.*, 1990; Chaffotte *et al.*, 1992; Elöve *et al.*, 1992; Evans & Radford, 1994; Itzhaki *et al.*, 1994). Structural studies of stable, partially folded species have also been important in giving insight into the nature of folding intermediates (Baum *et al.*, 1989; Harding *et al.*, 1991; Buck *et al.*, 1993; Fan *et al.*, 1993) and site-directed mutagenesis has been employed to study the effects of specific interactions during folding (Fersht, 1993; Fersht & Serrano, 1993; Matthews, 1993).

Several common features have emerged from these studies, particularly that compact collapsed states with a significant amount of secondary structure content form very rapidly for a number of proteins (Kuwajima *et al.*, 1985; Elöve *et al.*, 1992; Radford *et al.*, 1992b; Jennings & Wright, 1993; Khorasanizadeh *et al.*, 1993; Kuwajima *et al.*, 1993; Varley *et al.*, 1993). These species have at least some characteristics of partially folded states observed at equilibrium for a number of proteins under mildly denaturing conditions (Baum *et al.*, 1989; Hughson *et al.*, 1990; Harding *et al.*, 1991; Buck *et al.*, 1993). Similarly, pulse labelling experiments coupled with stopped flow circular dichroism in the far-UV have shown that helical regions and simple motifs are often formed faster than more complex structures such as β-sheets (Radford *et al.*, 1992b; Varley *et al.*, 1993).

The kinetic folding of hen egg-white lysozyme has been extensively studied using a variety of different techniques (Radford *et al.*, 1992b; Dobson *et al.*, 1994). The first detectable step (≤ 2 ms) involves very rapid formation of a collapsed state containing extensive secondary structure, although this structure is clearly not sufficiently stable to afford substantial protection from amide exchange. A majority of molecules in the refolding mixture pass through an intermediate species in which the four α-helices and a C-terminal $3_{10}$ helix (the α-domain) form persistent structure cooperatively, substantially protecting (protection factor ≥ 500) the amides in this domain from exchange. In a slower step, the remainder of the protein (the β-domain), including two β-sheet regions, a long loop and a $3_{10}$ helix, folds to form the native structure. The overall folding process, under the conditions employed for these experiments, is more complex, however, since parallel pathways are involved; a proportion of the molecules become substantially protected in both α- and β-domains forming apparently native-like structure on a faster time scale (Radford *et al.*, 1992b; Miranker *et al.*, 1993).

Lysozyme has four disulphide bridges which remain intact in the studies of the wild-type protein described in this chapter (Figure 1.2). Two of these (Cys6–Cys127 and Cys30–Cys115) are located in the α-domain, one (Cys64–Cys80) within the β-domain, and one (Cys76–Cys94) linking the two domains. Recent studies on the effects of disulphide bridges on folding rates have indicated that introduction of disulphides can dramatically alter the kinetics of folding. For example, an 850-fold increase in refolding rate was observed for a modified form of subtilisin BPN′, in which a disulphide was introduced (Strausberg *et al.*, 1993), and was explained in terms of a decrease in configurational entropy in the transition state for folding. Disulphide bridges introduced into barnase have also been shown to affect folding kinetics by stabilising either the intermediate or transition states of folding depending on their precise location in the structure (Clarke & Fersht, 1993). In this chapter detailed kinetic studies are described of the folding of a derivative of hen lysozyme, in which one of the disulphide bridges located in the α-domain, Cys6–Cys127, has been selectively removed by chemical modification; the three remaining disulphide bonds remain intact in their wild-type pairings.

We have previously shown by differential scanning calorimetry that the decreased thermal stability of this three-disulphide derivative with respect to the intact protein is primarily an entropic effect, as a result of increased disorder in the denatured state in the absence of the constraining disulphide bridge (Chapter 3). This linkage is of particular interest for the folding of lysozyme, since it connects the N- and C-terminal regions in the α-domain of the protein. Using hydrogen exchange pulse labelling techniques, in conjunction with stopped flow circular dichroism and intrinsic tryptophan fluorescence, we have investigated in detail the folding of this three-disulphide derivative. By comparison with the extensive data available for the unmodified protein we discuss the importance of this disulphide bridge in the folding pathway of hen lysozyme and, more generally, the importance of disulphide bonds in protein folding.

## 4.2 *Hydrogen exchange pulse labelling experiments monitored by NMR*

The thermal denaturation of CM$^{6,127}$-lysozyme has previously been shown to be highly co-operative (Chapter 3). Unfolding of CM$^{6,127}$-lysozyme by guanidinium chloride at pH 3.8 was monitored in the present work by far-UV circular dichroism, and found, like the unmodified protein, also to be a cooperative process under these conditions (Figure 4.1). However, consistent with the lower stability of CM$^{6,127}$-lysozyme the midpoint was reduced from 4.5M GuHCl for wild-type lysozyme to 1.5M for the three-disulphide derivative. From these data an initial denaturant concentration of 3M GuHCl was chosen for the following refolding experiments to ensure that CM$^{6,127}$-lysozyme was completely unfolded and yet a sufficiently low final concentration (0.27M) after an 11-fold dilution to allow complete refolding of the protein, and to maximise the possibility of detection of any transient, marginally stable intermediates.

In order to make a direct comparison of the folding of CM$^{6,127}$-lysozyme with that of the intact protein, it is vital that the final denaturant conditions employed for refolding are identical (Kato *et al.*, 1981; Kotik *et al.*, 1995). Control experiments were performed with each of the techniques described below, in which unmodified lysozyme, denatured in 6M GuHCl, was refolded by a 21-fold dilution into sodium acetate buffer, pH 4.5, in order to achieve a final denaturant concentration of 0.27M, identical to the conditions employed for the refolding studies on CM$^{6,127}$-lysozyme. Further preliminary experiments were carried out to investigate the dependence of wild type lysozyme refolding on GuHCl concentration. These results are discussed in Appendix C.

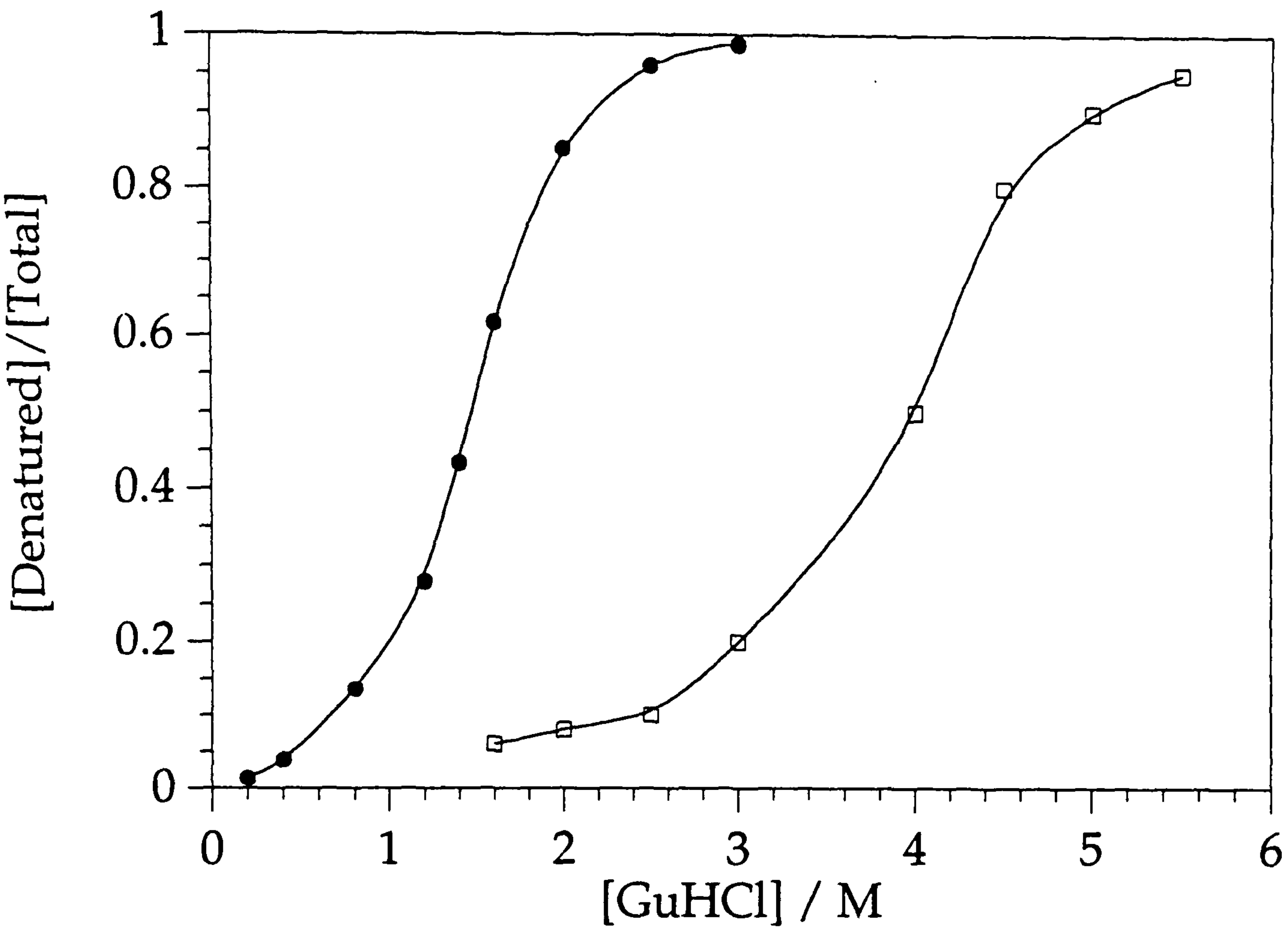


**Figure 4.1** Denaturation of $CM^{6,127}$-lysozyme (●) and unmodified hen lysozyme (□) by GuHCl at 20°C, pH 4.5, monitored by circular dichroism at 222 nm. Data were normalised to give concentration of denatured protein as a fraction of the total protein.

Refolding rates and kinetic amplitudes for wild-type lysozyme refolded into 0.27 M and 0.55 M GuHCl are compared in Table 4.1. No significant differences were detected as a result of refolding at pH 4.5, rather than pH 5.5 as previously published (Radford *et al.*, 1992b), consistent with there being few ionisable groups having $pK_a$ within this pH range.

Quenched flow experiments were performed to probe the kinetic refolding of CM$^{6,127}$-lysozyme at 20°C, pH 4.5, where the native protein is stable and hydrogen exchange in the refolding molecules is sufficiently slow compared with the time course studied. Kinetics of amide protection during refolding could thus be measured unambiguously for 42 backbone amides. A control in which native, deuterated CM$^{6,127}$-lysozyme was exposed to an 8.4 ms labelling pulse at pH 9.5, followed by the usual quench and washing steps, confirmed that labelling of the native protein by such a pulse is negligible for the 42 amides monitored.

| Protein refolding conditions/ Experiment | HEWL[a]<br>6M → 0.55M | HEWL<br>6M → 0.27M | $CM^{6,127}$<br>3M → 0.27M |
|---|---|---|---|
| CD (225 nm) | 80% < 2 ms<br>30% 25±12 ms<br>-10% 345±70 ms | 80% < 2 ms<br>30% 15±8ms<br>-10% 150±30 ms | 80% < 2 ms<br><br>20% 200±20ms |
| CD (289 nm) | 100% 330±60 ms | 100% 235 ± 10ms | 100% 205 ± 20ms |
| Trp fluorescence | 120 % < 1 ms<br>290 % 25±3 ms<br>-310 % 340±20 ms | 120% < 1 ms<br>290% 20±2 ms<br>-310% 190±20 ms | 80 % < 1 ms<br><br>20 % 267± 15ms |
| MeU-diNAG fluorescence | 100% 350±50 ms | 100% 210±30 ms | 100% 230±30 ms |
| NMR | α-48% 5±3 ms<br>52% 65±25 ms<br>β-29% 11±5 ms<br>71% 340±180 ms | α-45% 6±4 ms<br>55% 50±30 ms<br>β-30% 10±5 ms<br>70% 230±110 ms | α-34% 13±4 ms<br>66% 278±140 ms<br>β-35% 19±6 ms<br>65% 347±180 ms |
| ESI-MS | n.d.[b] | n.d. | 31% 16±5 ms<br>69% 300±30 ms |

**Table 4.1.** Comparison of folding time constants and kinetic amplitudes of unmodified hen lysozyme and $CM^{6,127}$-lysozyme. Errors indicated represent not only the error of the fit but also the average of several data sets.

[a] Results from previously published data are shown for comparison (Radford *et al.*, 1992b; Itzhaki *et al.*, 1994).

[b] n.d., not determined.

Kinetic profiles for a representative selection of individual amides are shown in Figure 4.2. All of the data fit to a double exponential process; 20-30% of the amplitude having a time constant of ~20 ms, with a slower phase ($\tau$ ~300ms) accounting for a further 50-60% of the total protection. These amplitudes and rates are similar for every measured amide, as shown in Table 4.2, suggesting a high degree of cooperativity throughout the folding molecule, with all amides which could be monitored being protected in a concerted fashion. No intermediate partly folded species, offering protection to only a proportion of the amides followed, was detected at any point during refolding. This is in marked contrast to the previously published results of experiments carried out on wild-type lysozyme (Radford *et al.*, 1992b), and to those carried out here on the unmodified protein under identical refolding conditions as to those of $CM^{6,127}$-lysozyme (see Figure 4.2), in which the hydrogens of the $\alpha$-domain of the protein are protected more rapidly ($\tau$ ~65ms) than those of the $\beta$-domain ($\tau$ ~300ms). Approximately 20% of the population folds very slowly in the case of both $CM^{6,127}$-lysozyme and the unmodified protein; this has been attributed previously at least in part to peptide bond *cis-trans* isomerisation involving the two proline residues in the lysozyme sequence (Kato *et al.*, 1982), but no further analysis of these very slow steps is made here.

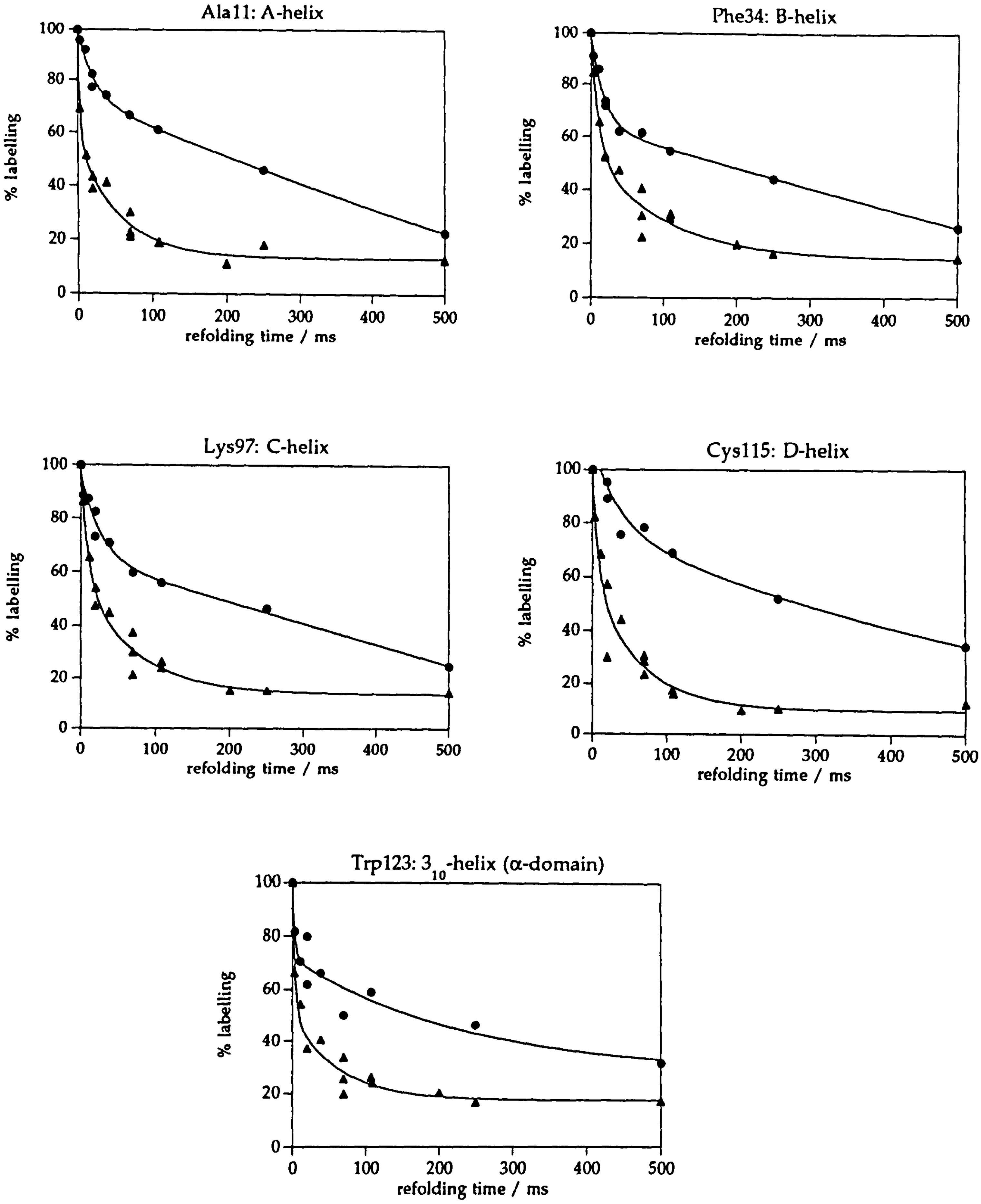


Figure 4.2a. Time courses for protection of representative amides in the α-domain during refolding of CM$^{6,127}$-lysozyme (●) and unmodified hen lysozyme (▲). The extent of labelling is plotted against time after the initiation of refolding. The curves drawn are the best fit to the sum of two exponentials over the time course studied.

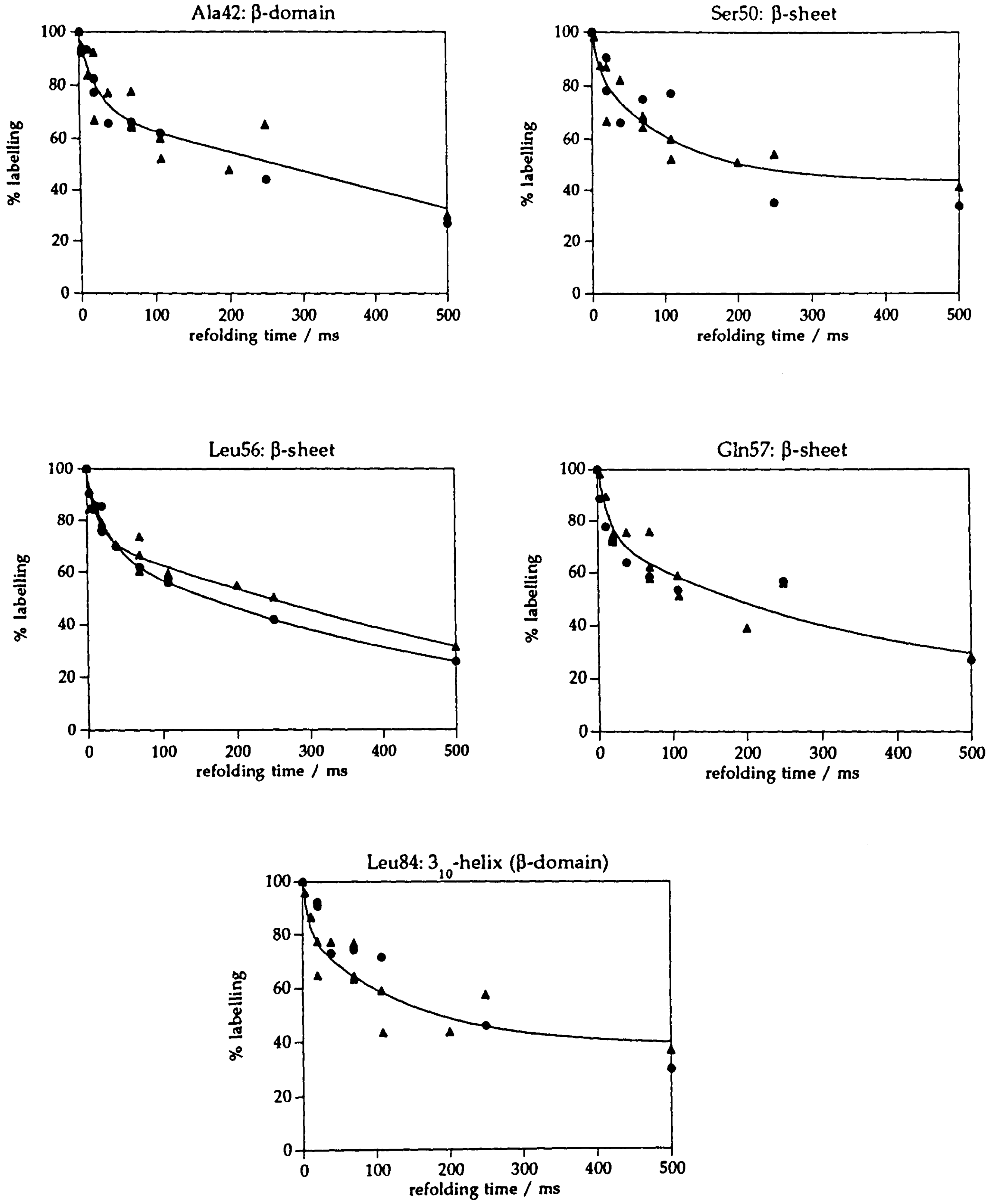


Figure 4.2b. Time courses for protection of representative amides in the β-domain during refolding of CM$^{6,127}$-lysozyme (●) and unmodified hen lysozyme (▲). The extent of labelling is plotted against time after the initiation of refolding. The curves drawn are the best fit to the sum of two exponentials over the time course studied.

| Residue | Fast amplitude (%) | Fast time constant (ms) | Slow amplitude (%) | Slow time constant (ms) | Residual (%) |
|---|---|---|---|---|---|
| Phe 3 | 17 | 26 | 65 | 253 | 16 |
| Ala 9 | 21 | 9 | 52 | 212 | 29 |
| Ala 10 | 34 | 14 | 37 | 295 | 27 |
| Ala 11 | 21 | 12 | 62 | 251 | 20 |
| His 15 | 44 | 8 | 34 | 384 | 27 |
| Leu 17 | 30 | 13 | 46 | 310 | 24 |
| Tyr 23 | 29 | 18 | 61 | 317 | 18 |
| Trp 28 | 36 | 4 | 53 | 358 | 24 |
| Val 29 | 33 | 5 | 40 | 215 | 26 |
| Ala 32 | 25 | 13 | 67 | 272 | 9 |
| Phe 34 | 32 | 15 | 60 | 468 | 7 |
| Glu 35 | 30 | 8 | 53 | 356 | 17 |
| Ser 36 | 35 | 14 | 52 | 290 | 33 |
| Asn 37 | 29 | 29 | 56 | 468 | 20 |
| Phe 38 | 25 | 13 | 56 | 347 | 15 |
| Asn 39 | 20 | 20 | 66 | 254 | 17 |
| Thr 40 | 24 | 19 | 53 | 388 | 24 |
| Gln 41 | 26 | 15 | 61 | 368 | 14 |
| Ala 42 | 25 | 18 | 64 | 383 | 11 |
| Asn 44 | 27 | 19 | 68 | 357 | 18 |
| Asn 46 | 29 | 13 | 59 | 326 | 14 |
| Ser 50 | 33 | 18 | 62 | 403 | 20 |
| Asp 52 | 19 | 27 | 71 | 441 | 9 |
| Leu 56 | 28 | 22 | 58 | 336 | 13 |
| Gln 57 | 32 | 12 | 46 | 403 | 21 |
| Ile 58 | 32 | 22 | 50 | 457 | 18 |
| Arg 61 | 30 | 25 | 57 | 288 | 16 |
| Trp 63 | 37 | 9 | 30 | 334 | 37 |
| Cys 64 | 25 | 7 | 55 | 185 | 18 |
| Asn 65 | 25 | 14 | 49 | 272 | 25 |
| Cys 76 | 18 | 23 | 68 | 257 | 19 |
| Ile 78 | 50 | 6 | 25 | 399 | 24 |
| Ala 82 | 29 | 20 | 53 | 327 | 18 |
| Leu 84 | 23 | 25 | 64 | 259 | 23 |
| Val 92 | 26 | 8 | 58 | 190 | 15 |
| Ala 95 | 30 | 24 | 53 | 266 | 20 |
| Lys 97 | 26 | 20 | 55 | 327 | 17 |
| Val 99 | 32 | 19 | 47 | 230 | 21 |
| Trp 108 | 13 | 10 | 72 | 349 | 15 |
| Trp 111 | 19 | 19 | 36 | 255 | 24 |
| Cys 115 | 38 | 17 | 59 | 303 | 26 |
| Ile 124 | 40 | 11 | 60 | 316 | 22 |

**Table 4.2.** Fits to the sum of two exponentials of kinetic refolding of $CM^{6,127}$-lysozyme, 3M->0.27M GuHCl, pH 4.5, 20°C.

### 4.2.1 Pulse intensity and pulse length experiments.

A common feature of the protection profiles of the two proteins is their biphasic nature. This suggests that CM$^{6,127}$-lysozyme, like its unmodified counterpart (Radford *et al.*, 1992b; Miranker *et al.*, 1993), folds along alternate, parallel folding pathways. This was investigated further by varying the intensity and the length of the labelling pulse. Pulse intensity experiments were performed at refolding times of 20 ms and 200 ms by varying the pH of the labelling pulse from pH 9.0 to pH 10.5. Data for representative amides are shown in Figure 4.3. At both refolding times no significant change in labelling (<5%) was observed over this pH range for the majority of amides monitored, enabling a lower limit of 500 to be put on protection factors of those amides in molecules which are protected on these time scales. The possibility of an $EX_1$ exchange regime (Hvidt & Nielsen, 1966) was eliminated by varying the length, rather than the pH of the labelling pulse (Roder *et al.*, 1988; Udgaonkar & Baldwin, 1988). Experiments were carried out in which a 16 ms labelling pulse was applied after 23 ms of refolding, and a 24 ms pulse after 108 ms of refolding. In both these cases, no significant dependence on labelling time was observed (<5 % difference in extent of labelling from experiments using an 8.4 ms pulse). This indicates that CM$^{6,127}$-lysozyme, like the wild type protein, folds via parallel folding pathways with the branch point being before any protection from exchange occurs. The Cys6–Cys127 disulphide bridge is not, therefore, the cause of these alternative pathways.

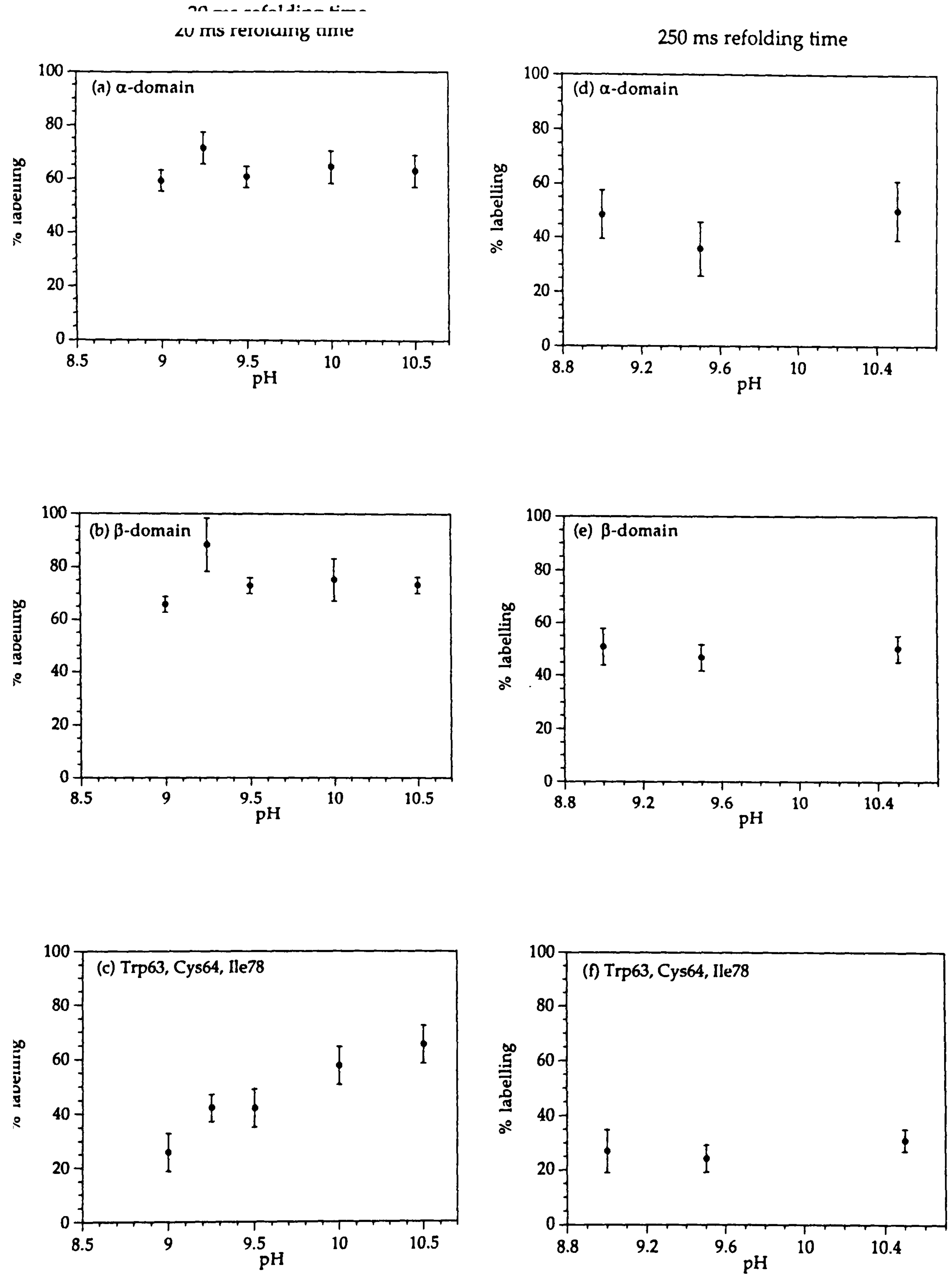


**Figure 4.3.** The pH dependence of the extent of labelling during an 8.3 ms pulse 20 ms or 250 ms after the initiation of refolding. The average proton occupancies for amide hydrogens in the α-domain (a,d), β-domain (b,e) and the three residues Trp63, Cys64, Ile78 (c,f) are shown.

Only three backbone amides were measurably affected by the intensity of the labelling pulse at the 20 ms refolding time point, namely those of Trp63, Cys64 and Ile78. This suggests that although the fast folding phase affords a high degree of protection to all of the amides in 20-30% of the folding population, these three residues are only partially protected in the species present at this time under these labelling conditions. Behaviour of this kind has also been observed for the same residues in the folding of unmodified lysozyme (Radford *et al.*, 1992b), and for the corresponding residues in human lysozyme (Hooke *et al.*, 1994). These three residues are located in the long loop of the native lysozyme structure (residues 61 to 80), a region constrained by the two disulphide bridges Cys64–Cys80 and Cys76–Cys94 (Figure 1.2). Partial protection of these three residues suggests that although protective structure involving these residues forms rapidly, it is not highly stable against hydrogen exchange under these labelling conditions. Partial protection in this region may be a result of a lack of persistent native-like structure in the vicinity of these disulphide bridges, requiring a rearrangement step to give the fully protected native structure.

### 4.2.2 Protection of Trp indole protons.

Protection of tryptophan indole protons is also a valuable probe of structure formation, particularly since they may be used as indications of tertiary structure formation in the refolding molecules. Two of the six indole protons (Trp111 and Trp108) of CM$^{6,127}$-lysozyme are sufficiently slow to exchange in the native state to be used as probes for this experiment. In the native state these indoles are buried within the hydrophobic box region of the protein (Smith *et al.*, 1993). Protection profiles for tryptophan indole protons were measured by integration of the N1H peaks of Trp108 and Trp111 from one-dimensional NMR spectra, and normalising the areas to those of two resolved resonances (Tyr23 εCH and Tyr53 εCH).

The indole proton of Trp108 was found to be protected very rapidly, indeed more rapidly than any of the backbone amides in the protein, but the amplitude of any further change was too small to quantify. Data for Trp111, however, could be fitted to a single exponential function with a time constant of 300 ± 10 ms, but the amplitude of this phase is only 50% of the total change expected (Figure 4.4), indicating that within the dead time of the experiment (≤3 ms), approximately half of the population is already sufficiently structured to afford protection to these indole hydrogens. This indicates that although the majority of the molecules do not form stable structure involving backbone amides on this time scale, there is clearly a rapid event, presumably involving substantial hydrophobic collapse, which renders these tryptophan side chains significantly protected from exchange on a very fast time scale. Varying the pH of the labelling pulse from pH 9.0 to pH 10.5 at 20 ms refolding time gave rise to a change of ~20% in the extent of labelling, indicative of partial protection of these indole protons. It was not possible to calculate protection factors for these indole protons since the pH dependence could not be fit to a simple mechanism, but an upper limit of 50 can be estimated. Thus extensive hydrophobic collapse occurs very rapidly, but at these early times the structure around these tryptophan side chains is less protective than that formed at later times for the amide hydrogens.

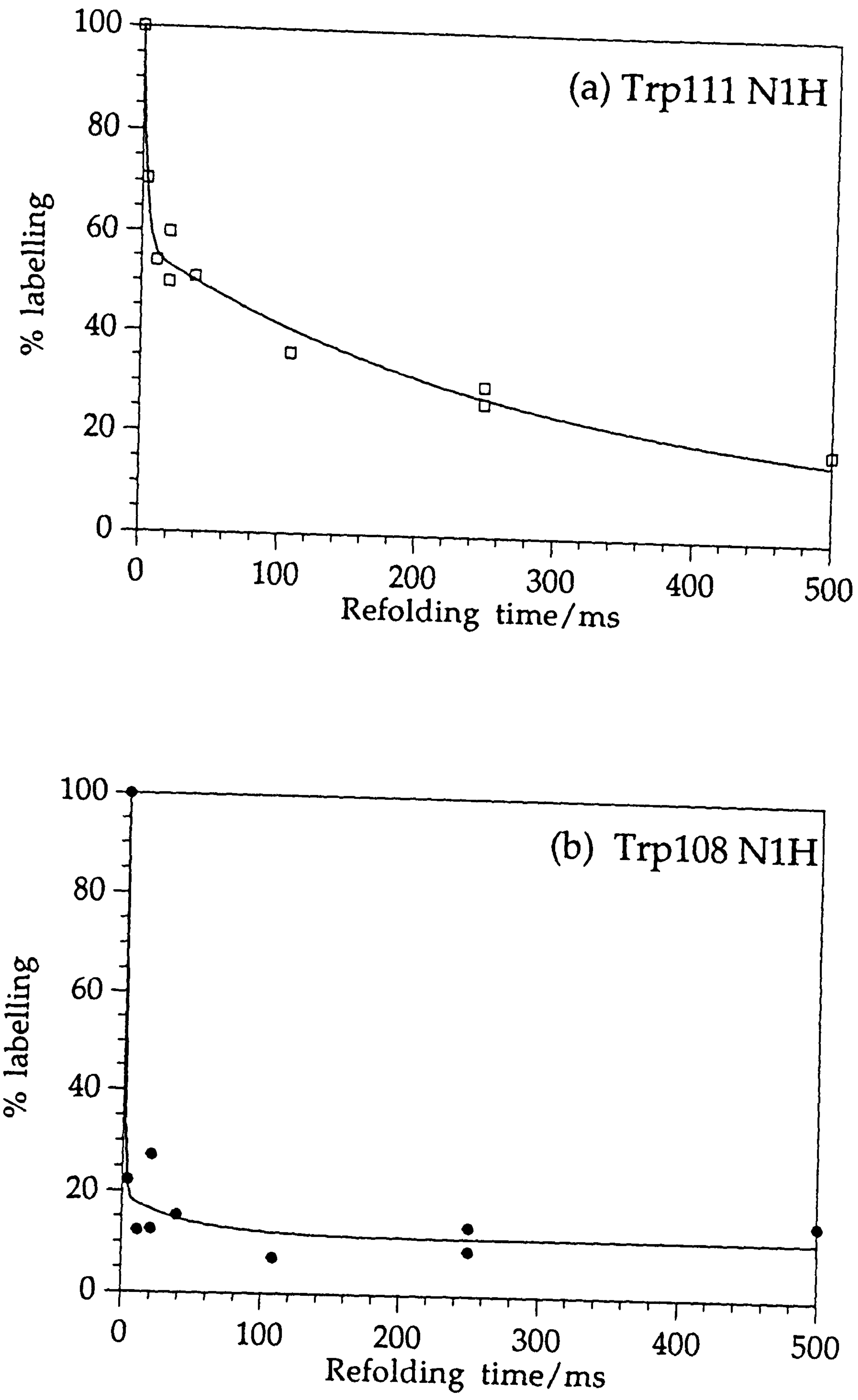


**Figure 4.4.** Protection profiles for indole protons of (a) Trp111 and (b) Trp108. The extent of labelling is normalised to the intensity of the zero time point spectrum.

## 4.3 *Hydrogen exchange pulse labelling monitored by ESI-MS*

ESI-MS has recently been developed as a means of studying protein folding through hydrogen exchange phenomena that is complementary to NMR (Miranker *et al.*, 1993). Although NMR is a residue-specific technique, it can only reveal average proton occupancy at each amide site averaged over all molecules in the sample. Mass spectrometry, by contrast, can distinguish populations of protein molecules in a sample, which differ in mass by only a few Da. This method was particularly important in folding studies of CM$^{6,127}$-lysozyme since the NMR data show that the α- and β-domains fold to protective structure with identical kinetics in both phases of folding. This behaviour is consistent with two fundamentally different folding mechanisms, involving either cooperative folding of the two domains, or alternatively, that the α- and β-domains fold independently but with identical kinetics. It is not possible to distinguish between these two possibilities by NMR alone. By contrast, as has been shown for intact lysozyme (Miranker *et al.*, 1993), ESI-MS can distinguish between these two possibilities by detection of any protected species which are only partially deuterium labelled.

Figure 4.5 shows mass spectra obtained from samples of CM$^{6,127}$-lysozyme labelled under identical conditions to those in the NMR experiments at various times after the initiation of refolding. At the zero time point, only a single peak is observed at mass 14431 Da, corresponding to the unprotected CM$^{6,127}$-lysozyme. The small peak at 14489 Da results from a species carboxymethylated at His15 which occurs as a minor impurity in the preparation of CM$^{6,127}$-lysozyme. The mass of CM$^{6,127}$-lysozyme calculated from the amino-acid sequence is 14423 Da, the difference in mass from that observed for the zero time point arising from the 5% residual deuterium content in the labelling buffer, which is assumed to be distributed randomly throughout the protein. As the refolding time is increased, a second peak emerges with a mass 55 Da greater than that observed for the zero time point. At the same time, the intensity of the peak arising from the fully unprotected protein diminishes. The peak at mass 14486 Da corresponds to a fully protected species in which 55 slowly exchanging hydrogens under the conditions employed, become protected from exchange. The mass of this species is heavier than the number of amides (38) which could be individually monitored by NMR, reflecting the problems of NMR spectral overlap and exchange of amide protons during the acquisition of the NMR spectra, which limits the number of amides which could be reliably monitored in the native protein. It is, however, consistent with the number of slowly exchanging amides anticipated under these solution conditions from the data in Chapter 3 (Figure 3.5), and may also include some slowly exchanging side chain hydrogens.

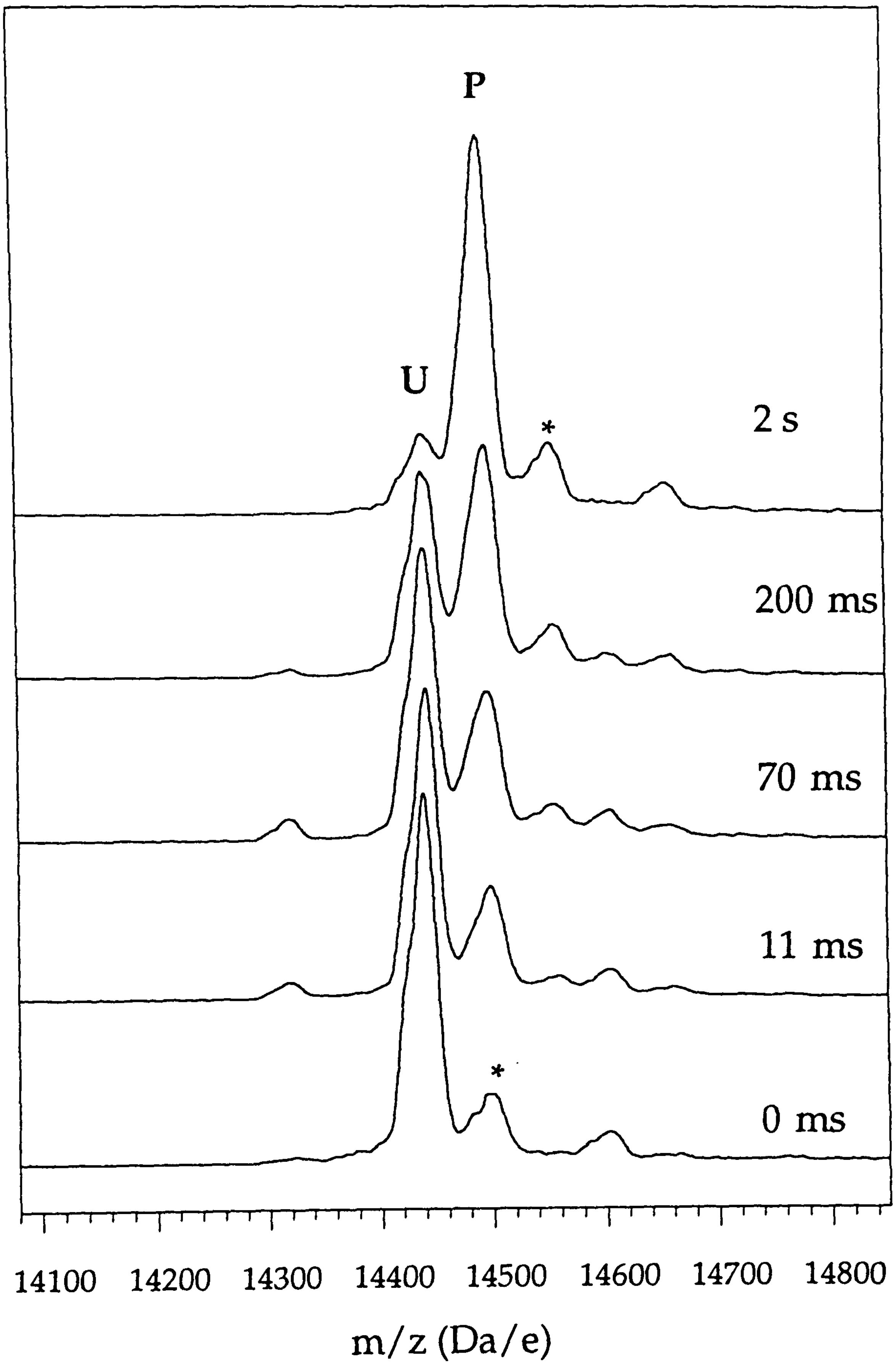


**Figure 4.5.** Time course for refolding of CM$^{6,127}$-lysozyme as monitored by pulsed amide hydrogen exchange labelling in conjunction with ESI-MS. The peak marked * arises from a species carboxymethylated at His15. The mass of this species is 14489 Da in the unprotected state and 14544 Da in the fully protected state. Peaks of low intensity at higher mass arise from Na$^+$ adducts.

The change with time of the relative heights of the two peaks present in the mass spectra is clearly biphasic and is fully consistent with a fast phase with a time constant of ~20 ms with an amplitude of 20-30%, and a slower phase protecting the remainder of the molecules with a time constant of ~300 ms, in accord with the NMR data presented here. No species of intermediate mass is observed, despite the fact that protection in the α- or β-domains should give mass shifts of 28 or 22 Da respectively. These species should be clearly evident in the mass spectra providing that they are populated to at least the 5% level (Miranker *et al.*, 1993). This result confirms that the protein indeed folds cooperatively without formation of any detectable protected transient intermediate states.

## 4.4 *Stopped flow optical measurements*

### 4.4.1 Circular dichroism in the far-UV

Circular dichroism of proteins in the far ultra-violet region arises primarily from secondary structure content and is dominated at 225 nm by α-helical structure. Refolding of CM$^{6,127}$-lysozyme monitored at 225 nm indicates the presence of at least two kinetic phases (Figure 4.6). The CD signal achieves some 80% of its native intensity within the 2 ms dead time of measurement, suggesting that a significant amount of the native helical content in the protein is developed within a very short time after the initiation of refolding. The kinetic trace observed for the remainder of the folding event can be fitted to a single exponential with a time constant of 200 ± 35 ms. These two characteristics are virtually identical to those seen in the refolding of the unmodified protein under identical refolding conditions. These data, however, demonstrate a marked difference between the folding of the two proteins in that for the wild-type protein a second fast phase is observed ($\tau$~15 ms) yielding an ellipticity at 225 nm actually greater than that of the native protein. This overshoot has previously been attributed to a contribution to the spectrum from disulphide bridges possibly in a non-native conformation in a partially folded species (Chaffotte *et al.*, 1992). These results, therefore, strongly suggest either that removal of the Cys6–Cys127 disulphide prevents formation of these interactions or that this disulphide bridge is the major chromophore contributing to the spectrum at this wavelength. The slow phase giving rise to ellipticity characteristic of a native-like state has, however, a remarkably similar time constant for both proteins.

### 4.4.2 Circular dichroism in the near-UV

The near-UV signal at 289 nm in the CD spectrum of hen lysozyme in the native state arises primarily from tryptophan residues in fixed orientations (Goux & Hooker, 1980). Kinetic refolding of CM$^{6,127}$-lysozyme monitored at this wavelength adequately fit to a single exponential function, with a time constant of 205 ± 20 ms (Figure 4.7), in good agreement with the slow phase observed by CD at 225 nm. It is interesting to note that although hydrogen exchange protection of tryptophan indoles occurs very rapidly as described above, the fact that these do not give rise to a signal in the near-UV CD suggests either that fixed tertiary interactions are not formed at this early stage of folding or that those tryptophans which could be monitored by NMR do not contribute to the CD signal.

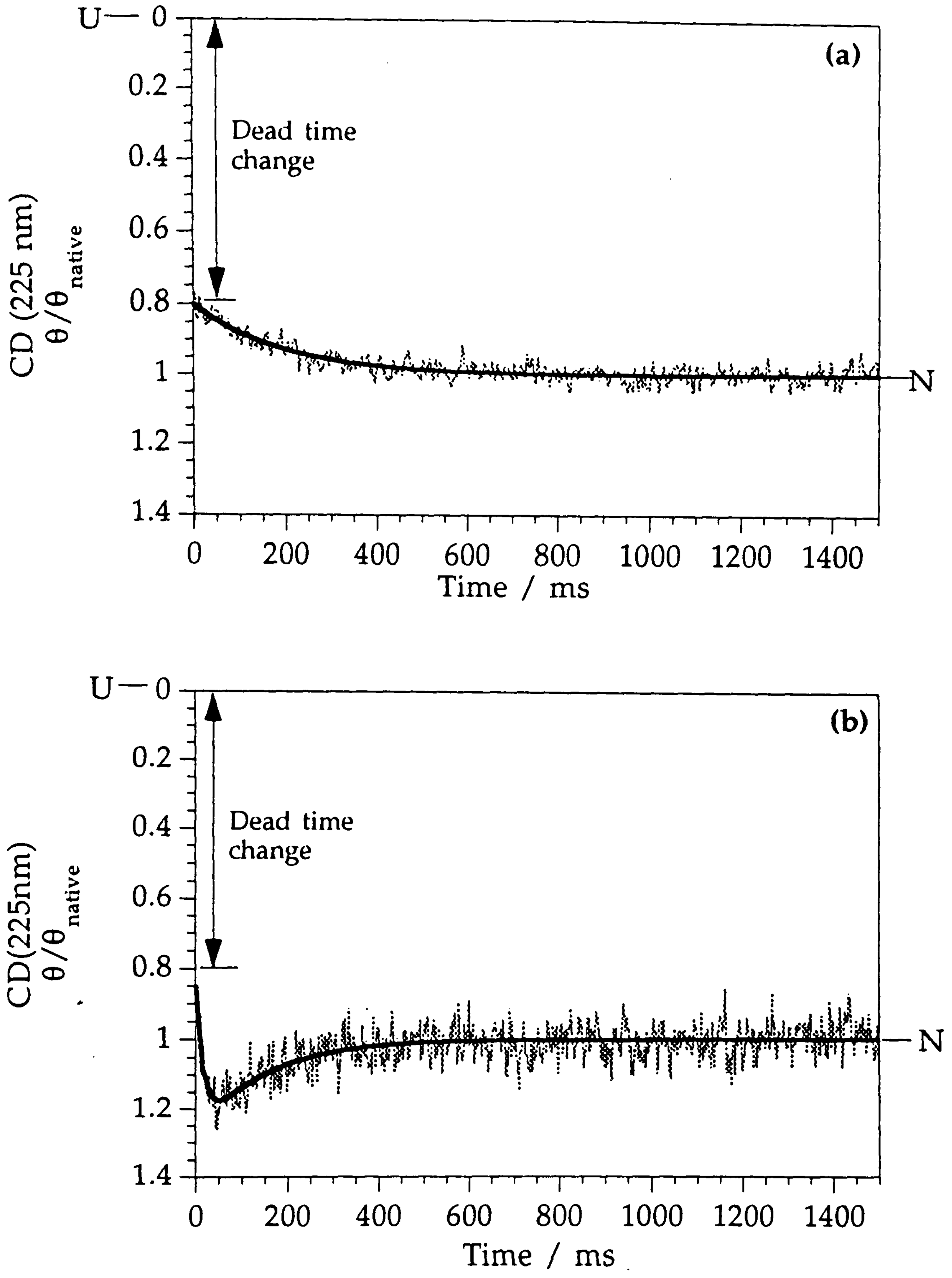


**Figure 4.6.** Refolding kinetics monitored by stopped flow circular dichroism in the far-UV at 225 nm. (a) CM$^{6,127}$-lysozyme, (b) hen lysozyme. Data have been normalised to the fraction of native state dichroism such that the value of the denatured state is 0 and that of the native state is 1.

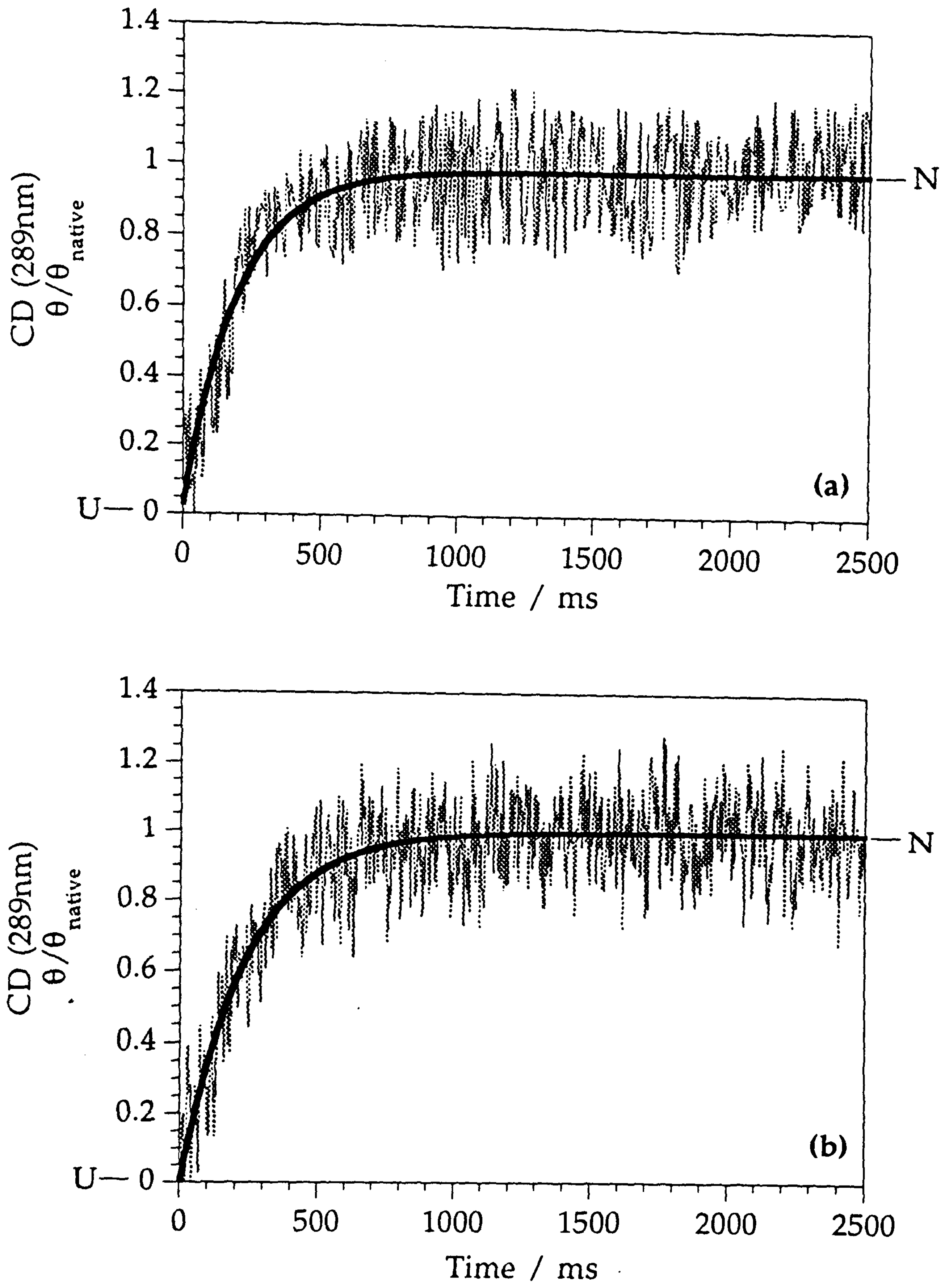


**Figure 4.7.** Refolding kinetics monitored by stopped flow circular dichroism in the near-UV at 289 nm. (a) CM$^{6,127}$-lysozyme, (b) hen lysozyme. Data have been normalised to the fraction of native state dichroism such that the value of the denatured state is 0 and that of the native state is 1.

### 4.4.3 Intrinsic tryptophan fluorescence

The intrinsic tryptophan fluorescence of lysozyme in the native state is thought to arise primarily from Trp62 and Trp108 (Imoto *et al.*, 1971). Figure 4.8 shows the kinetic trace of total intrinsic tryptophan fluorescence emission during refolding of CM$^{6,127}$-lysozyme. The kinetics exhibit two distinct phases, the first being a dead time change accounting for approximately 80% of the total change in fluorescence intensity. The remaining phase fitted well to a single exponential function, with time constant 267 ± 15 ms. This latter phase is similar in rate to that obtained by other optical techniques and to the slowest phase of protection observed by the pulse labelling technique. By contrast, refolding of the wild-type protein under these conditions exhibits three distinct kinetic phases. A significant dead time quenching event, accounting for almost all of the total expected amplitude, is followed by a fast phase ($\tau \sim 20$ ms) in which fluorescence is quenched beyond that of the native state. Recovery of native state fluorescence intensity then occurs on a 200 ms time scale. These observations are consistent with the rates of folding measured by CD described here, in data published previously (Radford *et al.*, 1992b; Itzhaki *et al.*, 1994), and are also in general accord with refolding studies of CM$^{6,127}$-lysozyme carried out under different conditions of pH and denaturant concentration (Denton *et al.*, 1994).

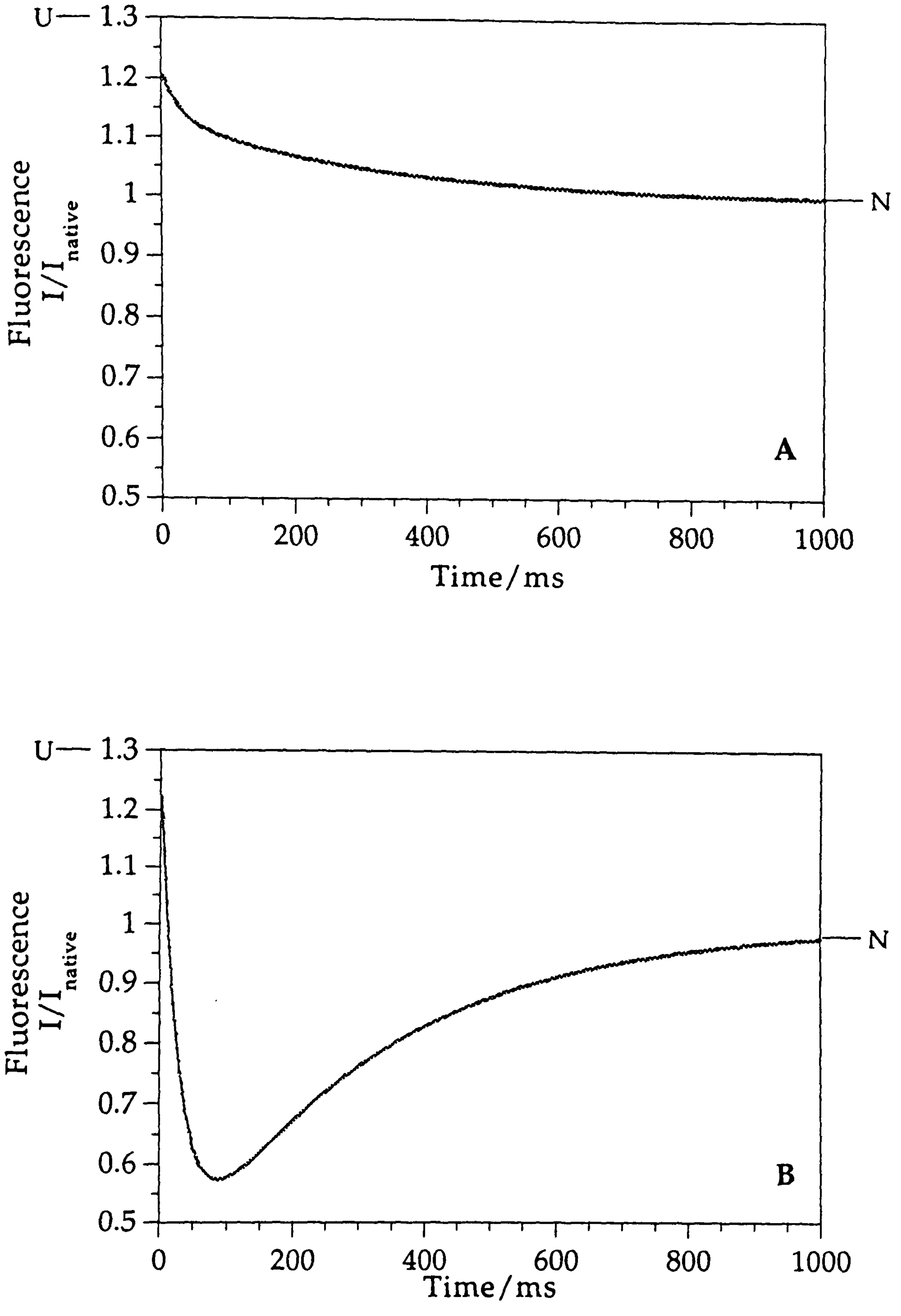


**Figure 4.8.** Refolding kinetics of CM$^{6,127}$-lysozyme (A) and unmodified lysozyme (B) monitored by stopped flow intrinsic tryptophan fluorescence. Data were normalised to the fraction of native state fluorescence, such that the value of the denatured state is 0 and that of the native state is 1.

### 4.4.4 Binding of MeUdiNAG

Binding of the fluorescently labelled inhibitor MeU-diNAG to CM$^{6,127}$-lysozyme as a probe for formation of the native protein was also monitored by stopped flow fluorescence (Figure 4.9). These data were fitted to a single exponential function with a time constant of 230 ± 30 ms, in very good agreement with the slow phase observed by intrinsic tryptophan fluorescence and circular dichroism in the near and far-UV No fast phase is observed, however, in this experiment, indicating that the fast phase of amide protection does not give rise to native active protein even though both the α- and β-domains are substantially protected from exchange. A similar result has been observed previously in the refolding process of the unmodified protein (Itzhaki *et al.*, 1994).

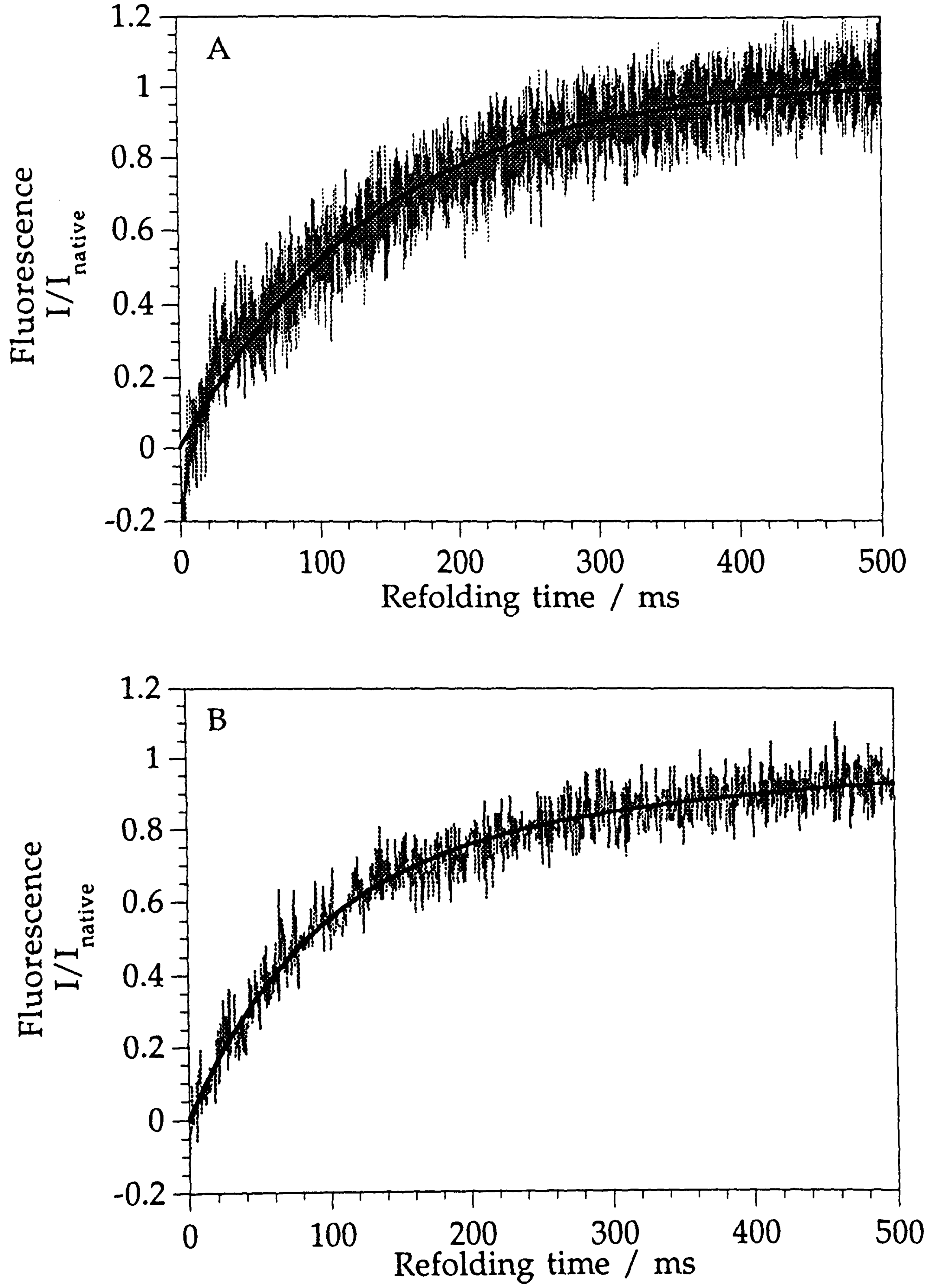


**Figure 4.9.** Refolding kinetics of CM$^{6,127}$-lysozyme (A) and unmodified lysozyme (B) monitored by binding of the fluorescent inhibitor MeUdiNAG measured by stopped flow fluorescence. Data were normalised to the fraction of native state fluorescence, such that the value of the denatured state is 0 and that of the native state is 1.

## 4.5 *Discussion*

The folding of lysozyme can be divided into at least three distinct stages, of which we have now some considerable understanding (Dobson *et al.*, 1994). The first steps occur within the dead time of measurement, on a millisecond time scale, and involve at least some degree of hydrophobic collapse as judged by intrinsic tryptophan fluorescence, fluorescence quenching with iodide, and ANS binding (Itzhaki *et al.*, 1994). The rapid, but partial, protection from exchange of tryptophan indole protons observed in this work is also consistent with such a conclusion. This collapse occurs concomitantly with formation of a significant amount of secondary structure, probably a near-native complement of α-helical structure, as indicated by CD in the far-UV region (Chaffotte *et al.*, 1992; Radford *et al.*, 1992b). No binding of MeU-diNAG inhibitor is observed on this time scale however, indicating, in accord with data in the near-UV CD and the lack of substantial amide hydrogen exchange protection, that few if any fixed tertiary interactions are present in the refolding mixture at this early stage. These states are thought to have similarities to equilibrium partially folded states, or molten globules, observed for some proteins in which a majority of the secondary structure content is present, yet no fixed tertiary interactions are detectable (Kuwajima, 1992; Baldwin, 1993; Buck *et al.*, 1993). These features are common both to the folding of wild-type lysozyme and to that of the three-disulphide derivative. Removal of a disulphide bridge in $CM^{6,127}$-lysozyme, therefore, appears to have no significant effect on the very rapid formation of the collapsed state. Interestingly in this regard, for all of the proteins whose refolding has been studied to date by stopped flow optical techniques a collapsed state has been inferred within a few milliseconds of the initiation of folding, despite the fact that the overall folding kinetics of these proteins differ by orders of magnitude (Kuwajima *et al.*, 1993). Thus in lysozyme the majority of molecules appear to fold completely within less than a second, while interleukin-1β takes several minutes to acquire its final native structure under the conditions studied (Varley *et al.*, 1993). Nevertheless, in both cases the majority of the secondary structural content of the native protein, at least as judged by far-UV CD, appears to be formed at the earliest possible point of detection, indicating that subsequent folding events are independent of the rate of this initial collapse.

The second observable stage in the folding process of intact lysozyme is the formation of an intermediate species in which the α-domain of the protein forms structure protected from exchange in the majority of molecules whereas the β-domain is still unstable and not protective towards hydrogen exchange. Protection of the α-domain occurs on a time scale similar to that of the development of an ellipticity in the far-UV CD which exceeds that of the native state and also to the development of non-native tertiary interactions which result in quenching of tryptophan fluorescence intensity to a value greater than that of the native state (Itzhaki *et al.*, 1994). During the refolding of CM$^{6,127}$-lysozyme none of these characteristics are observed. It is difficult to correlate quantitatively the results of the different techniques because, for example, the origin of the chromophores in the far-UV CD and fluorescence are not known and the different techniques may monitor very different molecular events that occur, by chance, on similar time scales. The fact that none of these characteristics is observed during the folding of CM$^{6,127}$-lysozyme suggests strongly, however, that formation of a stable α-domain intermediate is correlated with development of non-native interactions which quench the fluorescence of tryptophan residues, the majority of which are located in the α-domain, and gives rise to the overshoot in the ellipticity at 225 nm. It is interesting that this overshoot has been attributed, by reconstruction of the spectrum of the intermediate, to the contribution of disulphide bridge chromophores to the far-UV CD (Chaffotte *et al.*, 1992). It is possible that the lack of such an overshoot in the refolding of CM$^{6,127}$-lysozyme results from the fact that this interaction directly involves the Cys6–Cys127 disulphide bond. It is also possible, however, that the removal of the disulphide bridge destabilises the α-domain intermediate to such an extent that the non-native interactions giving rise to these optical properties, if they are populated, are too transient to be detected by these methods.

We have previously shown that removal of the Cys6–Cys127 disulphide bridge has a large effect on the configurational entropy of the native protein, favouring the unfolded state by some $70 \pm 5$ J K$^{-1}$ mol$^{-1}$ (Chapter 3, Cooper *et al.*, 1992). The effects of this are seen in the overall stability of the intact protein, and in the increased rate of hydrogen exchange (see above). It seems extremely likely, therefore, that a similar destabilisation will be experienced in the α-domain intermediate, in which this disulphide bridge is located. The conditions of the pulse labelling experiment require protection factors of any intermediate species to be greater than 20 in order to be detectable. Protection factors in the α-domain intermediate of intact lysozyme are >500, but the destabilisation arising from the removal of Cys6–Cys127 could well reduce these values such that the α-domain intermediate is too unstable to be detected in pulse labelling experiments. In a further series of experiments, therefore, refolding of CM$^{6,127}$-lysozyme was examined at low temperature and in the absence of denaturant (by dilution from the pH 2 denatured state of CM$^{6,127}$-lysozyme), where residual denaturant cannot be responsible for destabilisation of any intermediate species, in order to attempt to detect protection in an intermediate. Again, however, even under these conditions no evidence for a stable intermediate was observed. The stability of intermediate species in the absence of the Cys6–Cys127 disulphide bridge, therefore, is presumably too low even under these conditions to enable its detection by the experiments employed. Preliminary results (A. Matagne, personal communication) suggest that the addition of salt to the refolding buffer can afford sufficient stability to enable observation of the intermediate in CM$^{6,127}$-lysozyme. This is in accord with our findings that the pathway of folding is not significantly changed by the removal of the Cys6–Cys127 disulphide bridge, but that lack of this disulphide bridge destabilises the α-domain.

It is interesting to compare these findings with those of recent studies in which the structures of stable disulphide intermediates of α-lactalbumin were investigated (Ewbank & Creighton, 1993b). 3SS$_{cam}$ (in which the protein is reduced and carboxyamido-methylated at Cys6 and Cys120), is a three-disulphide derivative of α-lactalbumin, lacking the Cys6–Cys120 disulphide bridge homologous to that in CM$^{6,127}$-lysozyme. Structural studies by circular dichroism (Ewbank & Creighton, 1993b) and NMR (see Chapter 5) indicate that both the native state and the molten globule state of this protein (formed at low pH, or in the absence of the calcium ligand), is significantly less stable than its unmodified counterpart. Indeed the calculated difference in free energy of unfolding of the molten globule states from guanidinium chloride denaturation curves, $\Delta\Delta G(\alpha LA-CM^{6,120}\alpha LA)$ has been found to be 5 kJ $mol^{-1}$ (Ikeguchi *et al.*, 1992), which would result in a significant destabilisation of such a marginally stable species.

The third stage in the folding of lysozyme is formation of the native state of the protein. The overall folding rate of $CM^{6,127}$-lysozyme, remarkably, is quite similar to that of the unmodified protein ($\tau \sim 200$ ms), and is therefore apparently largely unaffected by the removal of the Cys6–Cys127 disulphide bridge. This suggests that the rate determining step for folding of both proteins does not reflect the stability of the α-domain intermediate, but involves stabilisation of the β-domain and docking of elements of secondary structure to form fixed tertiary contacts in the native folded state. This may in part reflect the fact that β-structure is intrinsically slower to form than α-helical structure (Dill *et al.*, 1993; Varley *et al.*, 1993). In the case of $CM^{6,127}$-lysozyme, the structure of the α-domain is sufficiently labile in the absence of the cross-linking disulphide bridge that it can only be observed concomitantly with stabilisation of the β-domain. This may be compared with the folding of intact lysozyme into high concentrations of denaturant, as described in Appendix C. In this case folding again appears highly cooperative, but the overall rate of folding is significantly slowed, presumably as a result of the effect which chemical denaturant has to destabilise the entire protein molecule rather than solely the α-domain as appears to be the case here. The fact, however, that some lysozyme molecules are able to protect amides rapidly in the fast phase of folding suggests that misfolding events may play an important role, and that only some topologies formed in this initial ensemble of collapsed states can proceed directly to the fully protected state.

It has been suggested that the most highly protected amides in the native state of a protein represent the earliest structural region to form during refolding (Woodward, 1994). This theory would predict, therefore, in the light of the hydrogen exchange data presented in Chapter 3, that the B- and C-helices and the third strand of the main β-sheet in CM$^{6,127}$-lysozyme, should be the earliest regions to fold. The results described here are not, however, consistent with this prediction; indeed residues Trp63, Cys64 and Ile78 are highly protected from exchange in the native protein but are in fact the last amides to attain full protection from exchange during folding. If this region is the earliest structure to form during folding therefore, it cannot be of sufficient persistent stability to protect against hydrogen exchange. This result could be a consequence of the rate limiting step in folding arising from misfolding events rather than intrinsic features of the native structure. Whatever the explanation, this result reaffirms the independent nature of the two folding domains in lysozyme. Further characterisation of the effects of specific residues in mutant and naturally occurring variant lysozymes will allow these individual submolecular kinetic events to be examined more fully as discussed in Chapter 6 (Hooke *et al.*, 1995).

One of the particularly interesting features in the study of protein folding is relating *in vitro* studies to the processes occurring within the living cell. Surprisingly little is known at a molecular level about how a protein attains its final native structure as it emerges from the ribosome and, for disulphide containing proteins, is transferred into the lumen of the endoplasmic reticulum. A number of helper proteins, including protein-disulphide isomerase, play an important part in folding and processing in this protein-rich cell compartment (Freedman, 1992; Jaenicke, 1993). As far as disulphide formation in the ER lumen is concerned, it is evident that the two cysteine residues must come into close proximity in a partly folded state in order to allow the oxidative formation of a disulphide bridge. As regions of the protein begin to fold, disulphides can form concomitantly, driving the folding process towards completion by stabilising local structure relative to the unfolded state. An important, perhaps dominant, contribution to the driving force is the greater reduction in entropy in more unfolded states that results from a cross-linking reaction.

In the case of hen lysozyme, extensive studies have suggested that the Cys6–Cys127 disulphide is the last to form, at least in the case of *in vitro* oxidative refolding experiments (Anderson & Wetlaufer, 1976). In this chapter, we have shown that the presence of this crosslink is not required for the complete folding of the protein since the three-disulphide species clearly folds efficiently to the native state. Instead its formation creates a constraining loop which confers stability on the final folded conformation and on partially folded intermediates by entropically disfavouring the unfolded state (Cooper *et al.*, 1992). The role of the disulphide bridge in this case appears to be one of maintenance of the folded structure rather than inducing structure into the polypeptide chain. Our data suggest that the Cys6–Cys127 disulphide bond does not dictate folding at all, but stabilises folded structure once formed. Further folding studies of proteins both in the presence and absence of disulphide bonds should determine whether this is a general role of disulphide bridges in the folding of proteins or whether disulphide bonds in different environments have very different roles.

# Chapter Five

# Structural studies of folding intermediates of bovine α–lactalbumin containing two, three or four disulphide bridges.

## 5.1 *Introduction*

The majority of small proteins are thought to unfold under equilibrium conditions via a highly cooperative two-state mechanism, involving no observable intermediates (Privalov & Khechinashvili, 1974; Pfeil & Privalov, 1976; Privalov, 1979). A number of proteins, however, exhibit more unusual unfolding characteristics, in that they do not exhibit classical two-state behaviour during the reversible unfolding of the protein (Wong & Tanford, 1973; Kuwajima *et al.*, 1976). Instead, under conditions such as high or low pH, high or low ionic strength, elevated temperature or intermediate denaturant concentration, a transition occurs from the well defined native structure to an ensemble of dynamic partially folded species, which have been termed "molten globules" (Dolgikh *et al.*, 1981; Ohgushi & Wada, 1983). The characteristic properties of these "compact globules with native-like secondary structure and with slowly fluctuating tertiary structure" were initially rigidly defined, but further examples have caused the definition to be somewhat relaxed, as reviewed by Christensen & Pain (1991). The fixed tertiary interactions of the native state are significantly reduced, whilst considerable amounts of secondary structure persist. This expansion of structure, although the protein remains globular, tends to lead to exposure of aromatic side chains and hence the ability of these partially folded states to bind hydrophobic dyes such as ANS (8-anilinonaphthalene sulphonate) (Semisotnov *et al.*, 1987). Exposure of non-polar groups may also be responsible for the lack of an observable heat capacity change on complete unfolding of many these intermediate states, since the major contribution to the $\Delta C_P$ is thought to be exposure to solvent of these buried hydrophobic side chains (Dolgikh *et al.*, 1981; Pfeil *et al.*, 1986; Kim & Baldwin, 1990; Lala & Kaul, 1992), although a $\Delta C_P$ has been measured for such a transition in the case of the unfolding of equine lysozyme (Morozova *et al.*, 1991; van Dael *et al.*, 1993). Such phenomena have been observed now for a variety of proteins, including carbonic anhydrase B (Wong & Tanford, 1973; Dolgikh *et al.*, 1984), $\alpha$-lactalbumin (Kuwajima *et al.*, 1976; Dolgikh *et al.*, 1981), bovine growth hormone (Brems & Havel, 1989), cytochrome *c* (Kuroda *et al.*, 1992), apomyoglobin (Hughson *et al.*, 1990) and, more recently, partially folded states of ubiquitin (Harding *et al.*, 1991), human lysozyme (Haezebrouck *et al.*, 1995) and hen lysozyme in trifluoroethanol (Buck *et al.*, 1993).

Partially folded states at equilibrium have been postulated to be similar to intermediates transiently observed during kinetic refolding experiments, and have been shown in several cases to exhibit very similar properties (Kuwajima, 1989; Kim & Baldwin, 1990; Ptitsyn *et al.*, 1990; Christensen & Pain, 1991; Dobson, 1992, 1994). For example the transient intermediate in the refolding of myoglobin identified by pulse labelling has been shown to resemble closely the partially folded state of apomyoglobin at low pH (Jennings & Wright, 1993). Similarly the partially folded structure of hen lysozyme in trifluoroethanol (Buck *et al.*, 1993) resembles the α-domain intermediate formed transiently during kinetic refolding from the denatured state in guanidinium chloride (Radford *et al.*, 1992b), in that the four native α-helices appear to be present and are protective against hydrogen exchange in this equilibrium state. As the structure and folding properties of more proteins in solution are investigated, the occurrence of these molten globule states, either as metastable states or as transient kinetic intermediates, is becoming more common and has been implicated as a general intermediate in the folding of small globular proteins (Ptitsyn *et al.*, 1990).

Since transient intermediates, by their very nature, cannot be directly studied, their characteristics can only be inferred from techniques such as stopped flow optical measurement or hydrogen exchange pulse labelling. By contrast, the study of partially folded states which are populated under equilibrium conditions can reveal information about the events which occur during protein folding if a link between their structure and that of kinetic intermediates can be made. Additionally, as research continues into the folding processes which occur in the cell *in vivo* (Freedman, 1992; Gething & Sambrook, 1992; Hendrick & Hartl, 1993), molten globules have been implicated as the structures which can bind to some molecular chaperones, and may also be significant in biological transport and degradation of proteins (Bychkova & Ptitsyn, 1993).

## 5.2 *Tools for studying molten globules.*

The characteristics which define a molten globule state may be identified by a variety of techniques, perhaps most clearly at a diagnostic level by circular dichroism. In the near-UV region, CD signals arise from the presence of fixed tertiary interactions involving aromatic amino acid side chains in a protein, whereas in the far-UV region, it is chromophores from the protein backbone in specific secondary structure which give rise to characteristic signals (Bolotina & Lugauskas, 1985; Woody, 1985; Bolotina, 1987; Johnson, 1990). A protein in its close packed native conformation has, in general, well defined secondary and tertiary structure, hence yielding characteristic signals in both near and far-UV. In a partially folded molten globule state, however, expansion of the structure and freeing of the aromatic side chains causes the signal in the near-UV region of the spectrum to be abolished or at least significantly reduced, whilst the persistence of secondary structure often leads to only slight differences in the far-UV CD between the molten globule and native states (Dolgikh *et al.*, 1981; Bolotina, 1987).

NMR is a highly sensitive probe of protein structure. Native structures characteristically yield NMR spectra which are well dispersed, due to fixed side chain interactions causing chemical shift perturbations as a result of ring currents induced by aromatic residues, and other effects such as polarity of hydrogen bonds (Bundi & Wüthrich, 1978; Wüthrich, 1986). In a 'classic' molten globule, these aromatic residues become mobile and the induced shifts resulting from them tend to be averaged, giving rise to generally featureless spectra with broad and unresolved resonances. This arises in part from conformational averaging at rates intermediate on the NMR time scale (Baum *et al.*, 1989), and in some cases also as a result of aggregation effects at the high concentrations required for NMR (Wormald, 1991). By contrast, fully unfolded proteins are more flexible and resonances are usually narrow with chemical shifts characteristic of the amino acids in an unstructured polypeptide chain (Wüthrich, 1986).

The techniques of CD and NMR are highly complementary since the former is sensitive to small global changes which may be difficult to observe in the NMR spectrum of a conformationally averaged molten globule state, whereas the latter can often be usefully employed to detect regions of local structure by perturbation of resonances if some of these resonances can be resolved. These techniques together can be used to give useful information about the structure and dynamic nature of molten globules and other partially folded states of proteins.

## 5.3 *Alpha -Lactalbumin*

Bovine α-lactalbumin is the most widely studied of proteins which form equilibrium molten globules (for primary amino acid sequence, see Appendix A). It consists of 123 residues, a calcium binding site comprising the four residues Asp82, Asp84, Asp87 and Asp88, the binding constant for calcium ($K_{Ca}$) being approximately $10^8$ $M^{-1}$ at pH 7.2, 25°C (Segawa & Sugai, 1983; Ikeguchi *et al.*, 1986b). The protein also has four disulphide bridge crosslinks: Cys6–Cys120, Cys28–Cys111, Cys61–Cys77 and Cys73–Cys91 (Hill & Brew, 1975). Previous studies have shown that the first of these disulphides is highly reactive (Iyer & Klee, 1973) and can be selectively cleaved under mildly reducing conditions and the free thiols chemically modified to yield a three-disulphide derivative. This three-disulphide protein ($3SS_{cam}$) has been shown by CD to be native-like in the presence of the calcium ligand (Kuwajima *et al.*, 1990), and to adopt a molten globule conformation in the *apo* form or at low pH (Ewbank & Creighton, 1991). If the reduced thiols are not modified, intramolecular disulphide rearrangement can occur in the molten globule state, with apparently no specific disulphide pairing preferred, indicating the non-specific nature of any tertiary interactions present in these partially folded states (Ewbank & Creighton, 1991). More recent results, however, suggest that native disulphide pairings are preferred, certainly in the later stages of folding (Ewbank & Creighton, 1993a), as has also been demonstrated recently for an analogue of the α-domain of α-lactalbumin (Peng & Kim, 1994) and for the intact protein (Peng *et al.*, 1995).

In this chapter, studies by $^1H$ NMR are described of the three disulphide derivative of bovine α-lactalbumin ($3SS_{cam}$) and, further, a two-disulphide species in which both Cys6–Cys120 and Cys28–Cys111 are reduced and blocked by carboxyamidomethylation. The structures and stabilities of folded and partially folded states of these proteins have been investigated in order to determine the contribution of disulphide bridges to the various structural states of α-lactalbumin.

## 5.4 *Materials and Methods*

Bovine α-lactalbumin (Type I) was obtained from the Sigma Chemical Co. Ltd., desalted by extensive dialysis at pH 2.0, lyophilised and stored at –20°C. Samples of a three-disulphide derivative of bovine α-lactalbumin ($3SS_{cam}$), in which the Cys6–Cys120 disulphide bridge has been selectively reduced and free thiols carboxyamidomethylated by treatment with iodoacetamide, and of a two-disulphide derivative in which Cys6–Cys120 and Cys28–Cys111 disulphide were similarly treated ($2SS_{cam}$), were prepared as described in Ewbank & Creighton (1993b), and were a generous gift from J.J. Ewbank (EMBL, Heidelberg). Samples of the *apo* protein were prepared by dissolving the lyophilised protein in 0.1 mM HCl and adjusting the pH to 2.0 as necessary with small amounts of HCl or NaOH. The high pH forms of the proteins were prepared by dissolving in 0.1M $NH_3$ solution and lowering the pH with HCl. This avoids any irreversible precipitation on passing the protein through its isoelectric point at ~pH 4 (McKenzie & White, 1991). Addition of $CaCl_2$ to a concentration of 1 mM was sufficient to ensure formation of the *holo* protein where required.

Samples for NMR were 2 mM for unmodified α-lactalbumin and $3SS_{cam}$; those of $2SS_{cam}$ were less concentrated (<1 mM) due to the low solubility of the protein. $^1H$ NMR experiments were performed as described in Chapter 2.

## 5.5 The structure of intact bovine α-lactalbumin.

α-Lactalbumin in its native state has been extensively studied by NMR (Bradbury & Norton, 1975; Koga & Berliner, 1985; Harushima & Sugai, 1989; Wormald, 1991). More recently, $^1H$ NMR assignments of the resonances of aromatic residues have been made at pH 10.5 by comparison of variants from different species (Alexandrescu *et al.*, 1992). The protein has an isoelectric point around pH 3.5, at which point aggregation occurs to a significant extent even at low concentrations (Wormald, 1991) . Some aggregation of $3SS_{cam}$ is still observed up to pH 5.0, possibly as a result of exposure of previously buried charged side chains on breaking of the Cys6–Cys120 disulphide bridge or some other conformational change. For this reason, pH 6.0 was chosen for the following studies, since at this pH, minimal self-association of the protein was apparent as judged by the quality of 1D NMR spectra.

The 1D $^1$H-NMR spectrum of *holo* bovine α-lactalbumin, pH 6.0, 30°C in 90% $H_2O$: 10% $D_2O$ is shown in Figure 5.1a. The spectrum is characteristic of that expected for a tightly folded native protein, with dispersed resonances from the backbone amides and aromatic side chains in the downfield region ($\delta$ 6–11 ppm), downfield shifted αCH resonances characteristic of β-sheet structure, and upfield shifted side chain methyl resonances arising from aliphatic residues ring current shifted from their random coil shifts presumably as a result of burial in a structured hydrophobic core. By contrast, Figure 5.1b shows the spectrum of α-lactalbumin denatured at 70°C. In this spectrum the dispersion is almost completely abolished due to the mobility of side chains and the extended nature of the unfolded polypeptide chain.

The third state available to this protein in solution is the molten globule, the spectrum of which at pH 2.0, 30°C is shown in Figure 5.1c. This state is intermediate in structure between the highly structured native state and the fully unfolded state. Although dispersion is minimal, there is some evidence for non-random behaviour (Baum *et al.*, 1989), and the line widths of resonances are significantly broader than in either the folded or unfolded states, signifying an ensemble of structural states in intermediate exchange on the NMR time scale.

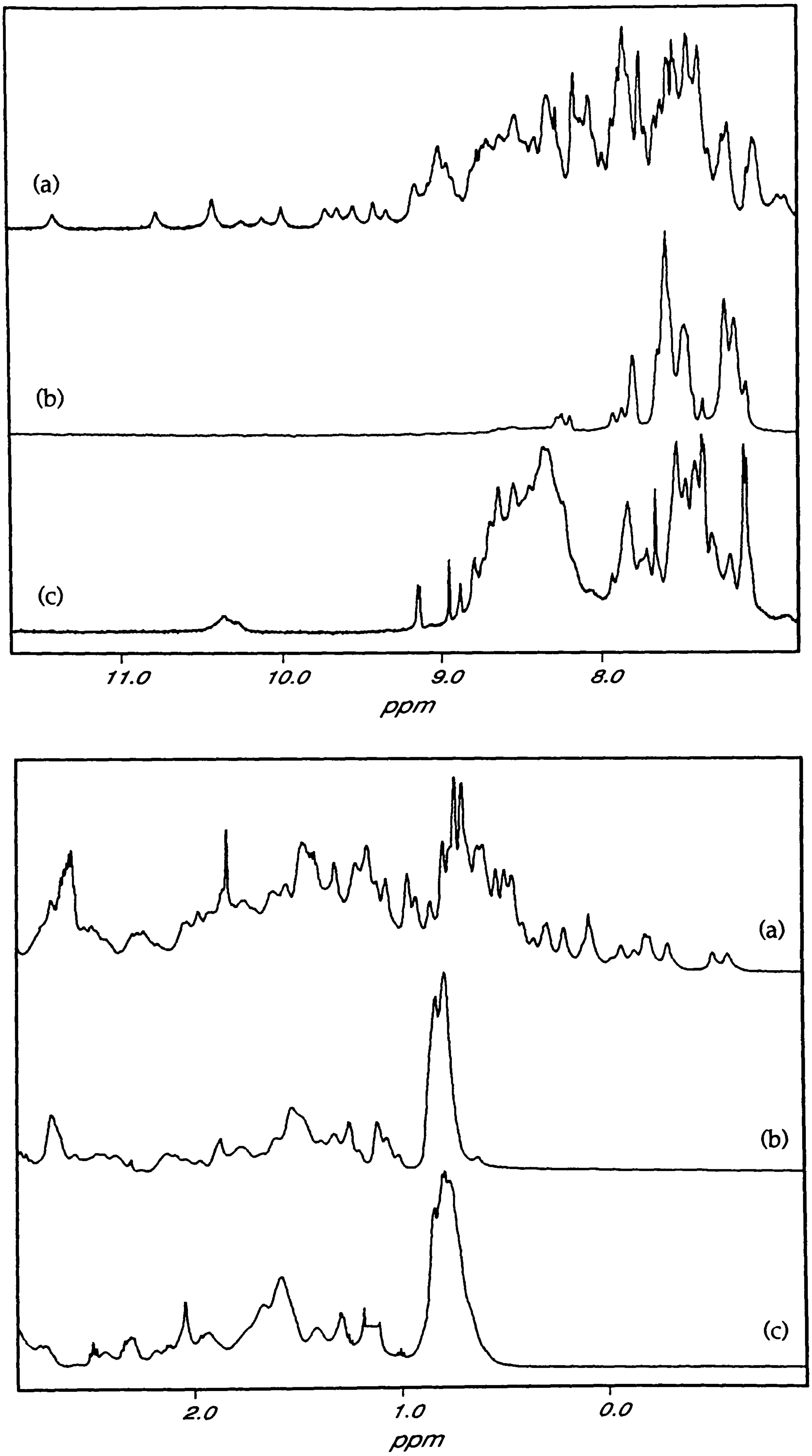


**Figure 5.1.** 1D $^1$H NMR spectra of bovine $\alpha$-lactalbumin at (a) pH 6.0, 30°C, (b) pH 6.0, 70°C, (c) pH 2.0, 30°C.

Circular dichrographs of these three structural states of α-lactalbumin are shown in Figure 5.2 [reproduced from Ewbank & Creighton (1993b)] The native *holo* form (labelled 6) shows characteristic double negative maxima in the far-UV CD at 208 and 222 nm, indicating the presence of a predominantly α-helical structure. In the near-UV region, the spectrum has a complex line shape caused by interaction of the many aromatic residues of the protein in a fixed conformation. The unfolded state in 8M urea shows no signal in the near-UV region, indicating the absence of fixed tertiary interactions, while in the far-UV, the band at 222 nm is absent, suggesting an unfolded structure in 8M urea. The strongly absorptive properties of urea at wavelengths below 215 nm prevent the observation of the maximum at 205 nm characteristic of an unfolded polypeptide (Woody, 1985) . The molten globule state at pH 2.0 has no near-UV CD signal, but the signal in the far-UV region is only slightly diminished relative to that of the native state, suggesting a native-like complement of secondary structure, but the complete absence of any fixed tertiary interactions involving aromatic residues.

We can now use NMR and CD spectra as a basis for comparison with the modified forms of α-lactalbumin in which disulphides have been selectively removed, in order to assess the effect of these disulphides on the stability and structural integrity of the native and molten globule states of the protein.

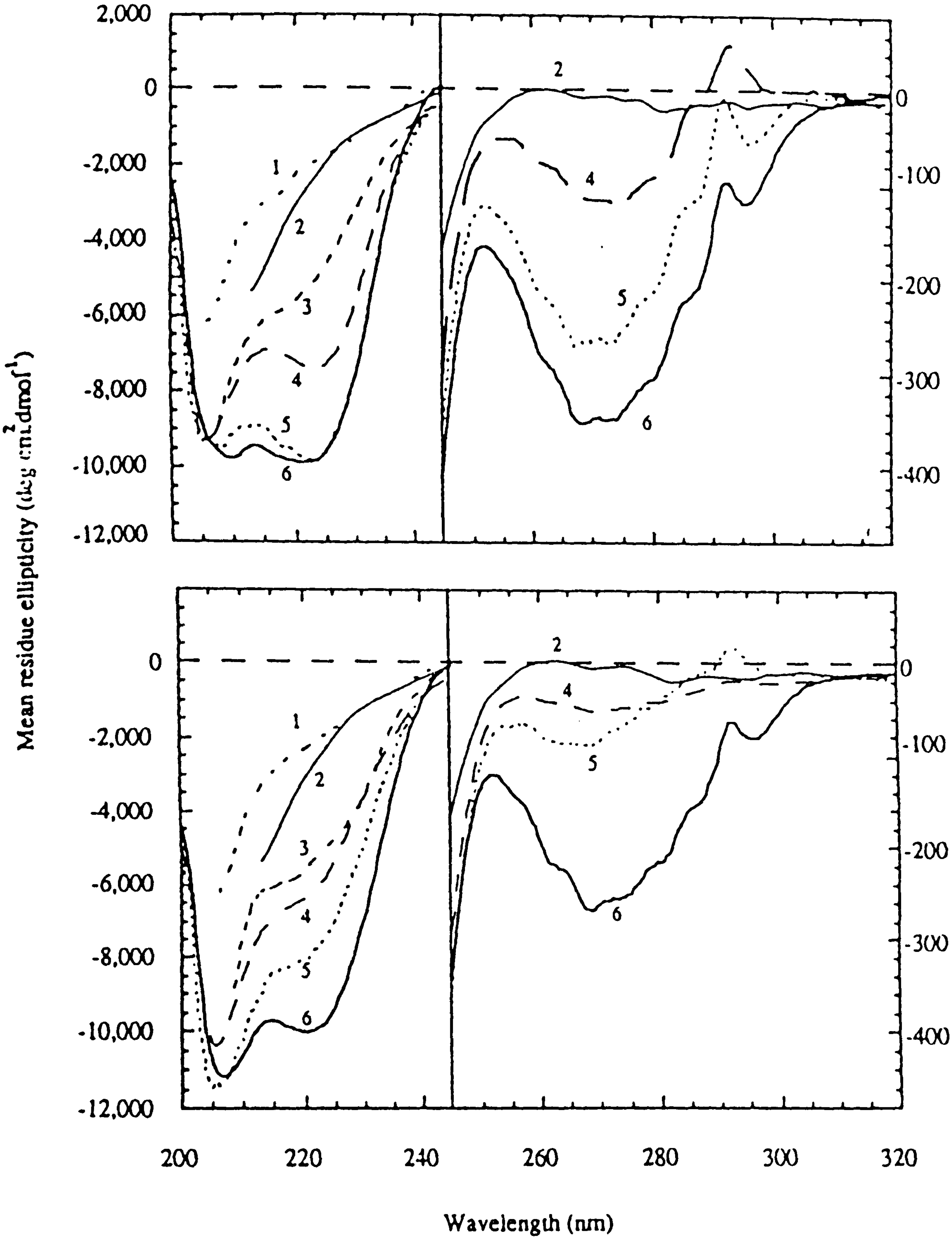


Figure 5.2. Far- and near-UV CD spectra of α-lactalbumin and its trapped derivatives in the presence and absence of $Ca^{2+}$ (upper and lower panels respectively). BLA (——,6), $3SS_{cam}$ (······,5), $2SS_{cam}$ (– –,4), fully reduced BLA with all free thiols carboxyamidomethylated (- - -,3), fully reduced BLA with all free thiols carboxymethylated (· · · ·,2) and BLA in 8M urea (thin line,1). All samples were buffered in 0.1M Tris (pH 8.7) 0.2M KCl and either 10 mM $CaCl_2$ or 1 mM EDTA, at 25°C. Reproduced from Ewbank & Creighton (1993b).

## 5.6 *The native state of holo $3SS_{cam}$*

Figure 5.3 shows the upfield and downfield regions of the $^1H$ NMR spectrum of *holo* $3SS_{cam}$ at pH 6.0, 20°C, compared with the unmodified protein under identical conditions. The dispersion in both regions is indicative of a highly structured folded protein. In the regions shown, where peaks are well resolved, only small differences in chemical shift are observed between the spectra of the two proteins, suggesting that the overall structure of the protein has not been significantly perturbed by removal of the Cys6–Cys120 disulphide bridge. In the downfield region, only one peak at 10.8 ppm in the unmodified protein is significantly perturbed in the three-disulphide protein and can no longer be resolved. This resonance has been assigned to Trp118 N1H (Alexandrescu *et al.*, 1992) and is close in the primary sequence to the site of modification. Small changes can also be observed in the pattern of upfield shifted methyl resonances, which presumably arise from small perturbations of the aromatic packing arrangement in $3SS_{cam}$ relative to wild-type α-lactalbumin. These chemical shift differences are all small, however ($\Delta\delta \leq 0.5$ ppm), suggesting only small perturbations in the environments of these methyl groups. The magnitude of these changes is similar to those observed following removal of the corresponding disulphide bridge (Cys6–Cys127) from hen lysozyme (Chapter 3). Again only small perturbations are observed in the NMR spectrum of $CM^{6,127}$-lysozyme compared with the unmodified protein, indicating again that only local structure is perturbed. In the case of $CM^{6,127}$-lysozyme, only a short region of the polypeptide chain close to the C- terminus is altered in

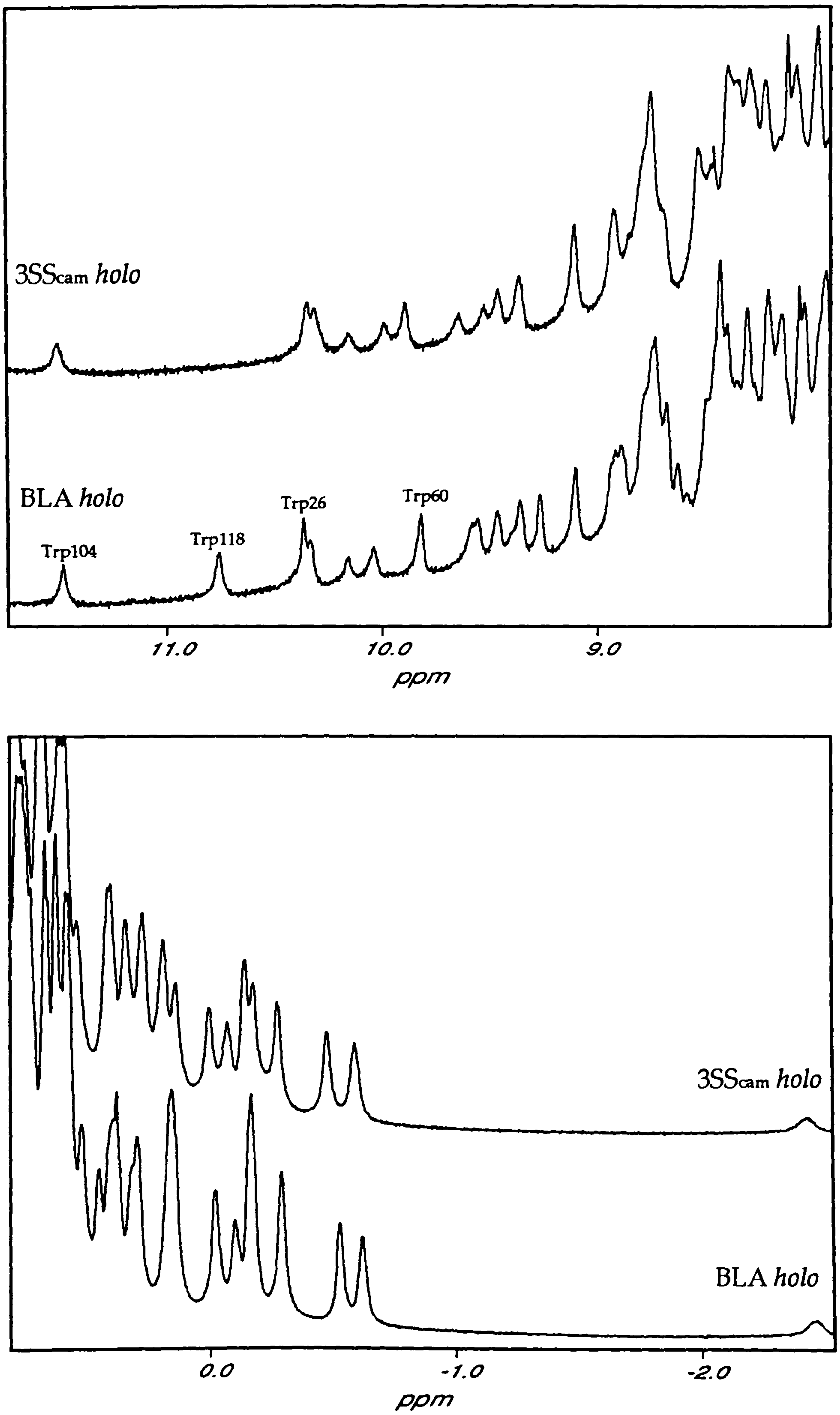


**Figure 5.3.** Comparison of the native *holo* states of intact bovine α-lactalbumin (BLA) and $3SS_{cam}$, pH 6, 30°C, $H_2O$. Upper panel: downfield region containing resonances from tryptophan indole protons, amides and aromatic side chains. Lower panel: upfield region including ring current shifted methyl side chain resonances.

conformation from that of the unmodified protein ((Radford *et al.*, 1991; Hill *et al.*, 1993) and see Chapter 3).

The aromatic side chain resonances of intact bovine α-lactalbumin have been completely assigned by comparison of the NMR spectra of bovine, human and guinea-pig α-lactalbumins, which each contain a different subset of aromatic residues (Alexandrescu *et al.*, 1992). The region of the COSY spectrum corresponding to resonances involving aromatic ring protons of bovine α-lactalbumin is compared with that of $3SS_{cam}$ in Figure 5.4. Many of the cross peaks occur at identical chemical shifts, again indicating that the majority of the structure of aromatic side chains is identical in the two proteins. Although a complete set of cross peaks for all of the aromatic residues was not resolved under the experimental conditions employed (not all of the 20 possible cross peaks arising from aromatic side chain hydrogens could be distinguished at 30°C, pH 6.0), it was possible to assign those peaks which could be resolved by direct comparison with spectra of the intact protein. Complete spin systems were observed for two of the four tryptophan residues: these have been assigned to Trp60 and Trp118. Partial assignment of Trp26 and Trp104 was also possible since these cross peaks occurred at very similar chemical shifts to corresponding resonances from hydrogens in the intact protein. An additional cross peak was observed in the spectrum of $3SS_{cam}$ ($\delta$ 7.31, 7.61 ppm). It is tempting to tentatively assign this resonance to Phe9, which is unresolvable from the diagonal in the unmodified protein ($\delta$ 7.32, 7.40 ppm), since this residue is very close to the site of modification. Spin system assignments of those aromatics which could be resolved are tabulated and compared with unmodified bovine α-lactalbumin in Table 5.1.

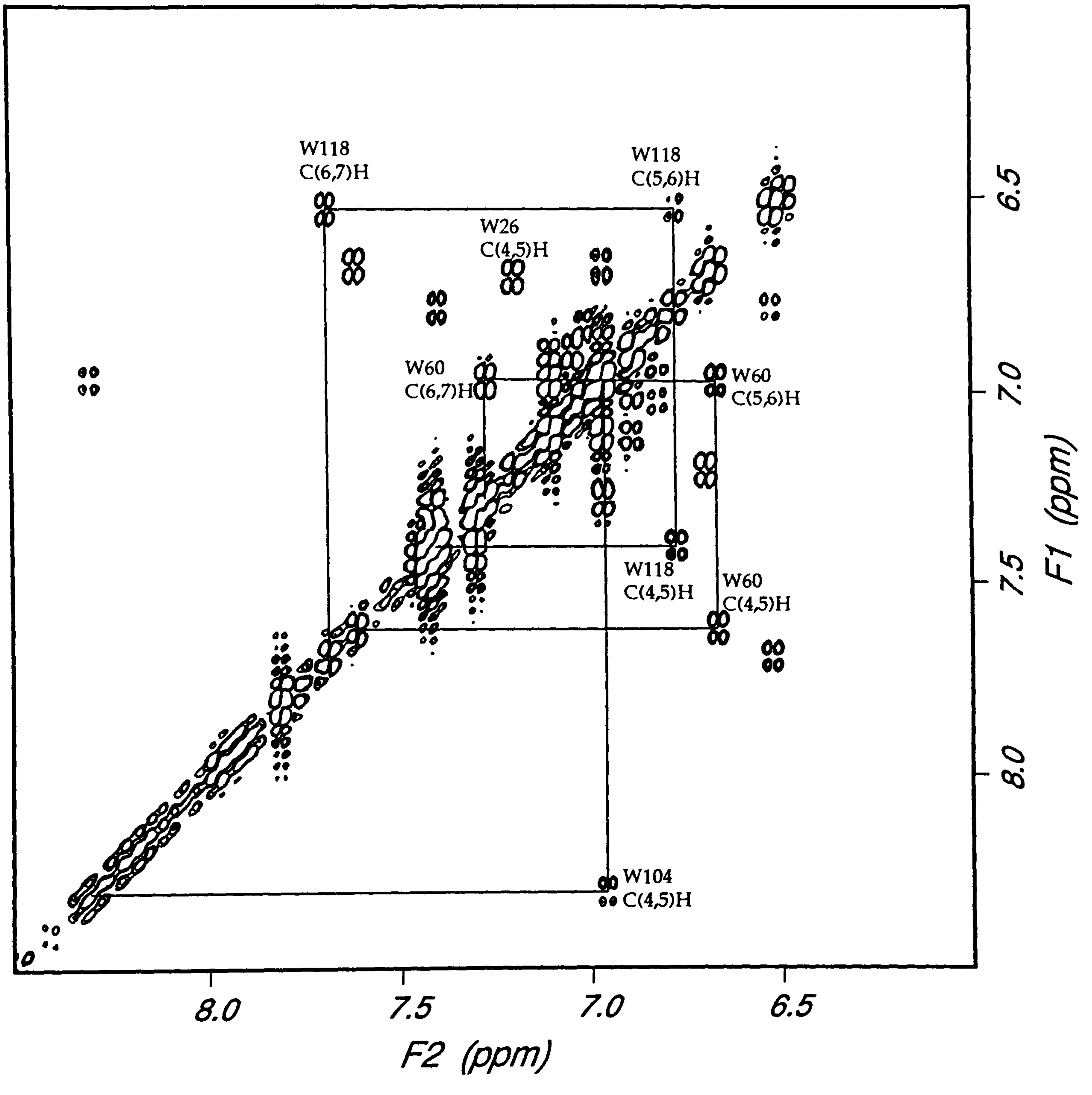


**Figure 5.4a.** Aromatic region of the DQF-COSY spectrum of bovine α-lactalbumin pH 6.0, 30°C, showing assignments for residues Trp26, Trp60, Trp104, and Trp118. Assignments taken from Alexandrescu *et al.* (1993).

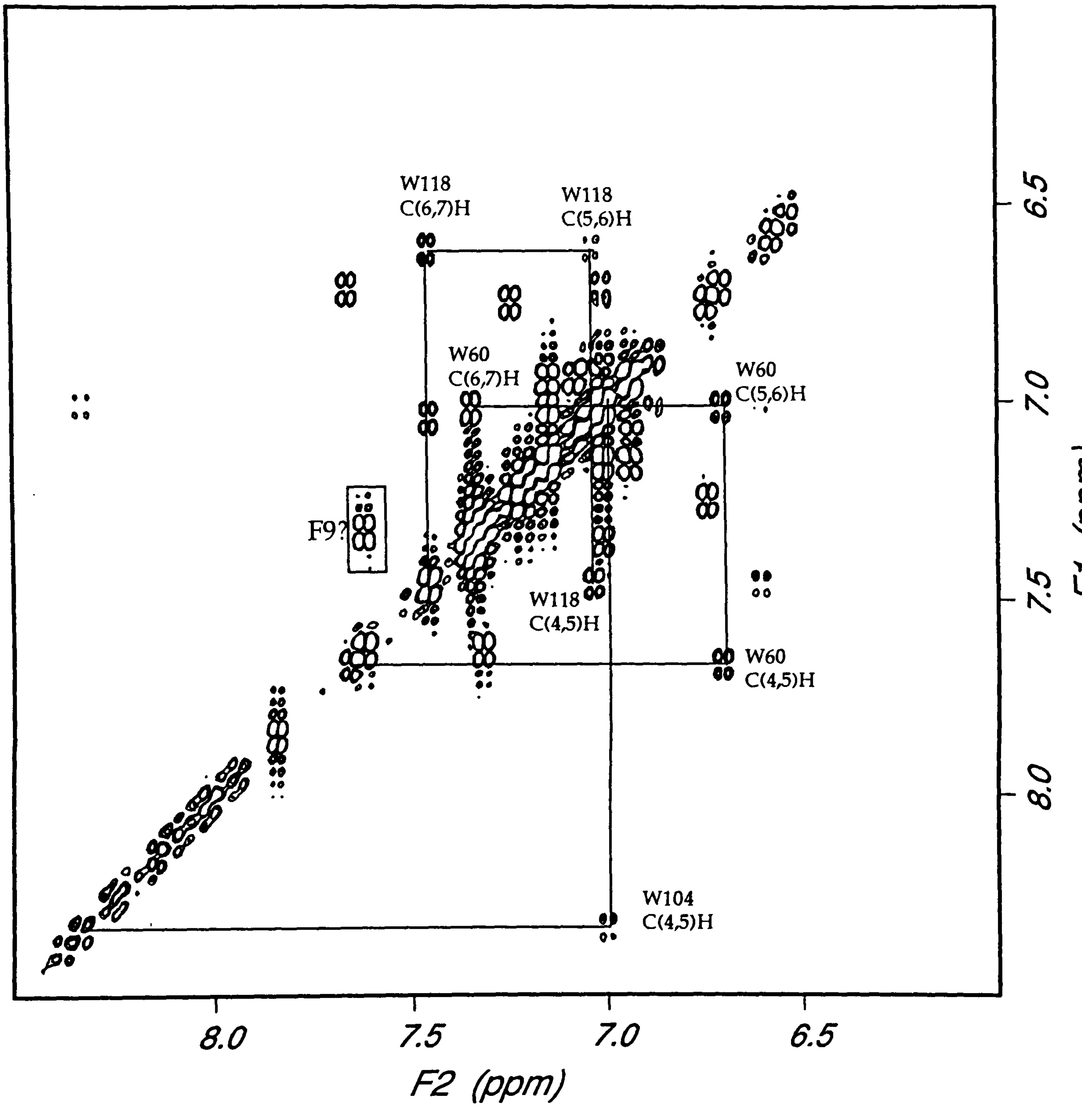


**Figure 5.4b**. Aromatic region of the DQF-COSY spectrum of 3SScam. Assignments of those cross peaks which are shifted significantly from the corresponding resonances in the wild type protein are indicated (Phe9, Trp60 and Trp118).

| BLA<br>Residue | Chemical shift of proton<br>C2H | <br>N1H | <br>C4H | <br>C5H | <br>C6H | <br>C7H | <br>C2,6H | <br>C3,5H |
|---|---|---|---|---|---|---|---|---|
| | ppm | | | | | | | |
| Trp26 | 6.59 | 10.41 | 7.26 | 5.70 | 5.54 | 6.97 | | |
| Trp60 | 7.43 | 9.86 | 7.64 | 6.71 | 7.02 | 7.32 | | |
| Trp104 | 7.49 | 11.41 | 8.26 | 6.98 | 6.85 | 7.00 | | |
| Trp118 | 7.19 | 10.73 | 7.42 | 6.80 | 6.50 | 7.63 | | |
| Phe9/80 | | | 7.47 | | | | 7.32 | 7.40 |
| Phe31 | | | 7.22 | | | | 5.92 | 7.19 |
| Phe53 | | | 7.06 | | | | 7.12 | 6.91 |
| Phe80/9 | | | 7.33 | | | | 7.23 | 7.27 |
| Tyr18 | | | | | | | 7.10 | 6.94 |
| Tyr36 | | | | | | | 6.95 | 7.17 |
| Tyr50 | | | | | | | 7.22 | 6.72 |
| Tyr103 | | | | | | | 5.46 | 6.55 |

| $3SS_{cam}$<br>Residue | Chemical shift of proton<br>C2H | <br>N1H | <br>C4H | <br>C5H | <br>C6H | <br>C7H | <br>C2,6H | <br>C3,5H |
|---|---|---|---|---|---|---|---|---|
| | ppm | | | | | | | |
| Trp26 | | | 7.31 | 5.72 | 5.39 | 6.93 | | |
| Trp60 | | | 7.67 | 6.71 | 7.01 | 7.35 | | |
| Trp104 | | | 8.33 | 7.00 | 6.87 | 7.04 | | |
| Trp118 | | | 7.45 | 7.04 | 6.61 | 7.45 | | |
| Phe9/80 | | | | | | | 7.32 | 7.62 |
| Phe31 | | | | | | | 5.91 | 7.19 |
| Phe53 | | | | | | | 7.07 | 6.93 |
| Phe80/9 | | | | | | | 7.23 | 7.27 |
| Tyr18 | | | | | | | 7.15 | 6.93 |
| Tyr36 | | | | | | | 6.94 | 7.16 |
| Tyr50 | | | | | | | 7.26 | 6.73 |
| Tyr103 | | | | | | | 5.49 | 6.56 |

**Table 5.1** (a) $^1$H NMR assignments and chemical shift values for aromatic ring protons in bovine α-lactalbumin, pH 10.5, 37°C, from Alexandrescu *et al.* (1992). (b) Corresponding assignments for $3SS_{cam}$, pH 6.0, 20°C.

The resonances corresponding to Phe9 and Trp118 are the only ones which are significantly perturbed in the spectrum of the three-disulphide derivative, and these are very close in the primary sequence to Cys6 and Cys120, the residues involved in the disulphide bond which has been reduced and carboxyamidomethylated. This suggests that any structural change induced in the native state of α-lactalbumin by removal of this disulphide bond is primarily a local one; the majority of the protein structure in the hydrophobic box region appears to remain intact and essentially native-like, as judged by few shifts in aromatics and aliphatic side chain methyl resonances, and appears unaffected by the absence of the Cys6–Cys120 disulphide bridge.

The fingerprint regions (NH-αCH correlations) of the COSY spectra of intact α-lactalbumin and $3SS_{cam}$ are shown in Figure 5.5. The spectrum of the four-disulphide protein contains far fewer peaks than one would expect from the amino acid sequence. Of 126 possible NH-αCH cross peaks (two cross peaks should be observed for each glycine residue (Wüthrich, 1986)), only 79 are observed in the spectrum of bovine α-lactalbumin under these conditions, and only 71 in the spectrum of $3SS_{cam}$. The small number of cross peaks may in part result from the relatively high pH required to maintain the protein in its native state without aggregation, at which the intrinsic rate of hydrogen exchange is much faster and may give rise to significant line broadening. Also, line broadening due to an equilibrium between the native state and partially folded or unfolded states is likely to reduce significantly the number of observable cross peaks. Investigation of a variety of different conditions of pH and temperature, however, did not enable any further cross peaks to be identified, in accord with data for the guinea pig protein (Wormald, 1991).

The dispersion of resonances in the fingerprint region of the spectrum of $3SS_{cam}$ is again indicative of a high degree of structure, with some αCH chemical shifts as low as 2.8 ppm [random coil αCH chemical shifts are in the range 4.2–4.8 ppm (Wüthrich, 1986; Wishart *et al.*, 1995)]. Many of the cross peaks clearly occur at closely similar chemical shifts in intact bovine α-lactalbumin and in $3SS_{cam}$, and others appear to have only small shifts induced by the modification. This again is strong evidence that the overall tertiary fold of the protein is not significantly perturbed by the removal of the Cys6–Cys120 disulphide bond, and that the peptide backbone retains the same overall topology. This is very similar to the case for $CM^{6,127}$-lysozyme, in which removal of the Cys6–Cys127 disulphide bridge only affects structure of the protein local to the site of modification, leaving the majority of the native structure intact (Radford *et al.*, 1991).

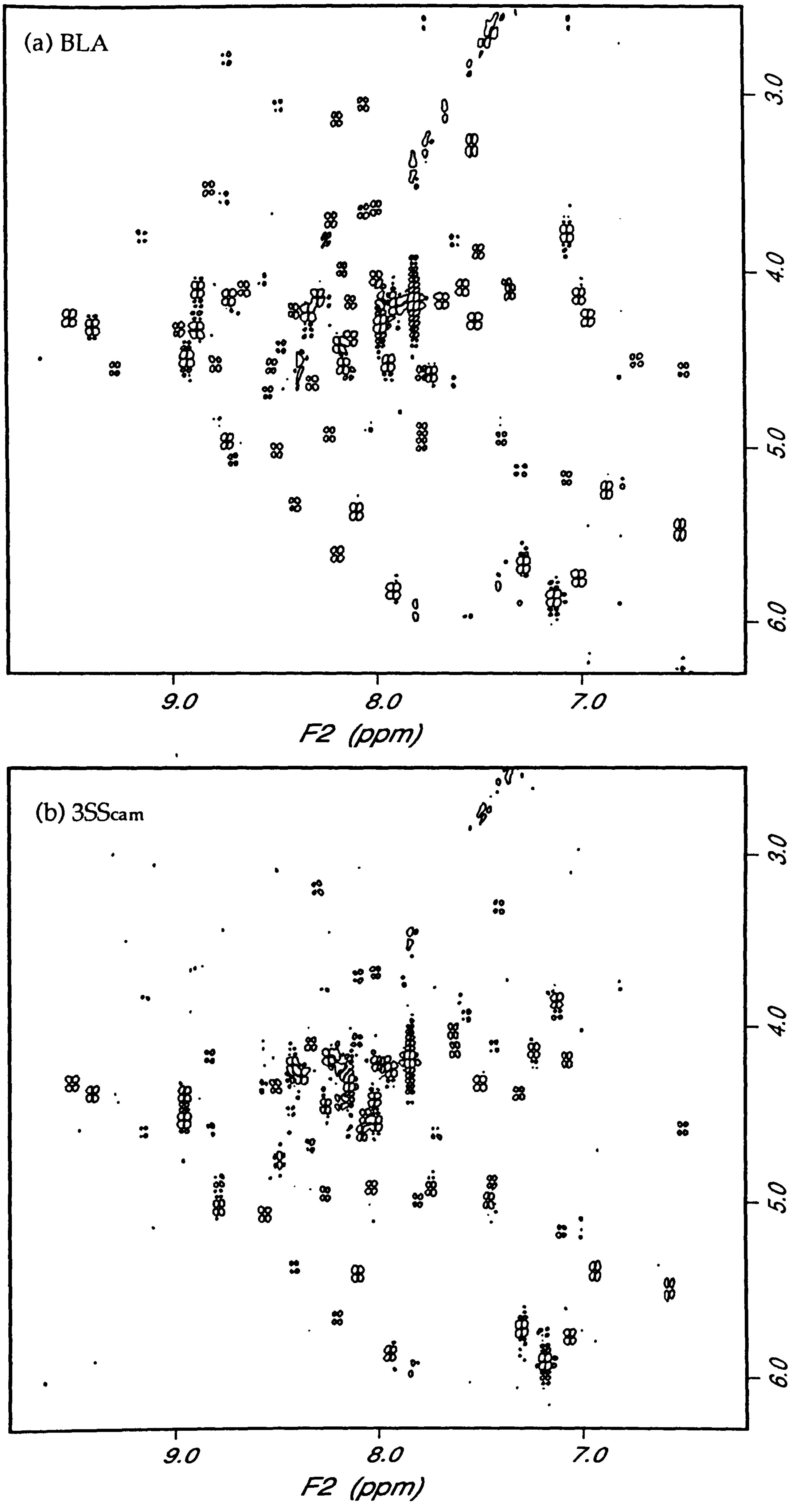


**Figure 5.5**. Comparison of the NH-αCH "fingerprint" region of the NMR spectrum of (a) bovine α-lactalbumin, pH 6.0, 30°C, $H_2O$ and (b) 3SScam, pH 6.0, 30°C, $H_2O$.

Sequential assignment of the NH-αCH resonances of $3SS_{cam}$ was not attempted from these spectra since an insufficient number of cross peaks were observed to enable its success; indeed it has already been concluded that assignment of the $^{1}H$ NMR spectrum even of intact bovine α-lactalbumin is not possible using homonuclear NMR methods alone (Wormald, 1991). Further work on assignment of human and bovine α-lactalbumins isotopically enriched with $^{15}N$ by heteronuclear methods is currently underway in this laboratory (C. Redfield, J. Balbach & C.M. Dobson, unpublished results). This should at some point in the future enable a more detailed characterisation of the structural changes occurring in the three-disulphide derivative.

## 5.7 *Temperature unfolding of $3SS_{cam}$*

In order to compare the thermal denaturation temperatures of unmodified bovine α-lactalbumin and $3SS_{cam}$ a temperature study was carried out at pH 6.0. Figure 5.6 shows representative $^{1}H$ NMR spectra of $3SS_{cam}$ acquired over the temperature range 20° to 55°C. Up to 45°C only small changes in the spectrum are observed, but above this temperature a dramatic change occurs in the appearance of the spectra, characteristic of unfolding of the protein. All of the upfield shifted methyl resonances disappear and downfield shifted αCH peaks collapse towards their approximate random coil values. The resonances from aromatic ring protons broaden to yield an unresolved envelope, but remain dispersed to some extent. This is thought to be due to exchange between an ensemble of partially structured states at a rate intermediate on the NMR time scale (Baum *et al.*, 1989). These features are all characteristic of the formation of a partially folded molten globule state.

Above this temperature, resonances begin to sharpen again as the protein unfolds completely, but this is accompanied by irreversible aggregation of $3SS_{cam}$, so spectra were not acquired above 55°C. Similar behaviour is observed for the unmodified protein, but again severe aggregation occurs above 70°C, so an accurate $T_m$ could not be measured for this protein.

The $T_m$ of $3SS_{cam}$ for unfolding of the native state to the molten globule was measured from this temperature study to be ~47°C at pH 6.0, by monitoring the disappearance of upfield shifted methyl resonances. The $T_m$ of $3SS_{cam}$ is, therefore, some 20°C lower than that of unmodified α-lactalbumin, consistent with a diminished thermal stability resulting from the removal of the covalent Cys6–Cys120 bond. Since the unfolding transition of α-lactalbumin is not a simple two-state mechanism, it did not prove possible to find conditions under which exchange between any two states was slow on the NMR time scale. This prevented the possibility of conducting magnetisation transfer experiments in order to extract any enthalpies of unfolding (Dobson & Evans, 1984). Similarly no resonances in the partially folded or unfolded states were sufficiently resolved to allow an equilibrium thermodynamic analysis. Studies of these species by differential scanning calorimetry are currently underway, however (A. Cooper, J.J. Ewbank & T.E. Creighton, personal communication), in which the effect on stability over a full range of pH and calcium concentration are being investigated. These should give insight into the factors which stabilise these species.

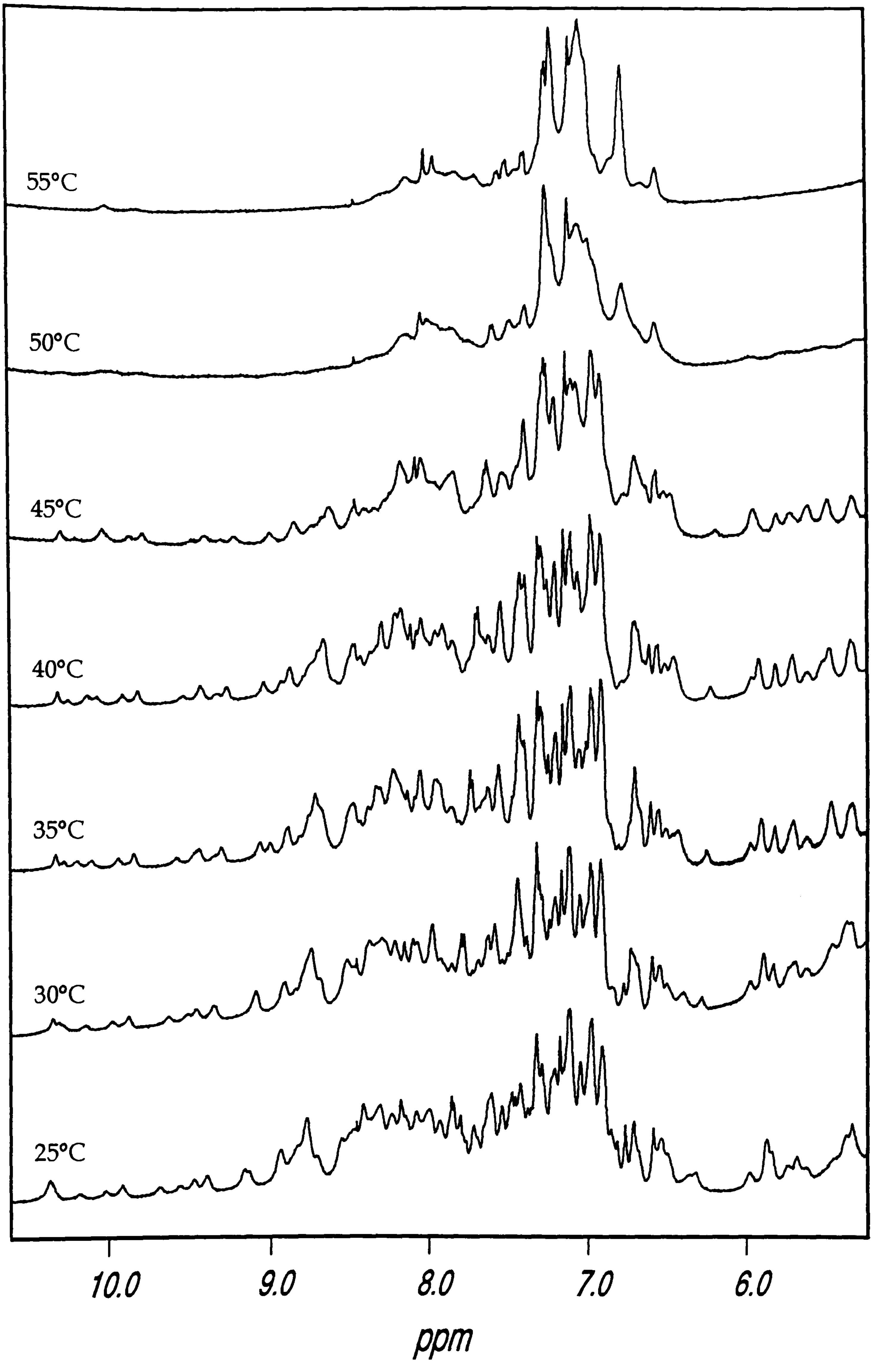


**Figure 5.6a**. Temperature denaturation of *holo* $3SS_{cam}$, pH 6.0, $H_2O$, monitored by NMR in the downfield region of the spectrum (including resonances from NH protons and aromatic side chains).

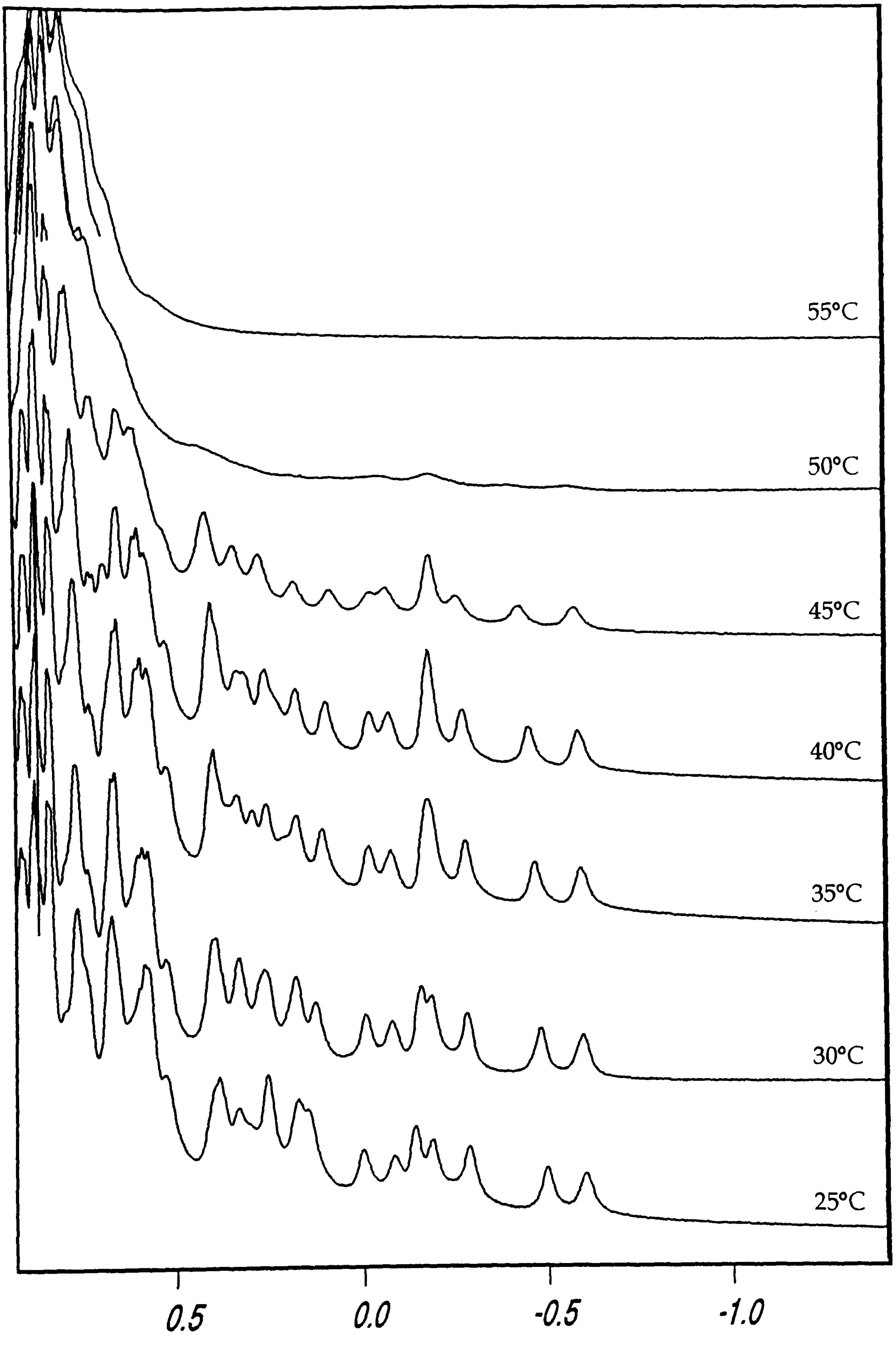


**Figure 5.6b**. Temperature denaturation of *holo* $3SS_{cam}$, pH 6.0, $H_2O$, monitored by NMR in the upfield region of the spectrum (including resonances from ring current shifted methyl side chains).

It is interesting at this point to summarise the results obtained by Ikeguchi *et al.* (1992) who investigated the unfolding transition of $3SS_{cam}$ denatured by guanidinium chloride. These authors were able to measure thermodynamic values for unfolding transitions to the molten globule state and to the unfolded state. Their conclusion was that the contribution of the Cys6–Cys120 disulphide bond to the stability of the native state relative to the unfolded state was primarily to reduce the configurational entropy of the unfolded state. This is in accord with our calorimetric studies on the homologous three-disulphide derivative of hen lysozyme (Cooper *et al.*, 1992) and Chapter 3.

Regarding the contribution of this disulphide to the stability of the molten globule state, Ikeguchi *et al.* (1992) concluded that secondary structure was indeed stabilised by the presence of this covalent bond, again mostly resulting from a reduction of entropy in the unfolded state. The increased configurational freedom on removal of the Cys6–Cys120 disulphide bridge is sufficient to enable shuffling of disulphides in the molten globule state of the unblocked three-disulphide protein (Ewbank & Creighton, 1991), which initially was thought to be completely random. More recent results, however, have suggested that there is more restriction to shuffling under different conditions, certainly in the case of the isolated α-domain fragment of α-lactalbumin (Peng & Kim, 1994) and indeed the three-disulphide protein in the presence of the chaperone GroEL (Hayer-Hartl *et al.*, 1994). The increased stability conferred on the protein by the disulphide Cys6–Cys120 may be primarily a local effect, affecting mainly the α-domain of the protein. Indeed Peng & Kim (1994) have suggested recently, in accord with earlier work on the molten globule state of α-lactalbumin (Baum *et al.*, 1989), that the β-domain may not be structured in the molten globule state. Indeed there are no residues in the β-domain which are protective against hydrogen exchange in the molten globule state of guinea pig α-lactalbumin (Chyan *et al.*, 1993). These results are consistent with the kinetic refolding pathway of hen lysozyme, in which the transiently formed α-domain intermediate has been postulated to resemble a molten globule state (Dobson *et al.*, 1994). In this case the α-domain of the protein forms stable structure protective against hydrogen exchange, whereas no protection is seen for residues in the β-domain. On reduction of the Cys6–Cys127 disulphide bridge, the α-domain kinetic intermediate can no longer be detected on the folding pathway of the three-disulphide species, possibly due to intrinsic lack of stability of the α-domain in the absence of this disulphide bridge (Eyles *et al.*, 1994) and Chapter 4.

## *5.8 NMR spectra of apo $3SS_{cam}$*

At low pH, or in the absence of the calcium ligand, unmodified α-lactalbumin partially unfolds to a molten globule state (Ikeguchi *et al.*, 1986b). Similar effects have been observed in the $3SS_{cam}$ protein by CD (Ewbank & Creighton, 1991; Ewbank & Creighton, 1993b), as shown in Figure 5.2, with loss of signal in the near-UV region but only a slight reduction in signal intensity in the far-UV. The NMR spectra of both intact α-lactalbumin and $3SS_{cam}$ in the absence of calcium become broad and featureless due to exchange between conformational states being in the intermediate regime on the NMR time scale. All of this evidence points to the formation of a molten globule ensemble of conformations on removal of the calcium ligand.

The NMR spectra of *apo* and *holo* $3SS_{cam}$ are compared in Figure 5.7. At 30°C and pH 6.0, the spectrum of *apo* $3SS_{cam}$ has the broad envelopes of resonances characteristic of a molten globule state. Several highly dispersed resonances may be seen at low intensity in the spectrum shown, but this may be attributed to leaching of $Ca^{2+}$ ions from the walls of the NMR tube as addition of small amounts of EDTA abolishes these resonances. Quality of spectra was seriously reduced, however, by the addition of EDTA (spectra not shown). The protein appears somewhat more unfolded than the molten globule state of intact α-lactalbumin, or at least some regions may be more dynamic, as indicated by the appearance of some sharp resolved peaks in the aromatic region of the spectrum (Figure 5.7). This is supported by the CD data above, in which the far-UV signal from *apo* $3SS_{cam}$ is diminished as compared to that from the *apo* state of unmodified α-lactalbumin (Figure 5.2). Nevertheless it is clear that removal of the Cys6–Cys120 disulphide bridge does not destabilise the protein to such a great extent as to prevent formation of the marginally stable molten globule species. This finding is supported by the evidence of Peng & Kim (1994) that a molten globule structure can form in the absence of the $Ca^{2+}$ binding site or indeed the entire β-domain of the protein.

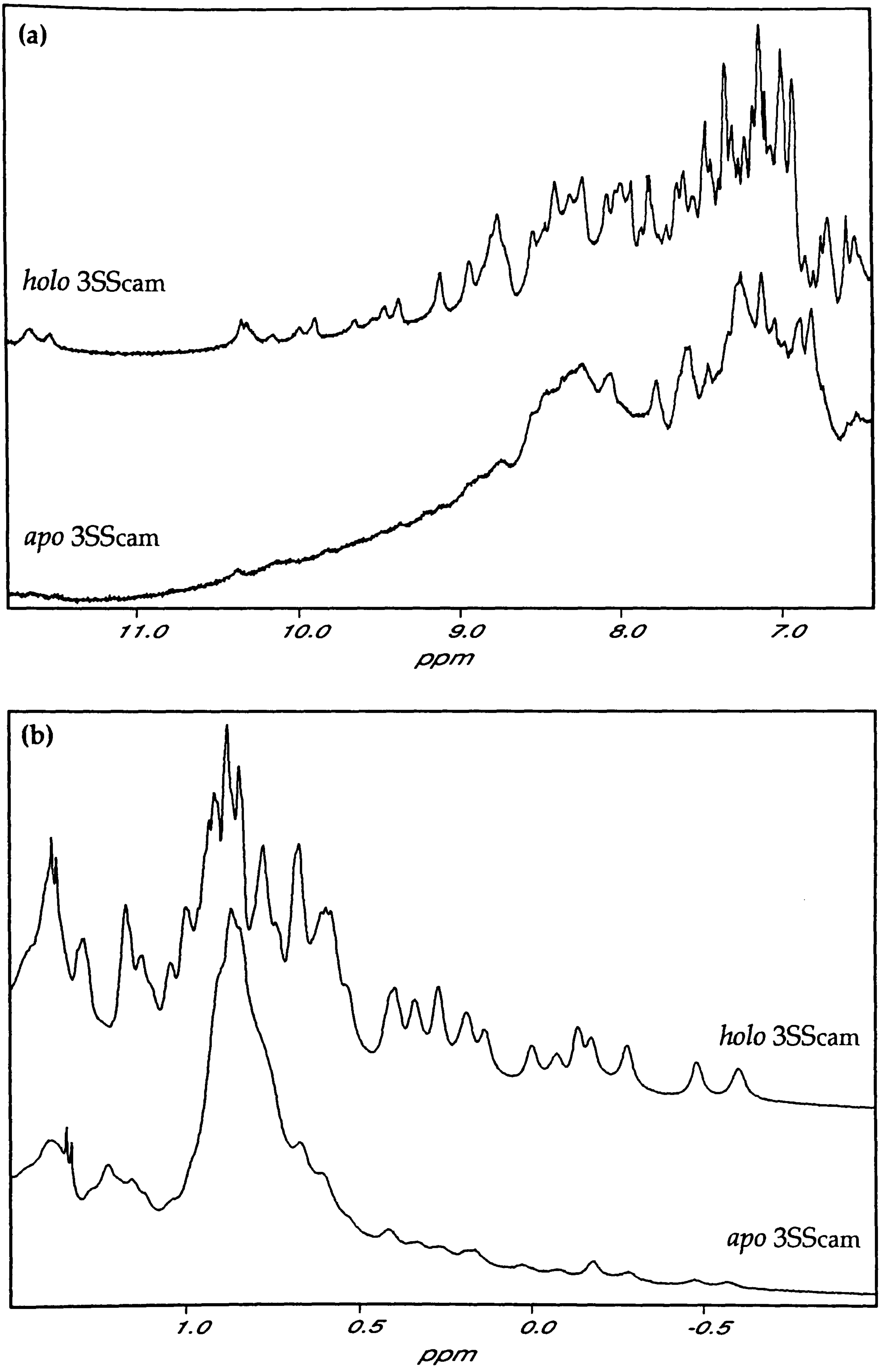


**Figure 5.7**. Comparison of *holo* and *apo* 3SScam, 30°C pH 6.0, 90% $H_2O$:10% $D_2O$. (a) Downfield and (b) upfield regions of the 1D NMR spectrum.

## 5.9 *Studies on 2SS$_{cam}$*

Removal of a second disulphide (Cys28–Cys111) from bovine α-lactalbumin results in a two-disulphide species which contains no disulphide crosslinks in the α-domain of the protein. This has a profound effect on the stability of the protein, presumably as a result of increased exposure of hydrophobic surface area to solvent on removal of the disulphide bridges. The protein has a high tendency towards aggregation in the range pH 3.0–8.0, hence it was necessary to acquire NMR spectra at very low protein concentration (0.3 mM, below which minimal association was detected as judged from the quality of spectra obtained). The $^1$H NMR spectrum of 2SS$_{\mathrm{cam}}$ in the presence of 1 mM Ca$^{2+}$, pH 8.7, is shown in Figure 5.8a. This spectrum shows very little dispersion of resonances from random coil chemical shifts, in stark contrast to spectra of the *holo* forms of intact α-lactalbumin or 3SS$_{\mathrm{cam}}$. The four- and three-disulphide forms of the protein adopt a tightly folded native structure in the presence of the calcium ligand, whereas this clearly is not the case for 2SS$_{\mathrm{cam}}$. There is some dispersion of resonances, suggestive of some structure, but this state is highly unstable and prone to precipitation, even at the low concentrations used.

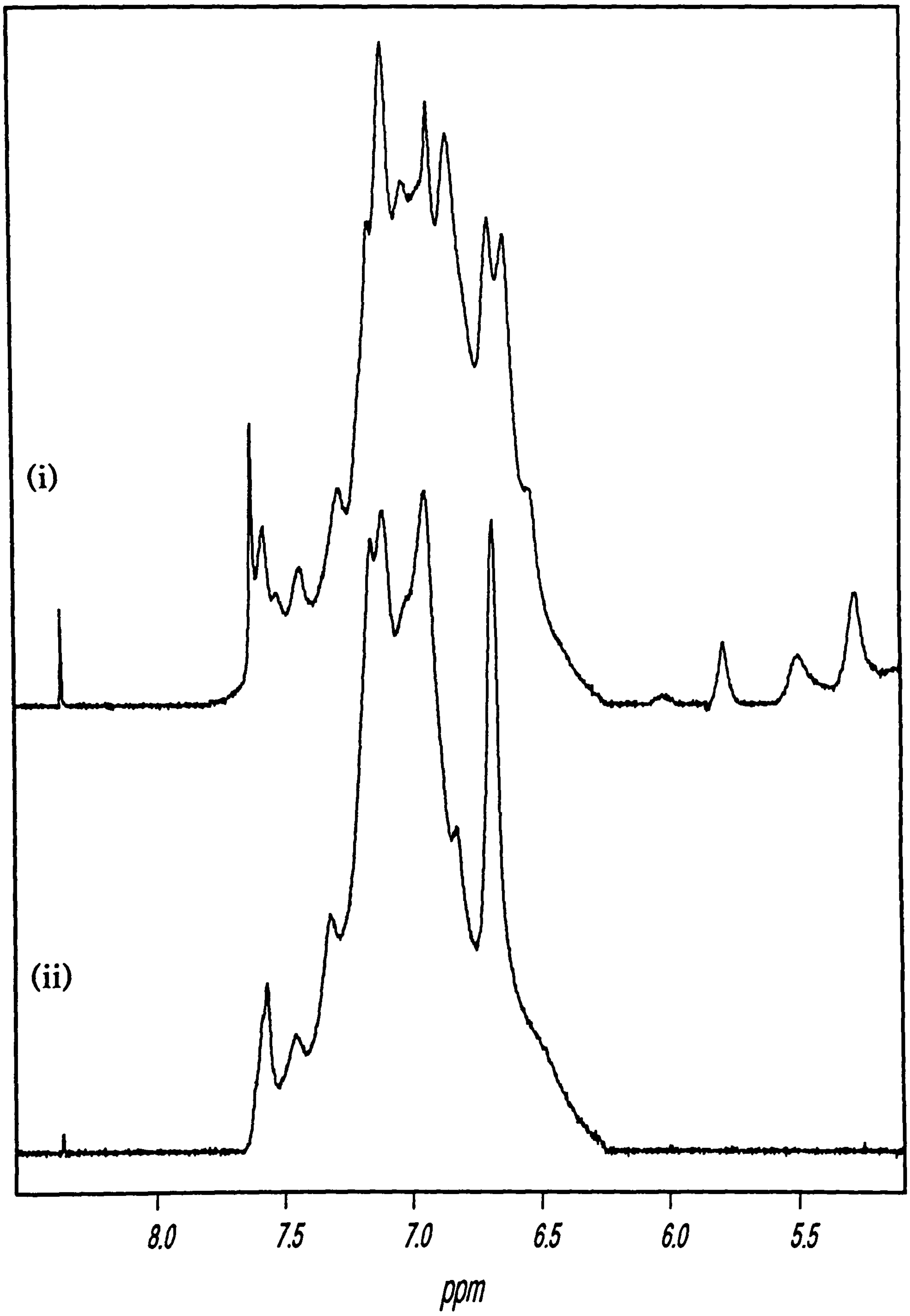


**Figure 5.8a.** Comparison of the NMR spectra of (i) *holo* 2SScam, pH 8.7, 20°C, and (ii) *apo* 2SScam, 0.2M KCl, pH 8.7, $D_2O$, 20°C. The region of the spectrum containing resonances from aromatic side chains and downfield shifted αCH resonances is shown.

Addition of salts has been shown to stabilise marginally stable protein structures in a number of cases; this may arise from a neutralisation of electrostatic attractions on the protein surface which promote aggregation (Goto *et al.*, 1990), or by increasing the surface tension of the solvent which can stabilise hydrophobic interactions in the protein (Collins & Washabaugh, 1985). For this reason, KCl was added to samples of the *holo* $2SS_{cam}$. As can be seen in Figure 5.8b, the spectrum of *holo* $2SS_{cam}$ in the presence of 0.2M KCl shows some interesting features. Although the protein appears to a large extent to be in a molten globule state under these conditions, there is clear evidence for some structure, as indicated by the presence of dispersed aromatic resonances, downfield shifted αCH resonances (δ 4.8 – 6 ppm), and upfield shifted methyl groups ($\delta < 0.0$ ppm) at low intensity. The possibility of specific binding of KCl inducing structure may be ruled out by the fact that such structure is not induced by the addition of salt to the *apo* protein, as shown in the next section. This suggests that structure is induced by the binding of calcium ions, but only in the presence of KCl as a stabilising agent. Two possibilities arise from this: either an equilibrium exists between a native-like structure and a molten globule, or the molecules are partly folded, containing a region of fixed tertiary structure which is induced by calcium binding, and which gives rise to the chemical shift dispersion observed. Addition of further $CaCl_2$, up to 10 mM, does not induce any further structural changes (precipitation occurred above 10 mM $Ca^{2+}$), consistent with the previously measured $K_d^{Ca}$ of $2\times10^{-5}$ M (Ewbank & Creighton, 1993b). Similarly, addition of further KCl does not further stabilise any structure such that it can be detected by NMR, lending support to the latter hypothesis.

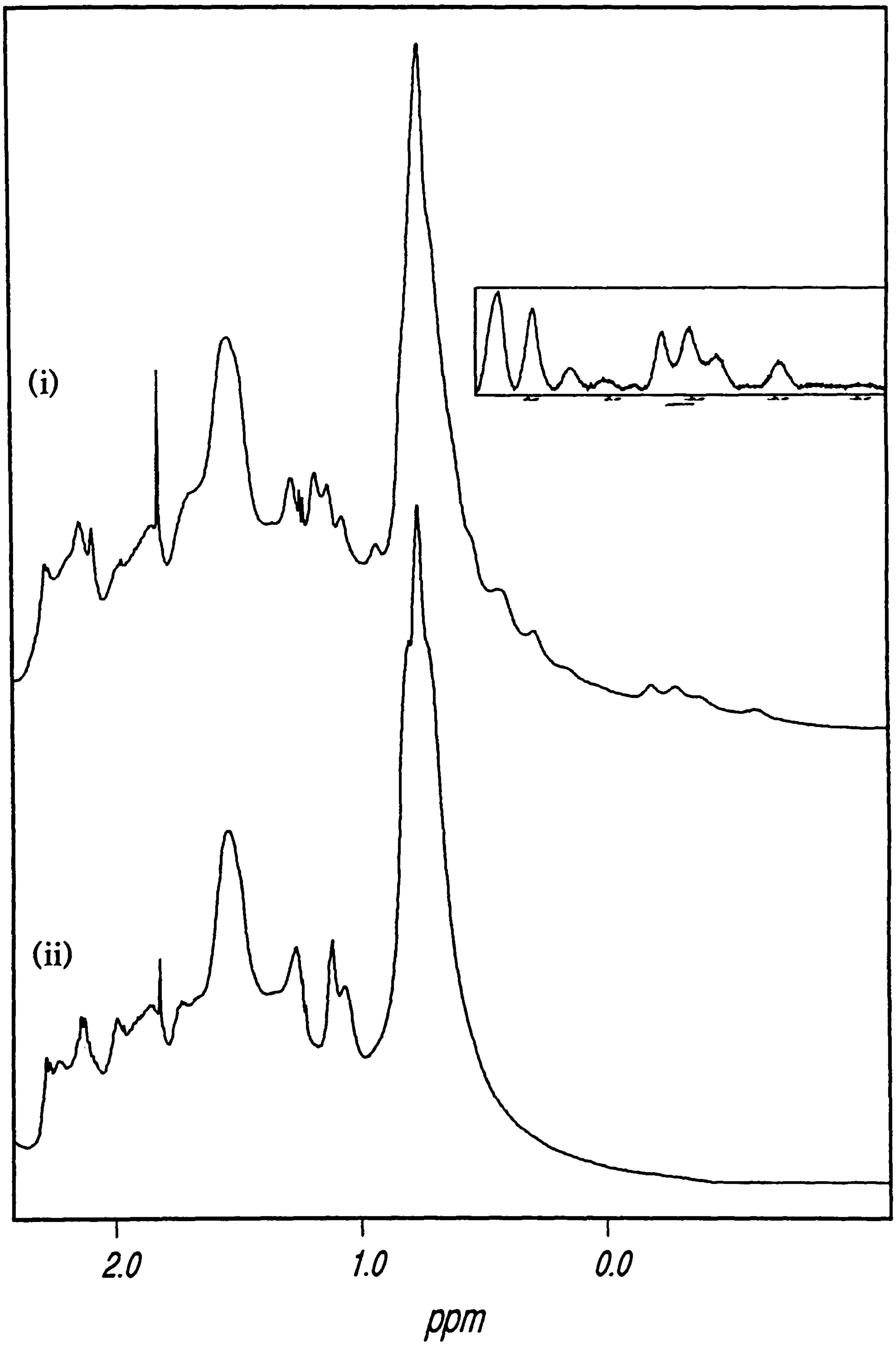


**Figure 5.8b**. Comparison of the NMR spectra of (a) *holo* 2SScam, pH 8.7, 20°C, and (ii) *apo* 2SScam, 0.2M KCl, pH 8.7, $H_2O$, 20°C. Inset: the upfield region is shown magnified after baseline correction to demonstrate the low intensity resonances.

Removal of the two disulphide bridges Cys6–Cys120 and Cys28–Cys111 destabilises the native structure apparently to such an extent that it cannot form, even in the presence of stabilising salts. These two disulphides occur in the α-domain of the native structure, so it may be this subdomain of the protein which ceases to be structured in $2SS_{cam}$. Even without these disulphides, however, there is a region of the protein which can bind a calcium ion, and, in doing so forms partial structure which gives rise to dispersion in the $^1H$ NMR spectrum. The calcium binding site in the native state comprises four aspartate residues (Asp82, Asp84, Asp87 and Asp88) which are adjacent in the primary sequence, and are situated in the β-domain, so it would seem reasonable to suggest that these may be the side chains involved in binding of calcium in this case also. This is particularly interesting since Peng & Kim (1994) have found that the two disulphides of the α-domain are unneccessary for formation of a molten globule structure in a peptide model of the α-domain of α-lactalbumin, and that the Cys28–Cys111 disulphide bridge plays an important role in stabilising the molten globule state (Peng *et al.*, 1995) The aromatic residues in α-lactalbumin occur mainly within the α-domain of the protein also, so there may be some local interactions in this domain which also give rise to some dispersion in the NMR spectrum (cf. Smith *et al.* (1994)).

The CD spectrum of *holo* $2SS_{cam}$, as demonstrated by (Ewbank & Creighton, 1993b), is significantly diminished in both the far and near-UV regions in comparison with the $Ca^{2+}$-bound form of the four-disulphide protein (Figure 5.2). The changes in the CD spectrum are greater than the contributions predicted solely for the breaking of two disulphide bonds (Kuwajima *et al.*, 1990), indicating a loss of secondary structure. The persistence, however, of some signal in the near-UV CD indicates that although the tertiary structure has been significantly disrupted, some constrained aromatic or disulphide chromophores continue to be present. These data are consistent with the NMR results presented here, and together are strong evidence that *holo* $2SS_{cam}$ under these conditions is a molecule which is partially structured. This state is distinct from the 'classical' molten globule states of $3SS_{cam}$ and the four-disulphide protein, which are an ensemble of interconverting states, however, but instead is a structural state in which a portion of the molecule is highly structured with at least some fixed tertiary interactions, possibly as a result of specific calcium binding. Further support for this conclusion comes from results of studies on a peptide fragment of bovine α-lactalbumin, which encompasses the β-sheet region of the protein and the $Ca^{2+}$ binding site (Creighton, T.E., personal communication). This peptide also becomes structured to some extent on binding of calcium ions, as judged by circular dichroism.

At elevated temperature, this unstable structure induced by calcium binding unfolds completely at ~30°C, with significant sample precipitation, as shown in Figure 5.9. It proved impossible to measure an accurate $T_m$ due to the effects of aggregation. This clearly demonstrates, however, the significant destabilising effects which removal of the two disulphide bonds Cys6–Cys120 and Cys28–Cys111 have on the protein structure.

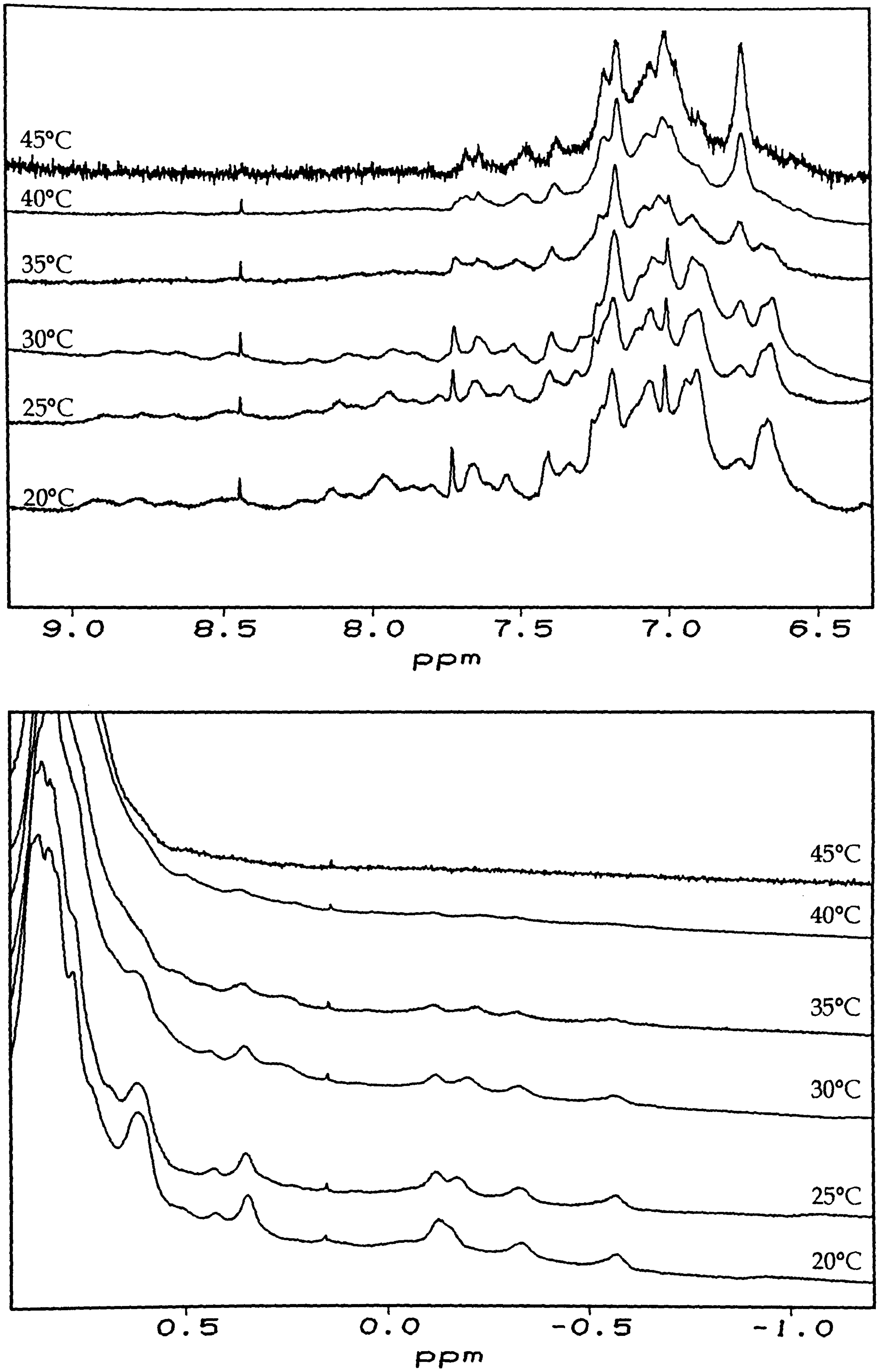


**Figure 5.9**. Temperature unfolding of 2SS$_{cam}$, pH 8.7, 0.2M KCl,monitored by 1D NMR in downfield (upper) and upfield (lower) regions of spectrum.

## 5.10 *The apo form of 2SS$_{cam}$.*

The NMR spectrum of 2SS$_{cam}$ in the absence of Ca$^{2+}$ appears to be characteristic of a molten globule (Figure 5.8a). There is very little dispersion of resonances from their random coil chemical shift values and peaks are broadened, indicative of the slow tumbling rate and intermediate exchange in a partially folded protein (Baum *et al.*, 1989). There are a number of sharper peaks present in the spectrum, however, suggesting that under these conditions an equilibrium exists between the molten globule and a more unfolded state, reflecting the reduction in stability of this partially structured state on removal of two of the four native disulphide bridges. In this case there are no resolved resonances arising from fixed tertiary interactions, which suggests that a native-like state is not appreciably populated under these conditions. Addition of KCl up to a concentration of 0.5M induces no further change in the NMR spectrum (data not shown), indicating that salt alone is not sufficient to induce folding of 2SS$_{cam}$ in the absence of Ca$^{2+}$.

The CD spectrum of *apo* 2SS$_{cam}$ is shown in Figure 5.2 (Ewbank & Creighton, 1993b). The signal in the far-UV region is significantly diminished relative to that of the *apo* four-disulphide protein, and the near-UV CD signal is almost completely absent, indicating, in accord with the NMR data presented above, that the majority of the sample population is unfolded. The small signal in the far-UV CD has been attributed to partial population of the molten globule conformation (Ewbank & Creighton, 1993b).

## 5.11 *Concluding remarks*

These disulphide derivatives of α-lactalbumin form a fascinating series of proteins. The parent four-disulphide protein unfolds *via* a three state process involving a molten globule intermediate. The three-disulphide derivative adopts a native-like conformation in the *holo* form, but this is significantly less stable than the unmodified protein, again unfolding *via* a molten globule state. The *apo* form, like that of the unmodified protein, exists as a molten globule state. Removal of a second disulphide bridge from the protein, however, destabilises the structure to such an extent that the native structural state is not detectably populated even in the presence of $Ca^{2+}$ and salt. Instead, in the presence of the $Ca^{2+}$ ligand and KCl, one region of the protein is folded, whilst the remainder of the molecule appears to be largely unstructured. This structure may include the native calcium binding loop and certainly involves a number of aromatic interactions, as judged by the signal in the near-UV CD and ring current induced shifts in the NMR spectrum of this species.

An understanding of the structures and stabilities of partially folded species can give useful insight into the processes which occur during protein folding. The molten globule state of four-disulphide α-lactalbumin has been shown to share strong resemblances with a transient kinetic intermediate formed during the refolding of lysozymes and α-lactalbumins as judged by circular dichroism and protection against hydrogen exchange. Structural characterisation of these intermediates at equilibrium can yield valuable information as to the interactions taking place during the kinetic refolding process. Studies of proteins lacking the full complement of native disulphide bridges are of particular interest because of their relationship to intermediates formed during oxidative refolding. This approach of dissecting the protein folding problem has been somewhat successful for BPTI. Isotopically enriched analogues of the species formed during oxidative refolding of BPTI (Creighton, 1974, 1988), have been structurally characterised by NMR (Darby *et al.*, 1991; van Mierlo *et al.*, 1991a,b, 1992, 1993, 1994). Since these intermediate species may be trapped at equilibrium, one can be optimistic that these structural states might be representative of the stages through which a folding protein passes as it forms its native structure from the unfolded polypeptide chain (Dobson, 1994). Further studies by heteronuclear NMR methods of isotopically enriched α-lactalbumin with one or more disulphides engineered out may enable more detailed characterisation of the effects which these disulphide bridges have on the structure and stability of the native and molten globule states of α-lactalbumin, and of the folding process in general.

# Chapter Six

# Electrospray ionisation mass spectrometry as a tool for studying protein folding.

## 6.1 *Introduction*

Over the past five years NMR spectroscopy has been demonstrated as one of the most valuable techniques for the investigation of protein folding. Pulse labelling techniques in conjunction with NMR have enabled characterisation of transient intermediates by the nature of their protection against hydrogen exchange (Baldwin, 1993). More recently, with the development of higher field instrumentation and multinuclear detection, structure elucidation of unfolded and partially folded states of proteins has become possible (Neri *et al.*, 1992b). No doubt advances will continue to be made in this field, but at present NMR has a number of limitations which restrict its usage more generally: the most important of these is molecular weight. Due to the line width and chemical shift degeneracy of proton resonances in proteins, there is an upper molecular weight limit on proteins which may be structurally characterised by NMR, even using heteronuclear multidimensional techniques (Clore *et al.*, 1991). This limits its use in the study of protein structure and also of folding to a subset of relatively small monomeric proteins, and to proteins with well defined structures. NMR has been used with some success as a tool for studying protein-

protein complexes. Elegant experiments have been performed which have allowed the nature of interactions of small proteins with larger ones, for example the cyclophilin–cyclosporin interaction has been investigated by perdeuteration of the host molecule (Spitzfaden *et al.*, 1994), antibody–antigen complexes (Scherf *et al.*, 1995) and the interaction of small peptides with chaperones has been studied by transferred nOe experiments (Landry & Gierasch, 1991; Landry *et al.*, 1992). In each of these cases, however, the amount of information which may be obtained is somewhat limited by the large size of the host molecule, and particularly in the latter case, specific conditions must be found at which the rates of association and dissociation are favourable for the experiment.

Mass spectrometry has been in use for many years as a means of accurately determining the mass of molecules which can be ionised (Fenn *et al.*, 1989). A wide variety of techniques of ionisation have been developed, including electron impact, chemical ionisation and fast atom bombardment, which can give masses to high accuracy from small molecules to large proteins, but these tend to require volatile samples and extreme conditions. Recent developments in methodology, in particular the introduction of electrospray ionisation mass spectrometry (ESI-MS), have led to an upsurge in the use of mass spectrometry for the study of protein interactions and folding. The gentle non-destructive nature of this technique has allowed experiments to be performed such as the observation of enzyme-substrate complexes (Ganem *et al.*, 1991; Lumb *et al.*, 1992), detection of conformational changes in proteins by changes in charge state distributions and hydrogen exchange (Chowdhury *et al.*, 1990; Katta & Chait, 1991; Katta & Chait, 1993), and, perhaps most interestingly, in the context of protein folding, detection of intermediates in folding by pulsed hydrogen exchange labelling (Miranker *et al.*, 1993; Eyles *et al.*, 1994). The difference in mass between the two hydrogen isotopes, hydrogen and deuterium, enables the extent of incorporation of deuterium into a sample to be determined by measuring the mass of the protein. This technique has the added advantage that heterogeneity of exchanging species within a sample may be measured from the line widths of peaks in mass spectra (Miranker *et al.*, 1993).

This chapter describes some of the techniques which we have developed, based on the use of electrospray ionisation mass spectrometry (ESI-MS), for the study of the stability and folding of proteins. The methods involve analysis of picomolar quantities of protein from aqueous solvent at low temperature, hence enabling the maintenance of protein structure and its extent of hydrogen/deuterium labelling during the MS experiment. The uses of ESI-MS to investigate cooperative elements in protein folding, the stability of native, unfolded and partially folded states of proteins, and the nature of interactions of partially folded proteins with chaperones are described here.

### 6.1.1 The technique of ESI-MS.

Conventional mass spectrometry techniques (electron impact, chemical ionisation and FAB) require heat and high energy processes to achieve ionisation of molecules in the gas phase. These give spectra routinely of high mass accuracy and may be used for a wide range of molecules from small organic molecules to involatile inorganic molecules and polymers. Such ionisation techniques are, however, inappropriate for the study of intact protein structures, or for monitoring hydrogen exchange events, since these conditions lead to complete denaturation of the protein and loss of hydrogen exchange protection. Alternative methods of ionisation have more recently been developed, however, which are capable of ionising molecules using much less harsh and invasive conditions. Matrix assisted laser desorption ionisation MS has been shown to be useful for the study of intact proteins (Katta & Chait, 1991; Senko & McLafferty, 1994), but the most versatile of these techniques is ESI-MS (Fenn *et al.*, 1989).

The principle of electrospray ionisation is to introduce a dilute solution of sample, conventionally dissolved in a mixture of water and acetonitrile, from a fine metal needle maintained at high positive potential (+5 kV) into a dry gas atmosphere with a strong electric field (see Figure 6.1). The droplets contain protein with bound solvent molecules which carry a positive charge, due to the low pH of the solution, and are attracted to the inlet of the mass analyser. The flow of dry gas causes these solvent molecules to evaporate, eventually leaving ionised protein molecules in the gas phase which then proceed into a quadrupole mass analyser. By careful adjustment of the potentials and gas flow conditions it is possible to produce completely desolvated single protein molecules for mass/charge (m/z) analysis. Using this technique very high mass accuracies (approximately 1 part in 10 000) can be obtained, and proteins with masses up to 60 kDa and beyond can be studied readily (Feng & Konishi, 1992).

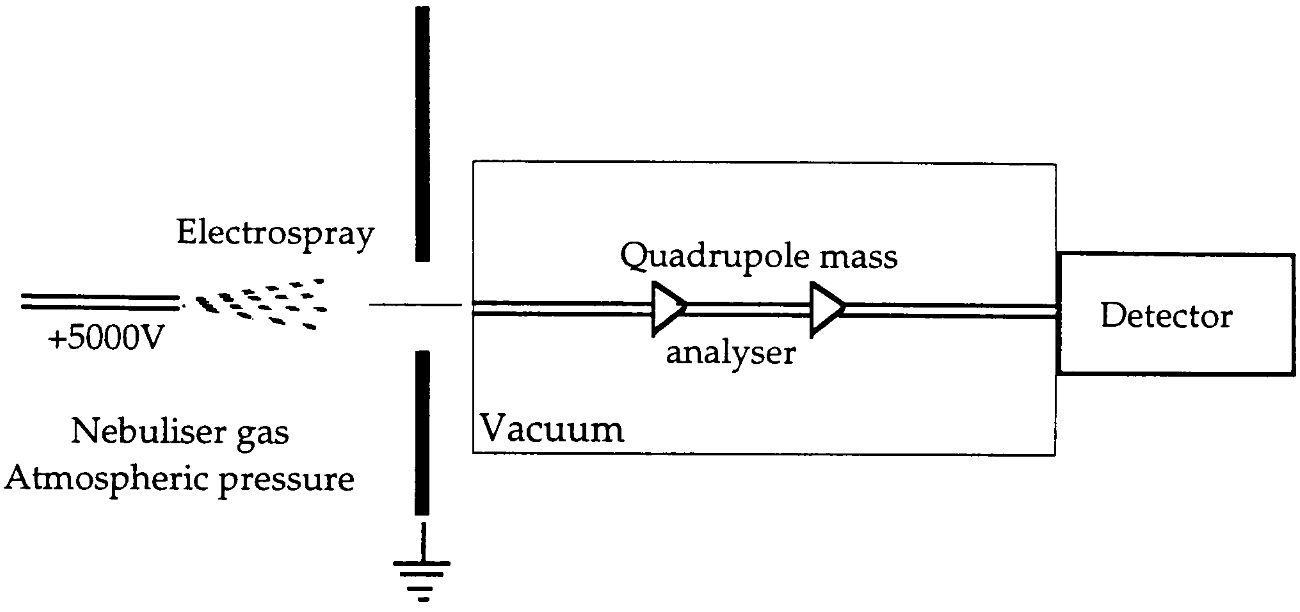


**Figure 6.1.** Schematic representation of an electrospray ionisation mass spectrometer.

---

In order to make ESI-MS a useful tool for the study of protein folding, it was necessary to make a number of changes to the conventional instrumentation. Hydrogen exchange in proteins occurs very readily, and particularly in situations where the structure is disrupted by destabilisation and partial denaturation (Hvidt & Nielsen, 1966; Englander & Kallenbach, 1984). To study native and molten globule as well as denatured states by mass spectrometry, therefore, we have developed a methodology which involves spraying protein samples from aqueous solution at low temperature. This is achieved by using water adjusted with formic acid to the required pH for study (usually pH<5) as the delivery solvent, pre-equilibrating the solvent delivery system and sample interface for several hours at 4°C, and by cooling the nebulizer gas over an salt-ice bath. The high tension voltage and nebulizer gas flow rate is carefully controlled to ensure complete desolvation of protein ions with minimal heating (20°C). This chapter describes how, using these conditions, we have been able to obtain consistent hydrogen exchange kinetics for proteins with relatively unstable native states such as $CM^{6,127}$-lysozyme (Chapter Four), and also of molten globule states and chaperone-bound partially folded states of proteins.

### 6.1.2 Cooperativity in protein folding monitored by ESI-MS.

NMR and ESI-MS have proved to be highly complementary techniques for the study of protein folding by hydrogen exchange pulse labelling (Miranker *et al.*, 1993) and see Chapter 4. NMR is a powerful tool for monitoring the rate at which individual amides become protective against hydrogen exchange, and can therefore enable determination of the order of formation of positive structure in regions of the three dimensional structure of the protein. The intensity of an amide resonance is, however, only a measure of the average proton occupancy in the entire sample population, so information about distinct populations of molecules with different levels of protection cannot be obtained. By contrast, although ESI-MS cannot monitor exchange on a residue specific basis, it can distinguish individual populations of protein molecules with distinct masses.

Figure 6.2 shows a simple model involving *n* amides with an average of 50% deuteration. In the first case the amides have completely exchanged in 50% of the population, but not at all in the other 50% (Figure 6.2a), while in the second case all of the amides in the sample have exchanged with 50% probability (Figure 6.2b). NMR is incapable of distinguishing these two situations since the amide intensity is an average of the proton occupancy over the whole population. ESI-MS can clearly distinguish between these situations, however, since two peaks would be detected in the former case, the lower mass peak corresponding to both amides protonated, the peak at higher mass resulting from deuteration at both amide sites. In the latter case a single peak is observed at a mass equivalent to 50% of the maximum deuterium content. In the case of a real protein, which will have many more amides, not only the peak mass but also the shape and width are informative, since they can give insight into the diversity and cooperativity of exchange mechanisms.

The use of ESI-MS in conjunction with hydrogen exchange pulse labelling has already been demonstrated in this laboratory for investigation of the kinetic refolding of hen lysozyme (Miranker *et al.*, 1993). In this study, mass spectrometry was used to show clearly the formation of the α-domain transient folding intermediate as a cooperative folding unit. NMR results have previously indicated that the rate of protection of amides monitored in the α-domain was closely similar (Radford *et al.*, 1992b). ESI-MS was able to demonstrate unequivocally, however, that these amides were indeed protected cooperatively as shown by the appearance of a peak corresponding to a population with 28 sites deuterated, the number of amides protected in the α-domain. The time scale for protection also corresponds well with that measured by NMR.

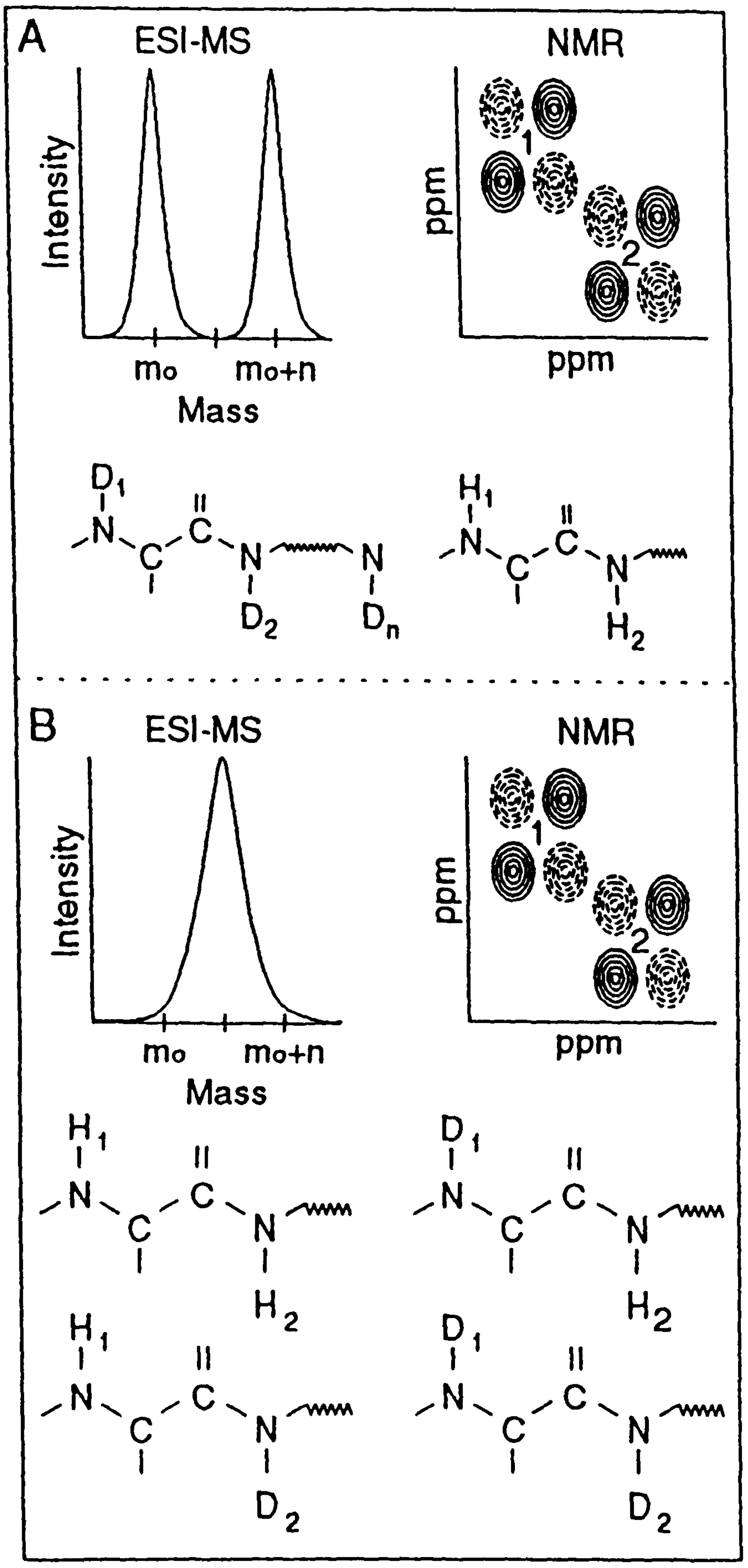


**Figure 6.2.** Schematic diagram to show that ESI-MS can distinguish between (A) 50% of population unexchanged, 50% completely exchanged, (B) 50% exchange evenly distributed throughout the entire population. Reproduced from Miranker *et al.* (1993)

In Chapter Four, ESI-MS again was shown to be essential in measuring cooperativity of folding events. In the refolding of CM$^{6,127}$-lysozyme, all of the amides become protected against exchange on a similar time scale (Eyles *et al.*, 1994). By NMR methods alone it was not possible to distinguish between two possibilities that either the entire protein molecule folds cooperatively, or that removal of the Cys6–Cys127 disulphide bridge merely slows formation of the α-domain to such an extent that α-domain and β-domain intermediates are formed independently but at similar rates. Again ESI-MS can distinguish between these two possibilities and the time course of protection clearly demonstrates that the former case is applicable, since no evidence of any intermediate species was observed. CM$^{6,127}$-lysozyme refolds as a cooperative unit involving no protective intermediates.

## *6.2 Determination of hydrogen exchange protection in the folding of lysozyme variants.*

In order to investigate further the effects of specific interactions in the native state on the folding pathway of proteins, a number of studies have employed protein engineering to introduce mutations which disrupt these interactions (Matouschek *et al.*, 1990; Eriksson *et al.*, 1992; Fersht & Winter, 1992; Fersht & Serrano, 1993; Matthews, 1993). An alternative approach is to make use of naturally occurring mutations in variant proteins. This approach was chosen for the study of the role of specific interactions in the folding pathway of lysozyme. Since the amino acid substitutions in these variant proteins involve only a small number of residues and their effect may be very small, it is crucial to ensure that the refolding conditions are identical in all cases, in order to be certain that any differences observed result solely from the changes in the amino acid sequence. The pulse labelling experiment is very sensitive to small differences in, for example, temperature, denaturant concentration and the buffer pH at each stage of the experiment; small differences are therefore difficult to distinguish from experimental errors. To overcome this, an ideal approach is to refold two or more proteins simultaneously in the same experiment. This of course is not possible if hydrogen exchange is monitored by NMR, since a large number of the resonances from a mutant protein will be coincident with those of the wild type, although this problem could theoretically be circumvented by complex isotope editing techniques. If, however, a variant protein is significantly different in mass from the wild type, then the two proteins may be resolved readily by mass spectrometry, enabling simultaneous measurement of the folding properties of the two proteins ensuring that the refolding conditions are identical throughout.

This approach is described here to compare the folding of hen lysozyme with variant *c*-type lysozymes obtained from Lady Amherst pheasant, Japanese quail and human sources, which differ from the hen protein by 6, 6 and 52 substitutions, respectively (see Appendix B). In the case of the former two proteins, the masses of the fully protonated species of 14303 and 14358 Da are too close to those of hen lysozyme (14305 Da fully protonated unprotected, and 14365 Da for the fully protected species) to allow direct comparison in the mass spectrometer. This was circumvented by the use of hen lysozyme uniformly isotopically labelled with $^{15}N$, which has a fully protonated mass of 14498 Da and hence may readily be resolved from the $^{14}N$-containing variant species by ESI-MS.

Figure 6.3 show three charge states of the ESI-MS spectrum of hen lysozyme, refolded and pulse labelled after 38 ms of refolding. The protein sample consists of a mixture of $^{14}N$ lysozyme (0.1% natural abundance $^{15}N$) and lysozyme with >98% incorporation of $^{15}N$ . The two series of peaks can clearly be resolved from one another by virtue of the mass difference arising from their differing $^{14}N/^{15}N$ content. Three peaks appear for each protein at each charge state as a result of the partial deuteration of species during the hydrogen exchange labelling experiment. The lowest mass peak in each charge state corresponds to the unfolded and hence unprotected protein molecules (14313.1±1.1 Da) that remained after 38 ms of folding (labelled U). This mass is slightly heavier than that of fully protonated hen lysozyme due to the residual deuterium content (6.25%) in the refolding buffer during the labelling pulse. The peak labelled N corresponds to those molecules which had folded to a fully protected state (14368.03±3.6 Da), trapping 50±1 deuterons prior to the labelling pulse. Finally the third peak (I, 14342.0±0.5 Da) corresponds to a species 28±1 Da heavier than the unprotected state. This has been shown by NMR to result from formation of a transient folding intermediate in which the α-domain of the protein becomes structured and protective against exchange while the remainder of the molecule remains unstructured (Radford *et al.*, 1992b). Identical populations of U, I and N species are observed for both $^{14}N$ and $^{15}N$ hen lysozyme within the same sample at this and all other time points measured from 3.5 ms to 2 s, clearly indicating that the folding behaviour is identical for both proteins.

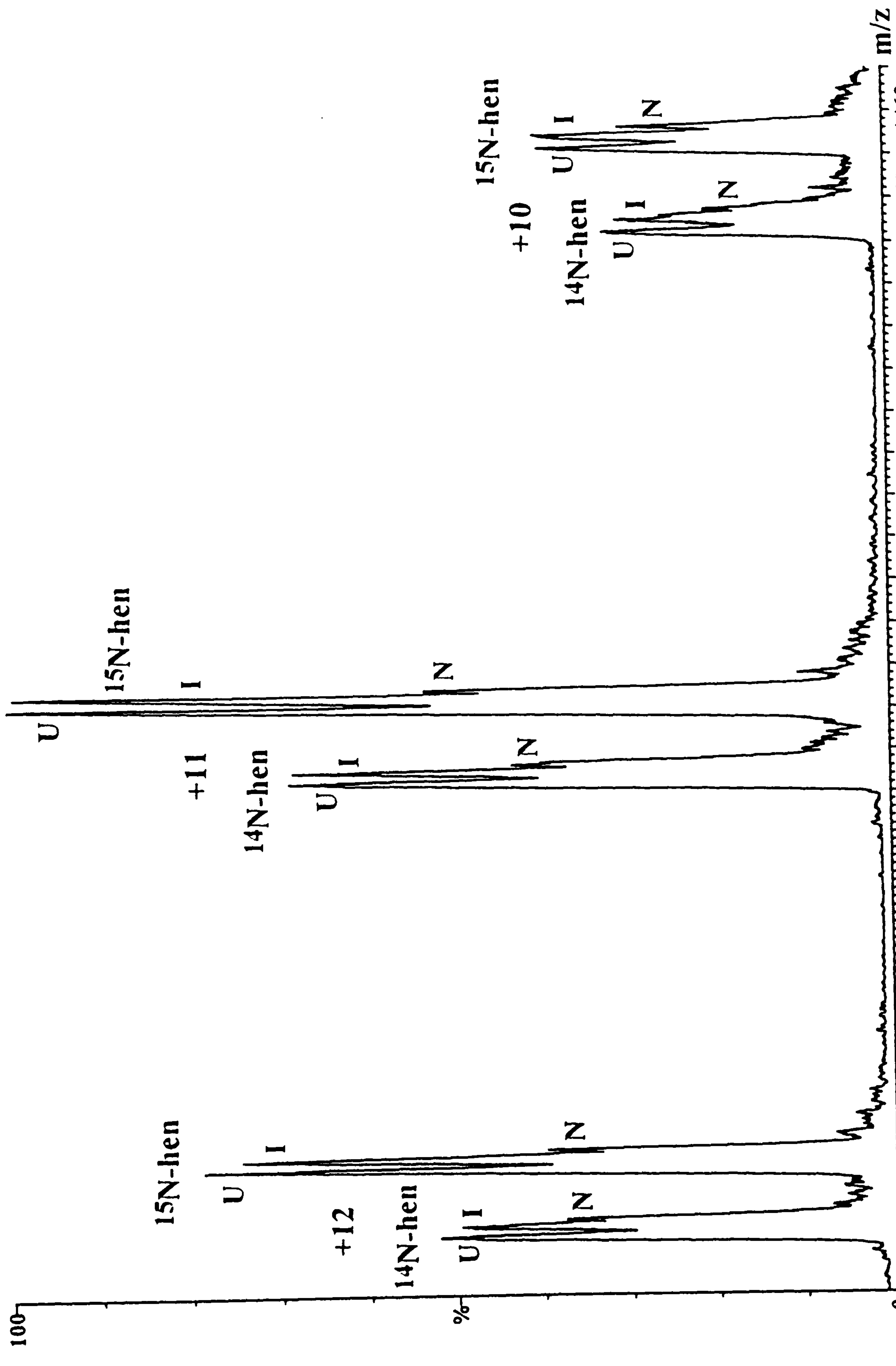


**Figure 6.3.** ESI-MS of approximately equimolar amounts of $^{14}$N-hen lysozyme and $^{15}$N hen lysozyme 38 ms after initiation of refolding. The charge states shown are the +10, +11 and +12 for both isotopic states.

The folding behaviour of hen lysozyme (HEWL) was then compared with that of the lysozymes from Lady Amherst pheasant (LAPL), Japanese quail (JQL) and human (HL) sources as shown in Figure 6.4. Similar folding behaviour exhibited by all four proteins, as seen by the formation of a well defined transient intermediate (labelled I) which affords protection to 28±1 deuterons, and is then converted to the native state as folding is completed. This clearly indicates that the α-domain intermediate forms in a highly cooperative manner in each of these variant lysozymes, despite their differences in sequence. Even in the case of human lysozyme, which shares only 40% identity with the hen protein, the pattern of protection in partially folded states is remarkably similar.

There are clear differences, however, in the rate of formation of both the α-domain intermediate and the fully folded state in each of these four proteins, as demonstrated by the relative populations of the species at a given refolding time (Hooke *et al.*, 1995). Protection of the α-domain in JQL, for instance, clearly occurs on a slower time scale than the hen protein or LAPL, although the overall folding rates to the fully protected states are very similar. It is interesting to note that the six substitutions in JQL relative to the hen protein all occur in the α-domain, indicating that the interaction of one or more of these residues with other regions of the protein influences the kinetics of formation of this intermediate. These differences in kinetics can be quite subtle and herein lies the beauty of this technique for direct comparison of refolding under identical conditions; refolding of each protein may be directly and conclusively compared with that of the hen protein.

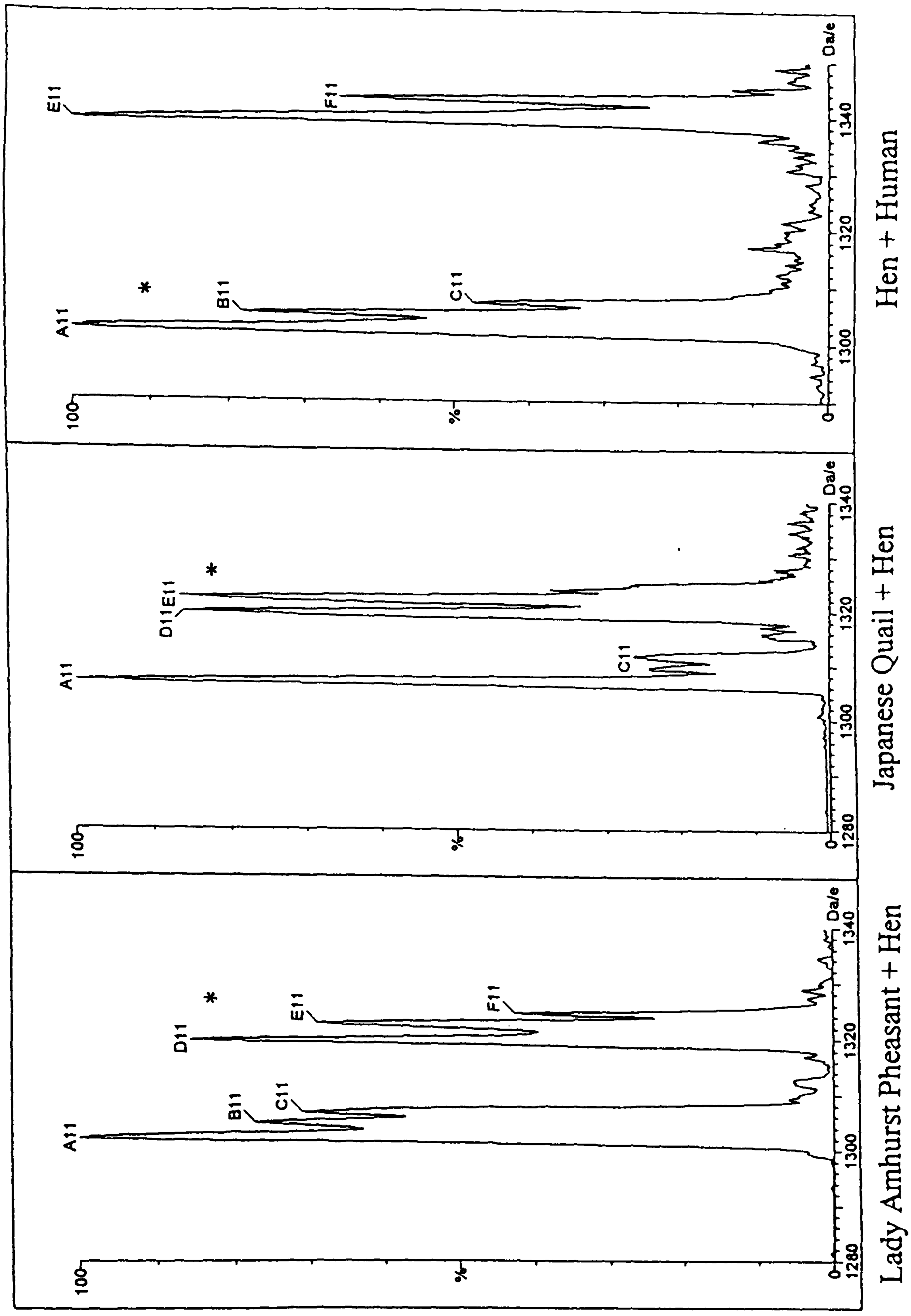


**Figure 6.4.** ESI-MS spectra of (a) LAP and $^{15}N$ hen lysozymes, (b) JQ and $^{15}N$ hen lysozyme and (c) human and $^{14}N$ hen lysozymes all pulse labelled 38ms after initiation of refolding. The +11 charge state is shown in all cases.

Comparison of the ESI-MS spectra of hen and human lysozymes demonstrates the most significant difference between the folding characteristics of these proteins (Figure 6.4). In human lysozyme after 3.5 ms of folding, the earliest possible time of measurement, a single population of molecules with 17±1 deuterons protected is observed. This folds to form the complete $\alpha$-domain within 38 ms. By contrast, in the hen protein a species with only 17 deuterons protected is not detected. ESI-MS alone cannot determine the nature of this very early intermediate. NMR experiments, however, have clearly shown that the kinetics for protection of amides in the A and B $\alpha$-helices and C-terminal $3_{10}$ helix of the protein are faster than those of the remaining two helices in the $\alpha$-domain (C and D), indicating formation of a subdomain comprising the A and B helices prior to complete packing together of the remainder to form the $\alpha$-domain (Hooke *et al.*, 1994). This accounts for the species observed with 17 protected deuterons by ESI-MS in the refolding of human lysozyme.

Detailed study of the folding of lysozymes by optical techniques has shown that collapsed states involving formation of significant amounts of secondary structure are formed within the dead time of the experiment (2 ms). It is interesting to note here that, in spite of these very rapid stages, there can be large variations in the time scale for later events which presumably involve reorganisation of secondary structure elements to form native structure, even for highly homologous proteins such as the variant lysozymes used for this study. Further investigation of the residues which are the origin of these differences in folding properties may be achieved by site directed mutagenesis. Clearly ESI-MS can be used as a rapid and effective tool for directly probing these differences using very small quantities of protein, suggesting that it may prove to be a very powerful technique for studies of this type.

## *6.3 Structural studies of lysozymes and α-lactalbumins by ESI-MS*

As well as the study of refolding by pulse labelling techniques, protection against hydrogen exchange has been used as a means of monitoring structural stability and measuring dynamic processes in various states of proteins (Linderstrøm-Lang, 1955; Woodward *et al.*, 1982; Englander & Kallenbach, 1984). Exchange in native proteins is thought to arise from a number of events, involving solvent penetration on the surface of the structure (Lumry & Rosenberg, 1975), and local unfolding of regions of structure, as well as global unfolding events (Englander *et al.*, 1980). The contribution of each of these processes varies as a function of the stability of the protein and its dynamic behaviour in solution. The theories of the mechanisms governing hydrogen exchange are complex and are discussed further in Chapter 3 and references therein.

Using the methodology described in the previous section such that proteins can be analysed at low temperature from aqueous solvents, we have been able to monitor hydrogen exchange in various lysozymes and α-lactalbumins in different conformational states, whilst ensuring that significant exchange at labile sites does not occur during the experimental measurement by comparison with NMR data obtained under identical experimental conditions.

The proteins selected were hen lysozyme, CM$^{6,127}$-lysozyme and α-lactalbumins obtained from bovine and guinea pig milk, and conditions chosen were pH 5.0 and pH 2.0. All of the proteins are in their native state at pH 5.0, whereas at pH 2.0, 20°C, CM$^{6,127}$-lysozyme is completely unfolded, α-lactalbumin adopts a molten globule state and unmodified lysozyme retains its native structure. The two α-lactalbumin species were chosen to allow investigation of any differences in the stability of their molten globule states. This gives a unique opportunity to study the structure and stability of proteins using very small quantities of sample. Useful information may be obtained not only from the time course of exchange, but also from the observed differences in isotope distribution (line width) and charge state distribution (Chowdhury *et al.*, 1990; Katta & Chait, 1991; Loo *et al.*, 1991; Katta & Chait, 1993).

Figure 6.5 shows mass spectra of bovine α-lactalbumin in the native state at pH 5.0 in the presence of the $Ca^{2+}$ ligand, in the molten globule state at pH 2.0, and in the unfolded state in 50% acetonitrile. The most striking differences between the three spectra are the distributions between the charge states of the protein ions detected. These molecular ions are highly protonated, reflecting in a qualitative way the large number of basic sites available for protonation at the protein surface (Fenn *et al.*, 1989). In the spectrum of the native state of bovine α-lactalbumin the charge state distribution is centred around the +10 state, and the mass of this series corresponds to 14216.8 ± 1.4 Da. This is equivalent to the mass calculated from the amino acid sequence (Shewale *et al.*, 1984) for the protein (14216 Da) with a single $Ca^{2+}$ ion bound and removal of two protons from the Asp residues involved in calcium binding. The mass obtained is in complete agreement with the published amino acid sequence (Shewale *et al.*, 1984), although a number of ambiguities were found in the different published sequences for bovine α-lactalbumin (Brew *et al.*, 1970; Shewale *et al.*, 1984; Acharya *et al.*, 1989; Acharya *et al.*, 1991), necessitating careful confirmation of this mass using a high resolution VG Autospec mass spectrometer equipped with an electrospray interface at Fisons Instruments Ltd, calibrated against hen lysozyme and equine cytochrome *c*. In this case we found it necessary to include $Ca^{2+}$ ions in the delivery solvent (1 μM) to maintain the protein in its calcium bound state, since the high dilution and formation of the positively charged molecular ion otherwise caused the metal ion to be abstracted from the protein. This clearly shows, however, that the protein can be maintained in a state in the spectrometer that continues to bind calcium throughout the electrospray process. From the spectrum of bovine α-lactalbumin in 50% acetonitrile, an $M_r$ of 14178.04 ± 0.83 is obtained, and the charge state maximum shifts to +12, consistent with loss of the $Ca^{2+}$ ligand (calculated $M_r$ 14178) and unfolding of the native structure of the protein under these conditions. The increased charge state observed presumably results from exposure of basic side chains which are buried in the native state, and changes in $pK_a$ values as a result of removing electrostatic interactions such as salt bridges on addition of acetonitrile and unfolding of the protein (Katta & Chait, 1993).

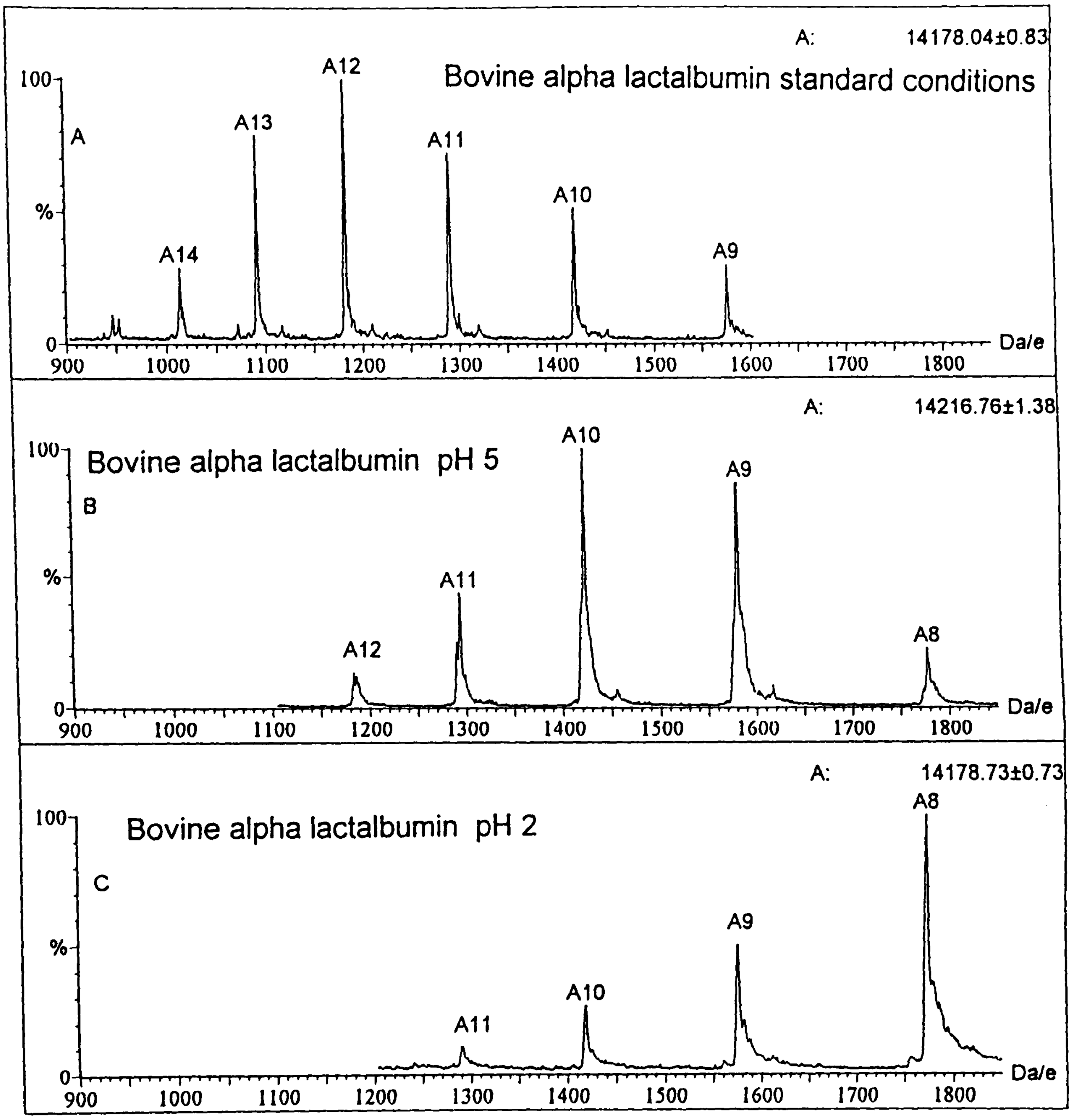


**Figure 6.5.** ESI-MS spectra of BLA obtained from (a) 50% (v/v) acetonitrile, 1% formic acid, (b) water pH 5.0 and (c) water pH 2.0.

The spectrum of α-lactalbumin at pH 2.0 is most interesting in this respect in that the charge state maximum occurs around +8 with a corresponding mass of 14178.8±0.7 Da, indicating in accord with NMR data (Ikeguchi *et al.*, 1986b) that the molecule does not bind calcium under these conditions. Although NMR and optical methods clearly show that the *apo* form of α-lactalbumin at pH 2.0 is expanded relative to the native state (Kuwajima *et al.*, 1976; Ikeguchi *et al.*, 1986b), the charge state distribution is shifted to lower charge in this partially folded state. This suggests that the molten globule state has distinct conformational preferences which expose fewer basic side chains to solvent than either the denatured or the native state, thus shifting the observed peaks to higher m/z. This phenomenon has been observed for a number of other proteins which adopt partially folded conformations (C.V. Robinson, unpublished results), although the charge state distribution is likely to be influenced by a number of factors such as the high concentration of formate counter ions present at low pH (Mirza & Chait, 1994). Self-association may be an important factor to consider, since an aggregated species would probably dissociate into monomers in the mass spectrometer, and aggregation of α-lactalbumin in the molten globule state occurs readily (see Chapter Five) (Wormald, 1991). Hence the partially folded state of α-lactalbumin may form structure in such a way as to cause some ionisable groups to be buried away from the surface of the protein.

Measurement of hydrogen exchange from these various structural states was achieved by first deuterating all labile sites on the proteins. Lysozyme was incubated at 80°C in $D_2O$, pD 4.0, $CM^{6,127}$-lysozyme at 60°C in $D_2O$, pD 4.0 for 10 min, and α-lactalbumin at 20°C in $D_2O$, pD 2.0

overnight. These conditions were controlled in each case to prevent aggregation and irreversible denaturation of the proteins. The samples were then lyophilised and the process repeated to ensure complete exchange at all labile sites. Before the final lyophilisation step, the pH was adjusted to that required for the exchange reaction and 1 mM $CaCl_2$ added to form the *holo* α-lactalbumin species. To initiate the exchange reaction, each protein (1.5 nmol/μl in $D_2O$) was diluted 100-fold into $H_2O$ at pH 5.0 or pH 2.0 at 20°C. After various times, an aliquot was removed and the mass spectrum recorded as previously described.

Two extremes of mechanism for hydrogen exchange have been described, depending on the relative rates of protein unfolding and intrinsic exchange rate of a labile proton (Hvidt & Nielsen, 1966). In the pseudo first order case ($EX_1$), the unfolding rate is so much slower than the intrinsic exchange rate that any opening process which exposes labile hydrogens to solvent will enable complete exchange at all sites to occur. In terms of mass spectrometry, after a given period of exchange this would manifest itself as two peaks, one corresponding to no exchange and a second corresponding to complete exchange of all sites. If on the other hand unfolding events are rapid compared with the intrinsic exchange rates, then a second order reaction mechanism will predominate ($EX_2$). In this case exchange is governed by the rates of both of these processes and proton occupancy will be uniform over the population of protein molecules. Mass spectra of these exchanging states will appear as a single peak for each charge state which shifts with time from a mass corresponding to no exchange, to a mass corresponding to complete exchange at long times. The distribution of masses (line width) is a measure of the distribution of intrinsic exchange rates over all of the amides. This can lead to quite complex line shapes since in practice many different unfolding processes may contribute to the exchange reactions (Miranker *et al.*, 1993).

In the case of all states of the structured proteins studied here, only a single mass peak was observed, indicating that all exchange occurs predominantly *via* a second order $EX_2$ process, except in the case of $CM^{6,127}$-lysozyme at pH 2.0, where exchange takes place at approximately the intrinsic rate for the constituent amino acids since it is completely unfolded under these conditions. The $EX_2$ process has been shown previously to be the primary mechanism of exchange even for hen lysozyme at 69°C, in which 10% of the protein population is present in the denatured state, suggesting that the situation of pure $EX_1$ kinetics is only rarely seen (Miranker *et al.*, 1993). In order to extract hydrogen exchange data from the mass spectra, the centroid of the transformed m/z peak was taken as the average mass, then the mass of the fully protonated protein (including bound $Ca^{2+}$ in the case of *holo* α-lactalbumin) was subtracted and corrected for 1% residual solvent deuterium content to yield the number of trapped deuterons remaining in the sample at each time point. The kinetic exchange profiles for the various states of each protein studied are shown in Figure 6.6.

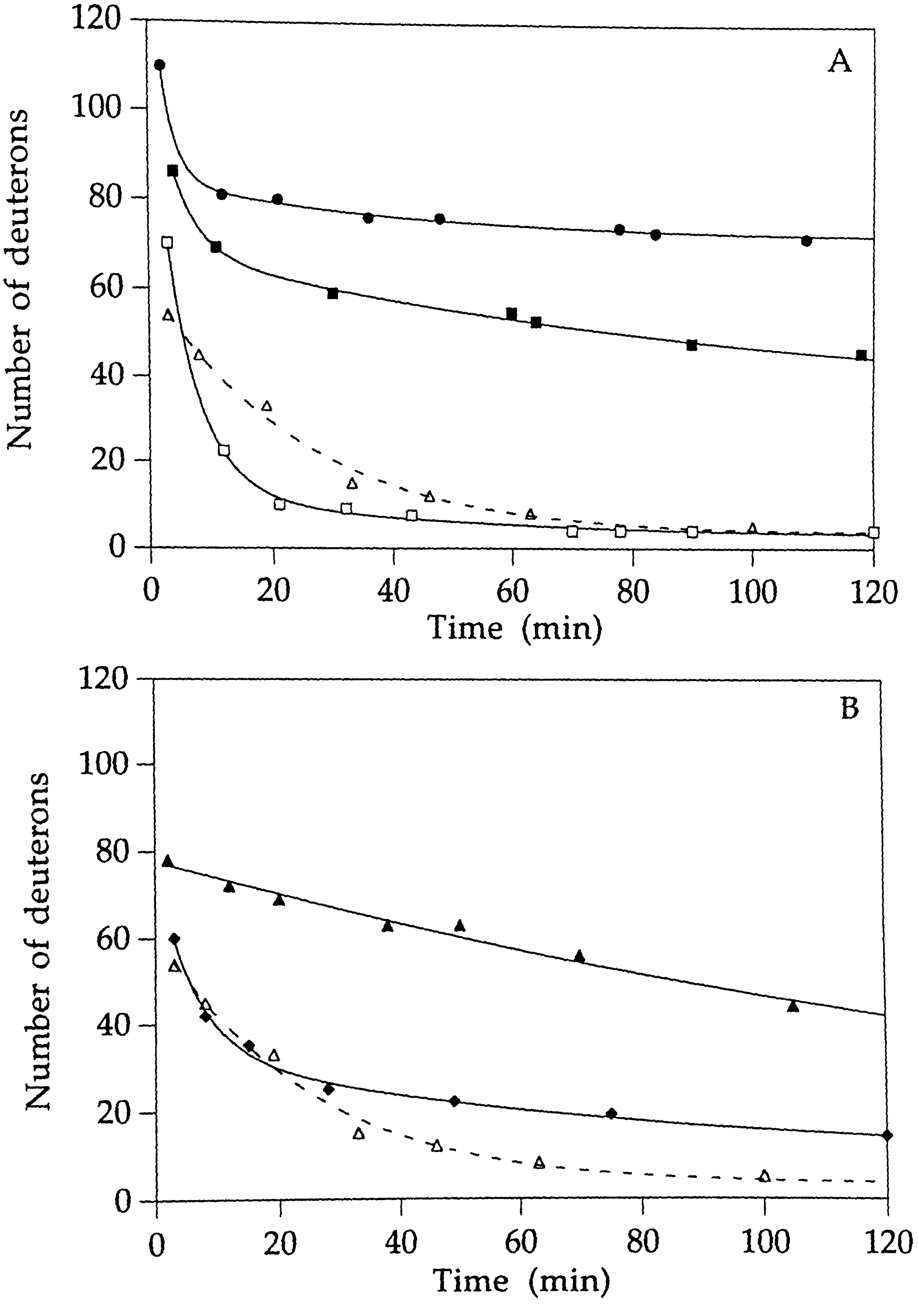


**Figure 6.6.** Hydrogen exchange kinetics of different conformational states of hen lysozymes and α-lactalbumins in water at 20°C.
(A) native hen lysozyme pH 5.0 (●), native CM$^{6,127}$-lysozyme, pH 5.0 (■), BLA molten globule state at pH 2.0 (Δ) and unfolded CM$^{6,127}$-lysozyme at pH 2.0 (□).
(B) native BLA at pH 5.0 (▲), BLA molten globule state, pH 2.0 (Δ) and GPLA molten globule state, pH 2.0 (◆).

■

The native state of hen lysozyme is by far the most protected state identified in this study, with 75 deuterons protected from exchange even after 2 hours of exchange. This confirms that hen lysozyme is stable with high resistance to solvent exchange under these conditions; more than half of the amides are protected against hydrogen exchange. Similarly, $CM^{6,127}$-lysozyme is in a native-like state at pH 5.0 and is also substantially protected against hydrogen exchange, indicative of a stable structure. The level of protection is not as great, however, as for the unmodified protein. Individual rates of exchange for the slowly exchanging amides in hen lysozyme (Pedersen *et al.*, 1991; Radford *et al.*, 1992a) and $CM^{6,127}$-lysozyme have been measured by NMR as described in Chapter Three. From these rates a bulk exchange curve was simulated for the slowly exchanging amides in the two proteins. Figure 6.7 shows an overlay of the bulk exchange curve simulated from NMR data with the bulk exchange kinetics measured by ESI-MS for $CM^{6,127}$-lysozyme. There are differences between the simulated and measured curve, particularly at early time points. This is because ESI-MS monitors exchange at all labile sites in a protein rather than only those which are exchanging slowly enough to be measured by NMR and which may be resolved clearly in the 2D NMR spectrum. Hence the ESI-MS kinetic curve starts with a higher number of deuterons, but converges with that obtained by NMR. Clearly, however, there is good agreement for the later time points between the data obtained by both techniques, thus further confirming the validity of the ESI-MS approach for measuring hydrogen exchange in proteins.

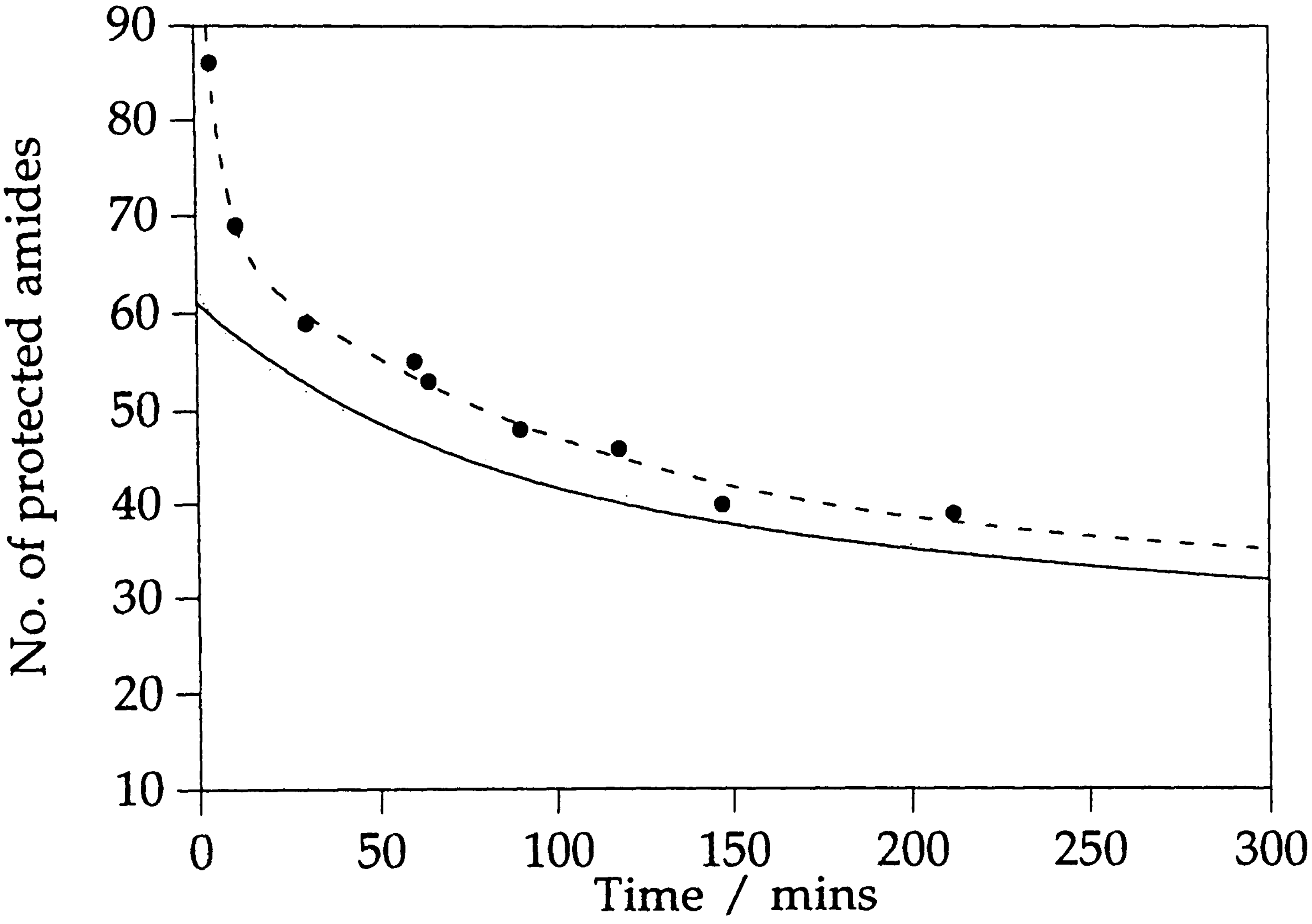


Figure 6.7. Comparison of bulk hydrogen exchange kinetics measured by ESI-MS with curve reconstructed from site specific NMR data measured for 65 slowly exchanging amides for CM$^{6,127}$-lysozyme, pH 5.0, 20°C.

By contrast, $CM^{6,127}$-lysozyme at pH 2.0 is known to be substantially denatured (Chapter 3) and this is reflected in the complete lack of protection against exchange of this state (Figure 6.6a); the observed kinetics agree closely with those calculated from theoretical exchange rates for a random structure of this sequence, accounting for inductive effects of neighbouring side chains (Bai *et al.*, 1993; Connelly *et al.*, 1993).

Protection of the various states of α-lactalbumin is particularly interesting. Again the native state of bovine α-lactalbumin is significantly protected from exchange, although less so than either hen lysozyme or $CM^{6,127}$-lysozyme. This reflects the lower stability to hydrogen exchange of the α-lactalbumins relative to the lysozymes in general, possibly due to the dynamic nature of the native structure of α-lactalbumin in equilibrium with the molten globule under these conditions (Baum *et al.*, 1989). At pH 2.0, α-lactalbumin adopts a molten globule conformation, as seen by CD and NMR (see Chapter 5). The extent of protection against hydrogen exchange was measured for the molten globule states of two different, but highly homologous α-lactalbumins (for sequences see Appendix A). It can be seen clearly from Figure 6.8 that all of these states are more highly protected from exchange than predicted for a random coil, but the number of hydrogens protected is small compared with that in the native proteins. A surprising feature of the exchange profiles for the three proteins is that the number of sites protected in their molten globule states is significantly different. At the earliest possible time of measurement, some 55 to 60 deuterons remain in the two molten globules, but these rapidly exchange such that for the bovine protein, only around 12 trapped deuterons remain after 60 minutes of exchange. By contrast, in the case of the guinea pig protein, some 25 deuterons remain after this time, indicating a much higher level of protection in the molten globule state of GPLA relative to the equivalent state in other $\alpha$-lactalbumins.

To confirm the differences in exchange properties of these three molten globule states, a second experiment was performed in which the two molten globules were monitored simultaneously, since the masses of the proteins are sufficiently different to be clearly resolved from one another in the same mass spectrum. The kinetics measured from these experiments were fully consistent with those measured for each protein individually, confirming the validity of the technique. In Figure 6.8 the exchange curves obtained by ESI-MS are compared with data obtained by bulk exchange or pH jump NMR experiments (Wormald, 1991; Chyan *et al.*, 1993). There is good agreement between the two techniques, indicating that ESI-MS can also be applied to the study not only of hydrogen exchange in native proteins, but also partially folded molten globule states. Protection factors of the most highly protected amides in the molten globule of GPLA are in the range 10–50 (Alexandrescu *et al.*, 1993; Chyan *et al.*, 1993), in contrast with values in the range $10^3$–$10^8$ for native hen lysozyme (Radford *et al.*, 1992a).

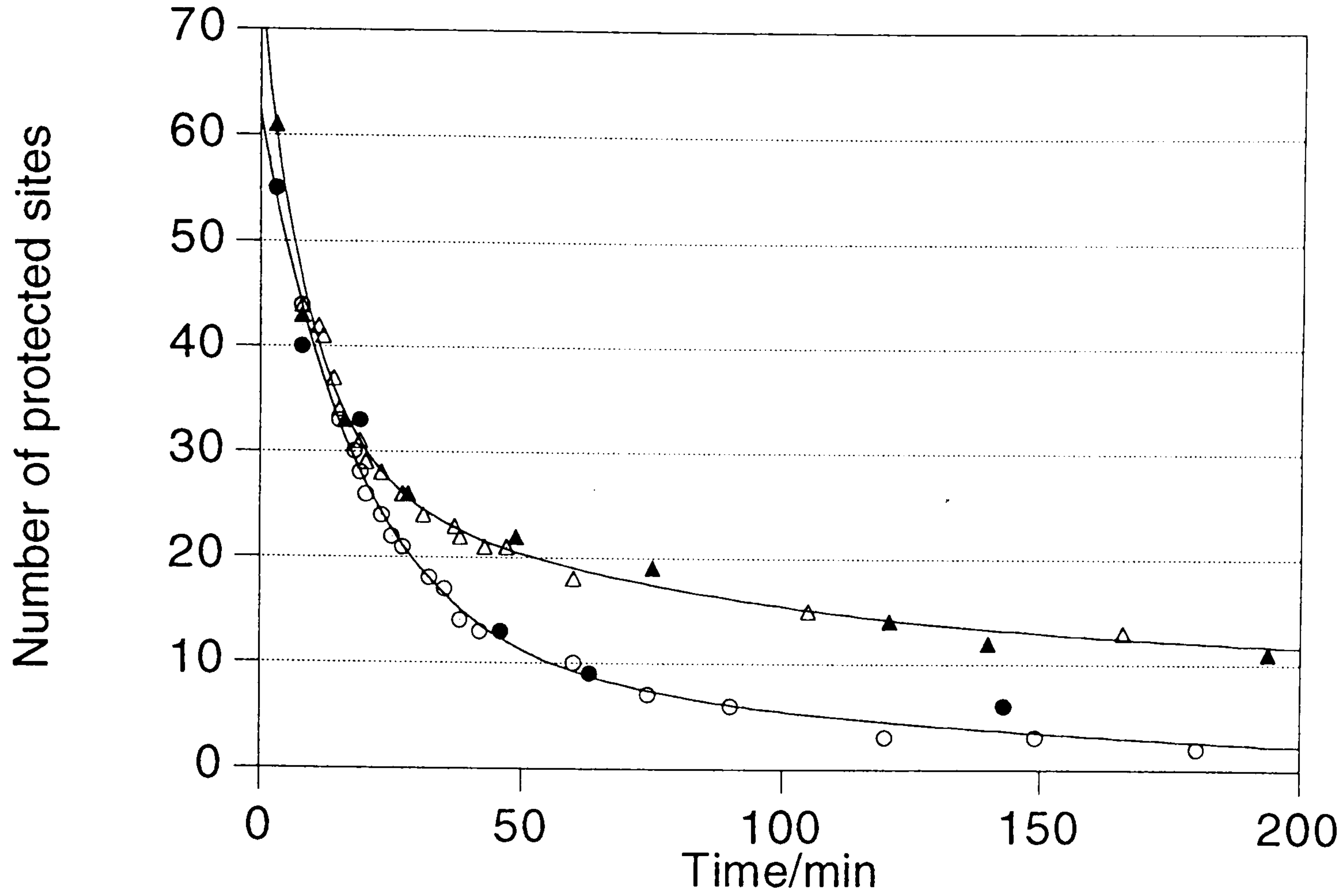


**Figure 6.8.** Comparison of hydrogen exchange kinetics measured by ESI-MS and NMR for molten globule states of $\alpha$-lactalbumin. BLA pH 2.0, 20°C (●) ESI-MS data compared with (○) bulk exchange NMR data (Wormald, 1991). GPLA pH 2.0, 20°C (▲) ESI-MS data compared with (△) bulk exchange NMR data (Chyan *et al.*, 1993).

## *6.4 Interaction of α-lactalbumin with GroEL detected by ESI-MS.*

The vast majority of kinetic refolding studies of proteins have investigated the unassisted folding reactions in which denatured protein is refolded in a buffer which favours the native state. A developing field of research involves study of the interactions between refolding proteins and chaperone proteins which are present within the cell and have been shown to have a role in assisting protein folding and preventing aggregation of folding intermediates (Jaenicke, 1993; Hartl *et al.*, 1994). One such example of a chaperone is GroEL, a large oligomeric protein consisting of 14 identical subunits, each with a mass of ~60 kDa. The structure of this protein has been shown by electron microscopy (Hutchinson *et al.*, 1989; Zwickl *et al.*, 1990) and recently also by X-ray crystallography (Braig *et al.*, 1994) to form a doughnut shape of two stacked rings, each consisting of 7 subunits. The central hole in this protein has a diameter of some 47 Å (Braig *et al.*, 1994), and it is thought that the interior of this hole provides a hydrophobic environment in which the protein can fold correctly without aggregation (Martin, 1991). GroEL has been shown to interact with a variety of protein substrates (Jaenicke, 1993). The structure in a nascent polypeptide chain recognised by the chaperone is unknown, but a number of structural features have been proposed, including α-helices (Landry & Gierasch, 1991; Landry *et al.*, 1992), β-sheets (Schmidt & Buchner, 1992), hydrophobic patches (Hartl *et al.*, 1994) and some molten globule states (Flynn *et al.*, 1994; Hayer-Hartl *et al.*, 1994), although other molten globules have been shown not to bind to GroEL (Okazaki *et al.*, 1994). Recent experiments have been designed to investigate the nature of the bound state of cyclophilin to GroEL using hydrogen exchange in conjunction with NMR (Zahn *et al.*, 1994). The complex was incubated for periods of time to allow amide hydrogen exchange to occur, then the two proteins were separated by chromatography. Protection in the complex was then assessed by measuring proton occupancy in free cyclophilin by NMR. This technique requires that amides protected in the chaperone bound state are also protective against exchange in the native state.

One of the best characterised molten globule states is that of α-lactalbumin. As previously described the native state of the protein consists of α-helical and β-sheet regions, separated into two domains, a tight calcium binding site, and four disulphide bridges which maintain this structure (Acharya *et al.*, 1991). The molten globule state may be induced by a combination of conditions including removal of the $Ca^{2+}$ ligand (Okazaki *et al.*, 1994), low pH (Kuwajima, 1989) or removal of one or more of the disulphide bridges (Ewbank & Creighton, 1991) (Chapter 5). This state retains extensive secondary structure, although tertiary interactions are significantly disrupted, and is thought to resemble the transient intermediate detected during the kinetic refolding of α-lactalbumin (Ikeguchi *et al.*, 1986a) and the homologous hen lysozyme (Dobson *et al.*, 1994). Hydrogen exchange NMR experiments have shown that in the molten globule state the most highly protected amides occur within the α-domain of the protein (Alexandrescu *et al.*, 1993; Chyan *et al.*, 1993). Similar patterns of protection have been observed in the α-domain folding intermediate in lysozyme (Radford *et al.*, 1992b; Miranker *et al.*, 1993). With the extensive knowledge of the nature of the molten globule state of α-lactalbumin, this protein was an ideal candidate for investigation of its interaction with the chaperone GroEL.

The size of the GroEL molecule (800 kDa) precludes detailed study by NMR, since a small protein molecule bound to it would take on its correlation time and become effectively invisible to NMR, although it has been possible to study peptides which are only transiently bound by transferred nOe techniques (Landry & Gierasch, 1991; Landry *et al.*, 1992). The protein concentrations required for NMR are also prohibitive (for a 1 mM sample a 1:1 complex would require 800 mg/ml GroEL). ESI-MS is an ideal technique for these studies, however, since very low protein concentrations are required, and species can readily be distinguished by their differences in mass. As demonstrated in the previous section, by careful control of the conditions employed for mass spectrometry it is possible to retain the calcium ligand within the structure of *holo* α-lactalbumin, and to measure hydrogen exchange in partially folded proteins, so this technique is sufficiently mild to enable studies of this kind. Using the data for hydrogen exchange in the native, molten globule and denatured states of bovine α-lactalbumin described in §6.3, ESI-MS may be used to determine the level of protection in the GroEL bound state of α-lactalbumin, and to compare this with the protection in the three equilibrium states previously described.

Previous experiments have indicated that the molten globule of α-lactalbumin formed by removal of the calcium ligand does not bind to the chaperone GroEL (Okazaki *et al.*, 1994). If the Cys6–Cys120 disulphide bridge in bovine α-lactalbumin is selectively reduced

(Chapter 5), but the free thiols left unblocked this species rearranges by disulphide shuffling to yield an ensemble of molten globule structures with different disulphide pairings (Ewbank & Creighton, 1991). Biochemical experiments have shown that this disulphide shuffled molten globule state of bovine α-lactalbumin (denoted [3SS]) does indeed bind to GroEL; this is thought to be due to the increased exposure of hydrophobic surface area in these states which do not contain a full complement of disulphide bridges (Hayer-Hartl *et al.*, 1994).

Experiments were designed in which the disulphide shuffled α-lactalbumin and GroEL in a 1:1 GroEL:[3SS] BLA complex can be visualised by ESI-MS (these experiments were performed by Drs Michael Groβ and Carol Robinson). It is then possible to measure hydrogen exchange in the GroEL bound α-lactalbumin by diluting a pre-deuterated sample into $H_2O$ buffer and allowing it to exchange for varying periods of time before introduction into the mass spectrometer. The conditions are such that the complex dissociates in the spectrometer, as GroEL itself dissociates readily into monomers under these conditions (Lissin & Hemmingsen, 1993), but the hydrogen exchange information in the α-lactalbumin released into the gas phase is retained. In this manner it is possible to infer from the deuterium occupancy in [3SS] the level of protection in the partially folded state bound in the central cavity of GroEL. The mass spectrum of the GroEl:[3SS] mixture after 27 mins of exchange by a 10-fold dilution into $H_2O$, pH 5.0, 4°C is shown in Figure 6.9a. Two series of peaks can be clearly resolved, with masses corresponding to α-lactalbumin (labelled A) and GroEL monomers (labelled B). Integration of the transformed spectrum (Figure 6.9b) gave a ratio of 16±4:1, consistent with the 1:1 stoichiometry of the GroEL 14mer and [3SS] in the complex.

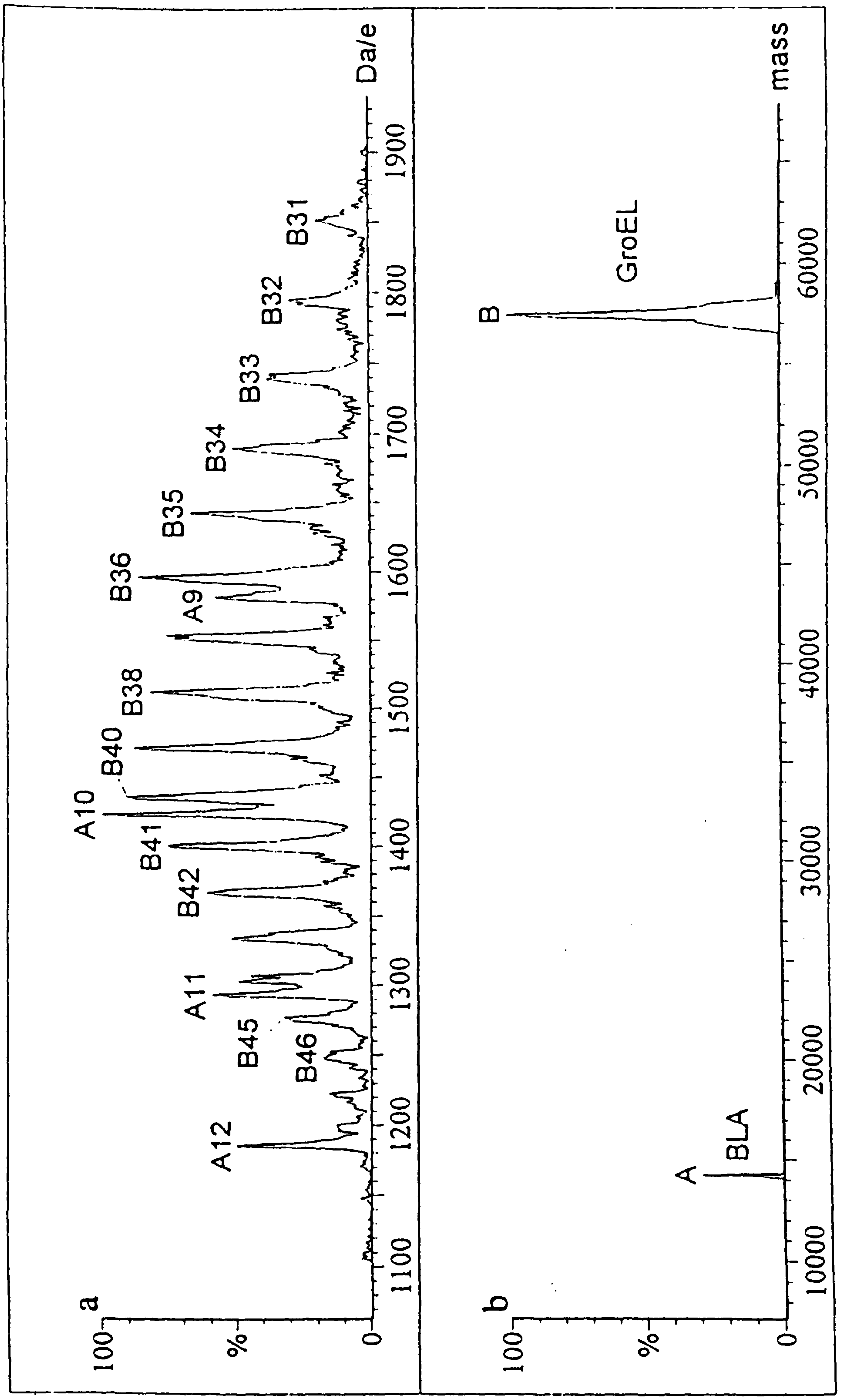


**Figure 6.9.** (a) ESI mass spectrum of GroEl:[3SS] BLA complex, 27 min after initiating exchange at 4°C in water at pH 5; (b) mass transformed spectrum of the complex.

Hydrogen exchange was measured in the GroEL:[3SS] complex using this experimental method. The time course of exchange in the chaperone bound state of [3SS] is shown in Figure 6.10, compared with exchange in the native state of BLA and the molten globule state of $3SS_{cam}$ under identical conditions. Clearly protection in the GroEL:[3SS] complex is higher than that which would be predicted for an unfolded protein but similar to that for $3SS_{cam}$ itself. It is intriguing to note that the level of protection in the complex is slightly but significantly lower than that in the partially folded state of disulphide blocked $3SS_{cam}$. It is difficult to draw conclusions from comparison of the different molten globule states, since the disulphide pairing is likely to affect the structure of the partially folded states to some degree. These data suggest, however, that the molten globule state of $3SS_{cam}$ is more persistently structured, and hence more protective against exchange than the disulphide shuffled species bound to the chaperone. It also suggests that binding of [3SS] is dynamic within the GroEL cavity since binding does not afford any greater protection to the molten globule than in free $3SS_{cam}$. These data are consistent with the conclusions of Zahn *et al.* (1994), that the protection in the chaperone bound state is significantly lower than that in the native state of the protein. Protection factors in GroEL bound cyclophilin were found to be less than $10^3$. The protection factors in molten globule states tend to be low, however, in many cases rarely exceeding 10 (Hughson *et al.*, 1990; Buck *et al.*, 1993; Chyan *et al.*, 1993). This clearly demonstrates the ability of ESI-MS to examine quantitatively a small number of marginally protected deuterons.

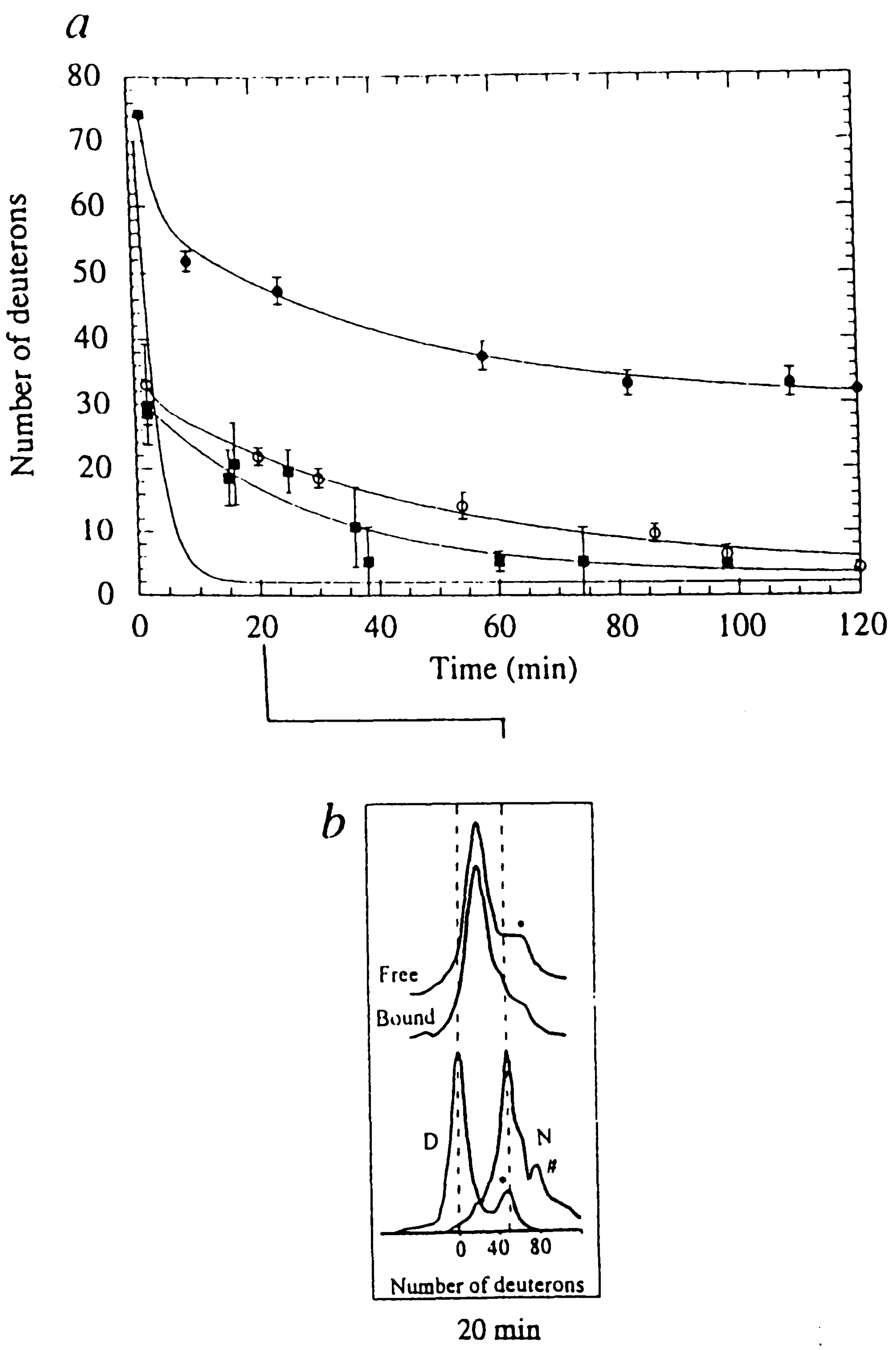


**Figure 6.10.** (a) Kinetic profiles of hydrogen exchange at pH 5.0, 4°C, monitored by ESI-MS. ● *holo* BLA (native state), O *apo* $3SS_{cam}$ BLA (molten globule state), ■ [3SS] BLA in a complex with GroEL; line, simulated curve expected for BLA in a completely unstructured conformation. (b) Representative charge states (+12) for *holo* BLA (N), *apo* $3SS_{cam}$ (Free), [3SS] BLA bound to GroEl (Bound) and in a simulated random coil structure of unfolded BLA (D), 20 mins after the initiation of exchange.

The technique has the added advantage over an indirect NMR approach that protection in the native state is not also required, hence allowing detection of non-native structure. These results suggest that [3SS] BLA binds to GroEL in a conformation which resembles the molten globule state, in accord with previous predictions (Bychkova *et al.*, 1988; Bychkova & Ptitsyn, 1993).

Comparison of the line widths of the peaks arising from the +12 charge state in the mass spectra of the free $3SS_{cam}$ and GroEL bound [3SS] states after 20 minutes of exchange with those of native BLA and fully exchanged $3SS_{cam}$ are shown in Figure 6.10. The peak width is significantly narrower than that which would be predicted for exchange in a range of states from fully protected (N) to fully unprotected (D). There is also remarkable similarity between the widths of the peaks from the free $3SS_{cam}$ molten globule and the GroEL bound states, indicating the similarity of the two states both in the degree of protection and in the conformations sampled by these states. Evidence from previous studies on the molten globule state of α-lactalbumin (Hughson *et al.*, 1990; Alexandrescu *et al.*, 1993; Chyan *et al.*, 1993), and on the transient intermediate formed during the kinetic refolding of hen lysozyme (Radford *et al.*, 1992b) indicates that the protected amides in these states occur in the α-domain of the protein, with the β-domain largely unstructured. On the basis of the results presented here, it is tempting to suggest that the α-domain is also structured in the GroEL bound state. The chaperone in this case recognises and binds to a state which is hydrophobically collapsed, and which contain a significant degree of secondary structure but few tertiary interactions, and which is highly dynamic. This is akin to the state populated within the burst phase (< 1ms) of kinetic refolding of lysozyme *in vitro*. Molten globules which are more highly ordered such as the *apo* form of intact four disulphide α-lactalbumin do not bind to the chaperone (Okazaki *et al.*, 1994), however, presumably since these states resemble late intermediates on the folding pathway. This suggests that GroEL acts at an early stage during protein folding *in vivo*.

The ability to perform experiments of this type opens the way to further investigation of the interaction of chaperones with states of proteins at various points on the folding pathway. In many cases the cofactors Mg and ATP have been required to effect chaperone assisted protein folding. The techniques described here lend themselves to study of the role of these cofactors also, in order to elucidate in further detail the mechanism of chaperone action.

## 6.5 *Future perspectives.*

In this chapter experimental techniques have been described which clearly demonstrate a number of novel applications of ESI-MS to the study of protein folding. In conjunction with hydrogen exchange pulse labelling, it is invaluable for determining the cooperativity of folding events which could only be inferred from NMR. Direct comparisons of the folding characteristics of variant proteins is possible by analysis of mixtures of proteins in the mass spectrometer, providing that the mass difference between each variant is sufficient, and this may be circumvented by the use of isotopically labelled protein. The ESI-MS technique may be controlled carefully to enable the study of hydrogen exchange protection in real time not only in native proteins but also in structurally unstable partially folded states. This has been recently extended to the study of the interactions between a folding polypeptide and chaperone proteins, which has given insight into the nature of the chaperone assisted folding pathways which occur *in vivo*. As further advances in technology and instrumentation are made, allowing protein fragmentation and even higher resolution, ESI-MS may become a tool as essential as NMR for the study of protein folding.

Recent developments in technology have enabled ESI-MS to be coupled with HPLC chromatographic instruments, such that protein fragments separated from peptic digests may be investigated by mass spectrometry. If conditions can be found whereby fragmentation can be achieved without amide hydrogen exchange, then the exact location of trapped deuterons within a protein can be determined. This has been achieved to some extent by peptic digest of cytochrome *c* at low pH where amide exchange is slow, followed by HPLC and ESI-MS of the resulting fragments (Zhang & Smith, 1993). An alternative and much more elegant approach is to fragment the protein in the gas phase, since theoretically there should be no solvent present at this point with which the protein could exchange. The development of FT-ICR (Fourier transform ion cyclotron radiation) technology has now made this possible at least in principle. A protein sample which has been labelled during folding can be sprayed and desolvated from an electrospray interface into a high magnetic field. Since the gaseous protein ions are charged they will precess in this magnetic field at a frequency characteristic of the m/z. Fragmentation may then achieved either by collision with inert gas molecules introduced into the chamber or by causing collision by accelerating the molecules at high field (CID, collision induced dissociation), while leaving the extent of deuterium labelling intact. Accuracy to within 0.2 Da can be achieved using FT-ICR ESI-MS, allowing observation of the natural abundance isotope distribution within a sample of molecules to be measured. This in principle enables each peptide fragment to be identified unambiguously and hence to measure the extent of deuterium incorporation in each fragment. This powerful technique promises to revolutionise the field of mass spectrometry, and has already been used with some success for identifying the cooperativity of hydrogen exchange in lysozyme both in the native state and during kinetic refolding (A.D. Miranker, C.V. Robinson & C.M. Dobson, manuscript in preparation).

# Appendix A

## Sequences of lysozymes and α-lactalbumins.

### *Aligned primary sequences of hen lysozyme (HEWL), bovine (BLA) and guinea pig (GPLA) α-lactalbumins.*

Gaps are left for the purposes of aligning the sequences of lysozymes and α-lactalbumins.

| | 1 | | | | | | | | | 10 | | | | | | | | | | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | K | V | F | G | R | C | E | L | A | A | A | M | K | R | H | G | L | D | N | Y |
| BLA: | E | Q | L | T | K | C | E | V | F | R | E | L | K | - | - | D | L | K | G | Y |
| GPLA: | K | Q | L | T | K | C | A | L | S | H | E | L | N | - | - | D | L | A | G | Y |

| | 21 | | | | | | | | | 30 | | | | | | | | | | 40 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | R | G | Y | S | L | G | N | W | V | C | A | A | K | F | E | S | N | F | N | T |
| BLA: | G | G | V | S | L | P | E | W | V | C | T | T | F | H | T | S | G | Y | D | T |
| GPLA: | R | D | I | T | L | P | E | W | L | C | I | I | F | H | T | S | G | Y | D | T |

| | 41 | | | | | | | | | 50 | | | | | | | | | | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | Q | A | T | N | R | N | T | D | G | S | T | D | Y | G | I | L | Q | I | N | S |
| BLA: | E | A | I | V | Q | N | - | N | D | S | T | E | Y | G | L | F | Q | I | N | N |
| GPLA: | Q | A | I | V | K | N | - | S | D | H | K | E | Y | G | L | F | Q | I | N | D |

| | 61 | | | | | | | | | 70 | | | | | | | | | | 80 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | R | W | W | C | N | D | G | R | T | P | G | S | R | N | L | C | N | I | P | C |
| BLA: | K | I | W | C | K | N | D | Q | D | P | H | S | D | N | I | C | N | I | S | C |
| GPLA: | K | D | F | C | E | S | S | I | T | V | Q | S | R | N | I | C | D | I | S | C |

| | 81 | | | | | | | | | 90 | | | | | | | | | | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | S | A | L | L | S | S | D | I | T | A | S | V | N | C | A | K | K | I | V | S |
| BLA: | D | K | F | L | D | D | D | L | T | D | D | I | M | C | V | K | K | I | L | - |
| GPLA: | D | K | L | L | D | D | D | L | T | D | D | I | M | C | V | K | K | I | L | - |

| | 101 | | | | | | | | | 110 | | | | | | | | | | 120 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | D | G | N | G | M | N | A | W | V | A | W | R | N | R | C | K | G | T | D | V |
| BLA: | D | K | V | G | I | N | Y | W | L | A | H | K | A | L | C | S | E | K | L | D |
| GPLA: | D | I | K | G | I | D | Y | W | L | A | H | K | P | L | C | S | D | K | L | E |

| | 121 | | | | | | | | | 130 |
|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | Q | A | W | I | R | G | C | - | R | L |
| BLA: | Q | - | W | L | - | - | C | E | K | L |
| GPLA: | Q | - | W | Y | - | - | C | E | A | Q |

# Appendix B

## Sequences of variant lysozymes.

### *Aligned primary sequences of hen lysozyme (HEWL), Lady Amherst pheasant (LAPL), Japanese quail (JQL) and human (HL) lysozymes.*

Numbering is according to the sequence of human lysozyme (Imoto *et al.*, 1972).

| | 1 | | | | | | | | | 10 | | | | | | | | | | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | K | V | F | G | R | C | E | L | A | A | A | M | K | R | H | G | L | D | N | Y |
| LAPL: | | | Y | | | | | | | | | | | | L | | | | | |
| JQL: | | | Y | | | | | | | | | | | | | | | | K | |
| HL: | | | | E | | | | | | R | T | L | | | L | | M | | G | |

| | 21 | | | | | | | | | 30 | | | | | | | | | | 40 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | R | G | Y | S | L | G | N | W | V | C | A | A | K | F | E | S | N | F | N | T |
| LAPL: | | | | | | | | | | | | | | | | | | | | |
| JQL: | Q | | | | | | | | | | | | | | | | | | | |
| HL: | | | I | | | A | | | M | | L | | | W | | | G | Y | | |

| | 41 | | | | | | | | | 50 | | | | | | | | | | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | Q | A | T | N | R | N | T | - | D | G | S | T | D | Y | G | I | L | Q | I | N |
| LAPL: | H | | | | | | | | | | | | | | | | | | | |
| JQL: | | | | | | | | | | | | | | | | | | | | |
| HL: | R | | | | Y | | A | G | | R | | | | | | | F | | | |

| | 61 | | | | | | | | | 70 | | | | | | | | | | 80 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | S | R | W | W | C | N | D | G | R | T | P | G | S | R | N | L | C | N | I | P |
| LAPL: | | | | | | | | | | | | | | | | | | H | | |
| JQL: | | | | | | | | | | | | | | | | | | | | |
| HL: | | | Y | | | | | | K | | | | A | V | | A | | H | L | S |

| | 81 | | | | | | | | | 90 | | | | | | | | | | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | C | S | A | L | L | S | S | D | I | T | A | S | V | N | C | A | K | K | I | V |
| LAPL: | | | | | | | | | | | | | | | | | | | | |
| JQL: | | | | | | | | | | | | | | | | | | | | |
| HL: | | | | | | Q | D | N | | A | D | A | | A | | | | R | V | |

| | 101 | | | | | | | | | 110 | | | | | | | | | | 120 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | S | D | G | N | G | M | N | A | W | V | A | W | R | N | R | C | K | G | T | D |
| LAPL: | | | | | | | | | | | | | | | | | | | | |
| JQL: | | | V | H | | | | | | | | | | | | | | | | |
| HL: | R | | P | Q | | I | R | | | | | | | | | | Q | N | R | |

| | 121 | | | | | | | | | 130 |
|---|---|---|---|---|---|---|---|---|---|---|
| HEWL: | V | Q | A | W | I | R | G | C | R | L |
| LAPL: | | N | | | | | | | | |
| JQL: | | N | | | | | | | | |
| HL: | | R | Q | Y | V | Q | | | G | V |

# Appendix C

# Kinetic refolding of hen lysozyme into varying final concentrations of guanidinium chloride.

## *C.1. Introduction*

In Chapter Four, the kinetic refolding of a derivative of hen lysozyme which lacked one of the four native disulphide bridges was described. This appendix describes a preliminary investigation of an alternative means of destabilising the native state of lysozyme, namely the addition of the chemical denaturant guanidinium chloride, to determine its effect on the refolding characteristics of the protein. We have shown (Chapter 4) that kinetic refolding of the three disulphide derivative proceeds at a similar overall rate to that of the unmodified protein. The pathway of folding is somewhat altered, however, since the kinetic refolding intermediate comprising the $\alpha$-domain of the protein is no longer populated during the refolding of CM$^{6,127}$-lysozyme.

Guanidinium chloride is known to cause destabilisation of proteins and denaturation at high concentration. This has been postulated to be a combination of the binding of denaturant molecules to the peptide backbone and the change in viscosity or dielectric constant of the solvent (Tanford, 1970; Pace, 1986). This section describes refolding experiments which were performed including varying concentrations of

GuHCl in the refolding buffer to investigate any change in mechanism caused by destabilising the protein in this manner.

## C.2. *Methods*

The refolding of lysozyme from the GuHCl denatured state into a final concentration of 0.55 M GuHCl, pH 5.5 has been extensively studied in this laboratory, as outlined in Chapter 4 and references therein. For simplicity, these conditions will hereinafter be referred to as the 'standard conditions' for refolding. The object of these experiments was to determine what effect varying the final concentration of denaturant in the refolding buffer has on the refolding characteristics of the protein. This was investigated initially by stopped flow circular dichroism, and then a number of conditions were chosen for more extensive study by hydrogen exchange pulse labelling.

Stopped flow circular dichroism experiments were performed as described in Chapter Two, but with GuHCl added to the acetate refolding buffer to vary the final refolding conditions. For pulsed hydrogen exchange labelling experiments, hen lysozyme which had been deuterated previously at all exchangeable amide sites as described in Chapter 2, was denatured in 6 M GuDCl. Refolding was initiated by a 5-fold dilution into 20 mM acetate buffer containing varying concentrations of GuHCl. All other conditions were maintained identical to those described in Chapter Two, i.e. pulse labelling for 8.3 ms at pH 9.5, followed by quenching to pH 3.8.

## C.3. *Circular dichroism in the near-UV*

Figure C.1 shows kinetic traces for refolding of lysozyme under varying final denaturant concentrations, monitored by circular dichroism in the near-UV region at 289 nm. At this wavelength signal arises from the interactions of aromatic chromophores and hence monitors formation of fixed tertiary structure. It is clear that as the concentration of GuHCl in the refolding buffer is increased, refolding is slowed. In the region of concentrations around 'standard conditions', refolding is complete within 1s, but as the concentration of GuHCl is increased beyond 1 M, folding is slowed dramatically, being incomplete in 1.9 M GuHCl even after 10 seconds. This indicates that, just as the native state of the protein is destabilised relative to the unfolded state, so too is the transition state for the rate determining step, giving a higher activation energy and hence a slower rate for the refolding reaction, as predicted by the Arrhenius relationship.

As in the case of refolding under 'standard conditions', the kinetics of refolding could be fitted in all cases adequately to a single exponential decay, suggesting that any intermediates formed during refolding do not involve formation of regions of static tertiary structure. The variation in the fitted rate constant is plotted as a function of GuHCl concentration in Figure C.2. Below around 1.0 M final [GuHCl], the rate does not appear to be strongly dependent on denaturant concentration, but above this concentration, folding is significantly retarded as the amount of denaturant present slows formation of the native structure.

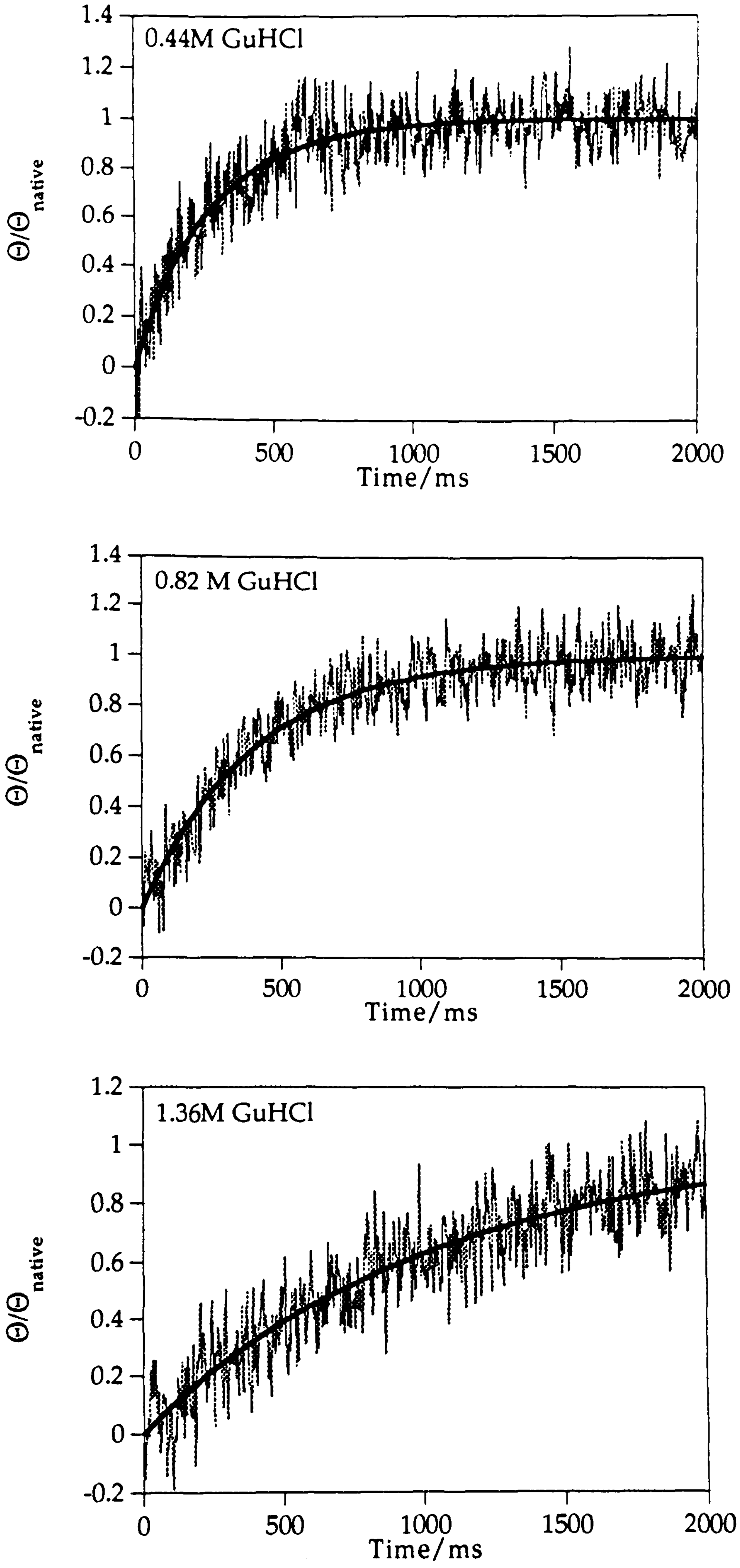


**Figure C.1.** Representative kinetic traces for refolding of hen lysozyme into GuHCl concentrations of 0.44 M, 0.82 M and 1.36 M monitored by circular dichroism in the near-UV at 289 nm.

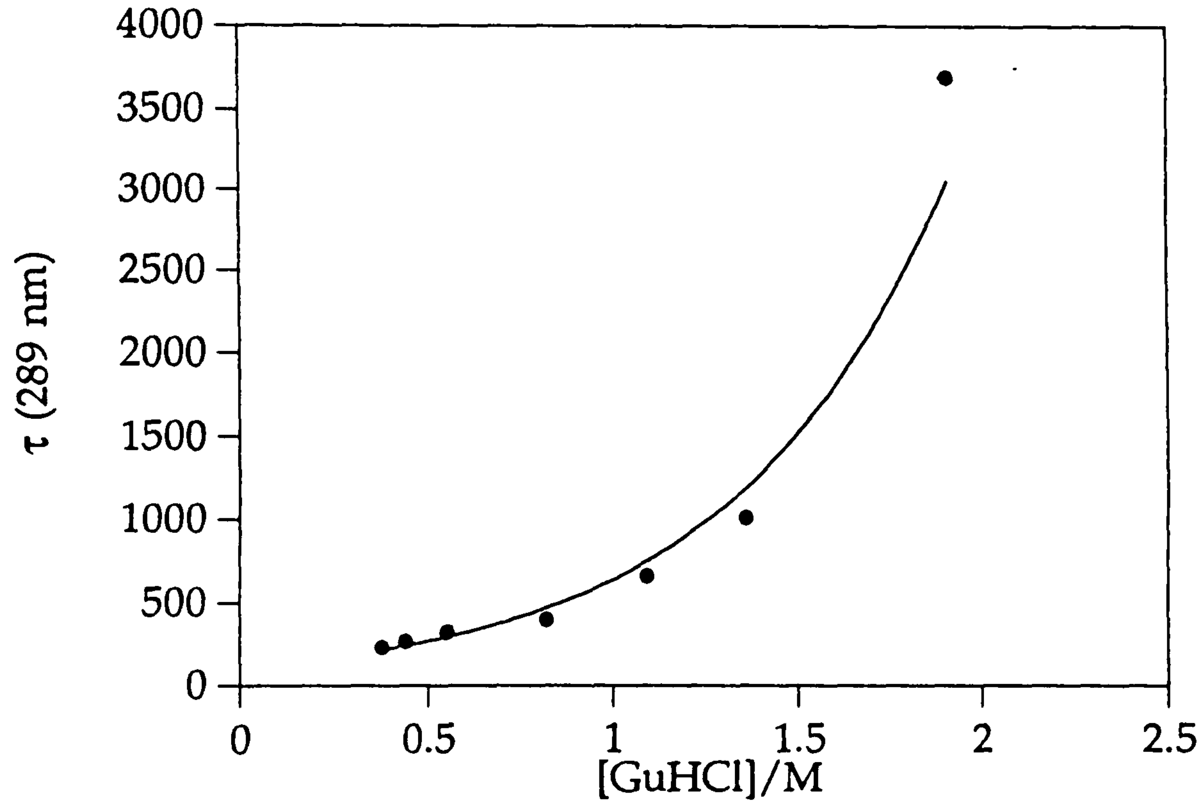


**Figure C.2.** Plot of rate of refolding measured by CD at 289 nm *versus* concentration of GuHCl.

## C.4. *Circular dichroism in the far-UV*

Folding of lysozyme under 'standard conditions' as monitored by far-UV CD exhibits three kinetic phases. Within the dead time of measurement, 80% of the native ellipticity at 225 nm is achieved, indicating rapid formation of a significant complement of secondary structure. This burst phase is followed by production of excess ellipticity at this wavelength with a time constant of 50 ms, and has been attributed to formation of the α-domain intermediate (Radford *et al.*, 1992b). The explanation for this excess ellipticity is unclear, but has been attributed to a positive contribution at this wavelength from disulphide bridges or from aromatic residues which are structured in the native state, but which are not in fixed orientations in this intermediate species. The third phase is the slower recovery of ellipticity characteristic of the native state with a time constant of 200 ms. Thus we have a very rapid collapse to a molten globule like state, followed by formation of the α-domain intermediate and then rearrangement to the native state as tertiary contacts are made. The effect of GuHCl concentration on these three phases is investigated here.

Figure C.3 shows kinetic traces of the refolding of lysozyme into varying final concentrations of GuHCl. At concentrations below 1.0 M GuHCl, as in the case of the CD signal in the near-UV, the kinetic curves appear similar for all three phases of refolding. Clearly, however, the overall rate of formation of the native structure is significantly slowed as the final concentration of denaturant in the refolding buffer is increased. Under 'standard conditions', folding is virtually complete within one second, whereas refolding into 1.9 M GuHCl the native complement of secondary structure is not fully regained even after 5 seconds.

Perhaps the most striking feature, however, is that as the GuHCl concentration is increased, the amplitude of the fast phase, corresponding to formation of the α-domain, diminishes such that above 1.0 M GuHCl, no overshoot can be detected, and at 1.9 M GuHCl, the fast phase is abolished completely. The amplitude of the dead time phase similarly decreases with increased denaturant concentration, suggesting that formation of a collapsed state with significant secondary structure is also influenced by the presence of GuHCl, although this may be partially explained by the effect of the intrinsic dichroism of the denaturant.

The fits of the kinetic data to a sum of two exponentials are tabulated in Table C.1. It can be clearly seen that although the rate of the fast phase is not significantly affected by GuHCl concentration, the amplitude of this phase decreases markedly with increased denaturant. The slow phase, however, is significantly affected. Plots of $\log_{10}$(rate constant) *versus* concentration of GuHCl for the slow phase observed by CD in the far- and the near-UV regions are shown in Figure C.4. These plots show good linearity over the whole range of denaturant concentrations studied and the plots from the data obtained at both wavelengths are parallel, suggesting that the two measurements are monitoring the same transition, or at least two transitions with fortuitously identical activation energies. Complete thermodynamic of these folding processes is possible using kinetic and equilibrium data for the unfolding transition at higher denaturant concentrations (Matouschek *et al.*, 1990) and references therein. Experiments are currently underway in this laboratory to complete this task.

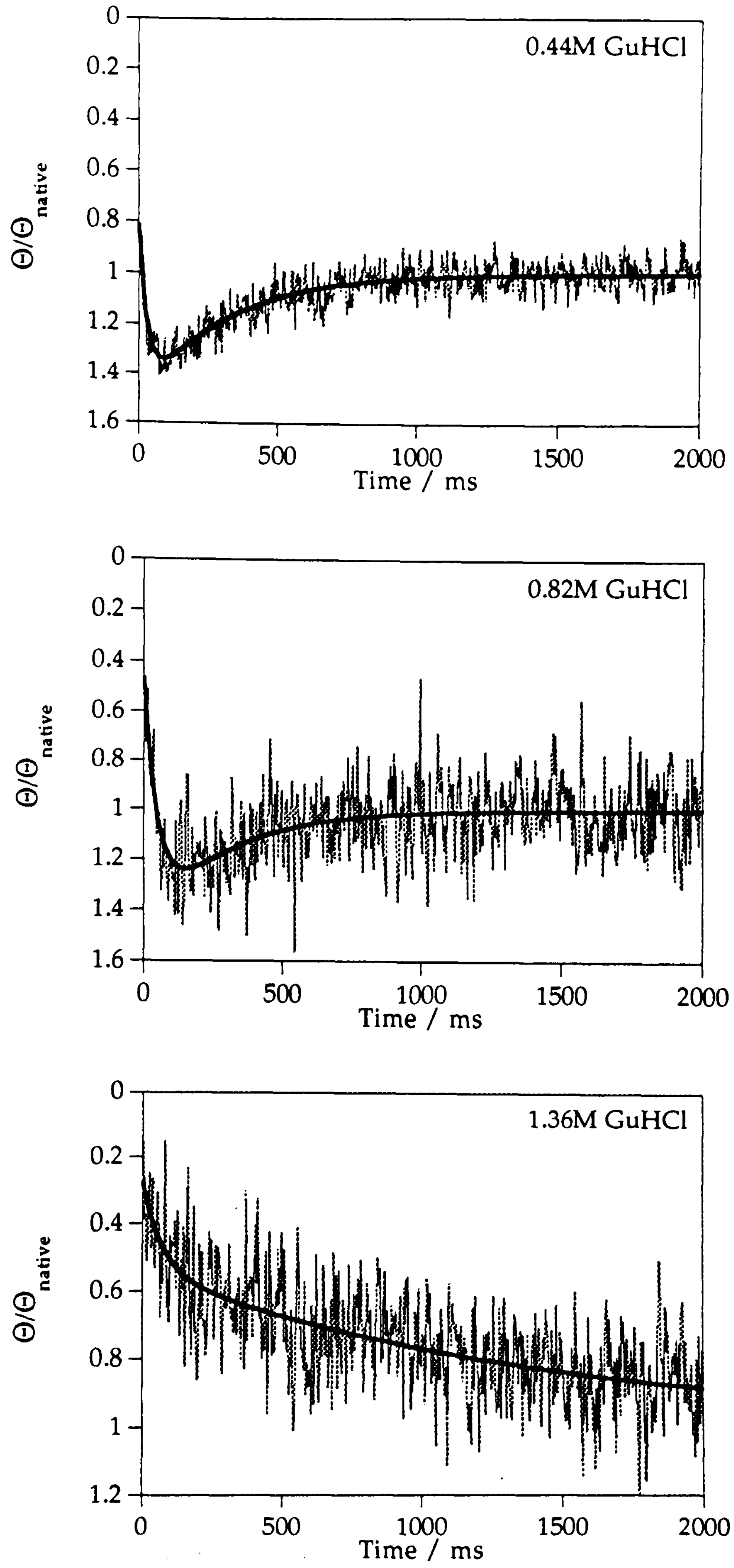


**Figure C.3.** Representative kinetic traces for refolding of hen lysozyme into GuHCl concentrations of 0.44 M, 0.82 M and 1.36 M monitored by circular dichroism in the far-UV at 225 nm.

| 225 nm | | | | | | 289 nm |
|---|---|---|---|---|---|---|
| [GuHCl] M | Amp Burst | Amp. Fast ($\tau_1$) | Fast rate $\tau_1$ | Amp. Slow ($\tau_2$) | Slow rate $\tau_2$ | Time constant |
| 0.44 | 83 % | 69 % | 31 ms | -52 % | 299±30 ms | 271±35 ms |
| 0.55 | 86 % | 60 % | 27 ms | -46 % | 370±35 ms | 325±40 ms |
| 0.82 | 82 % | 69 % | 50 ms | -51 % | 382±35 ms | 405±53 ms |
| 1.09 | 34 % | 39 % | 31 ms | +27 % | 412±40 ms | 665±80 ms |
| 1.36 | 28 % | 26 % | 69 ms | +46 % | 1436±150 | 1010±200 ms |

**Table C.1** Rates of refolding of hen lysozyme into varying concentrations of GuHCl monitored by CD in the far- and near-UV regions.

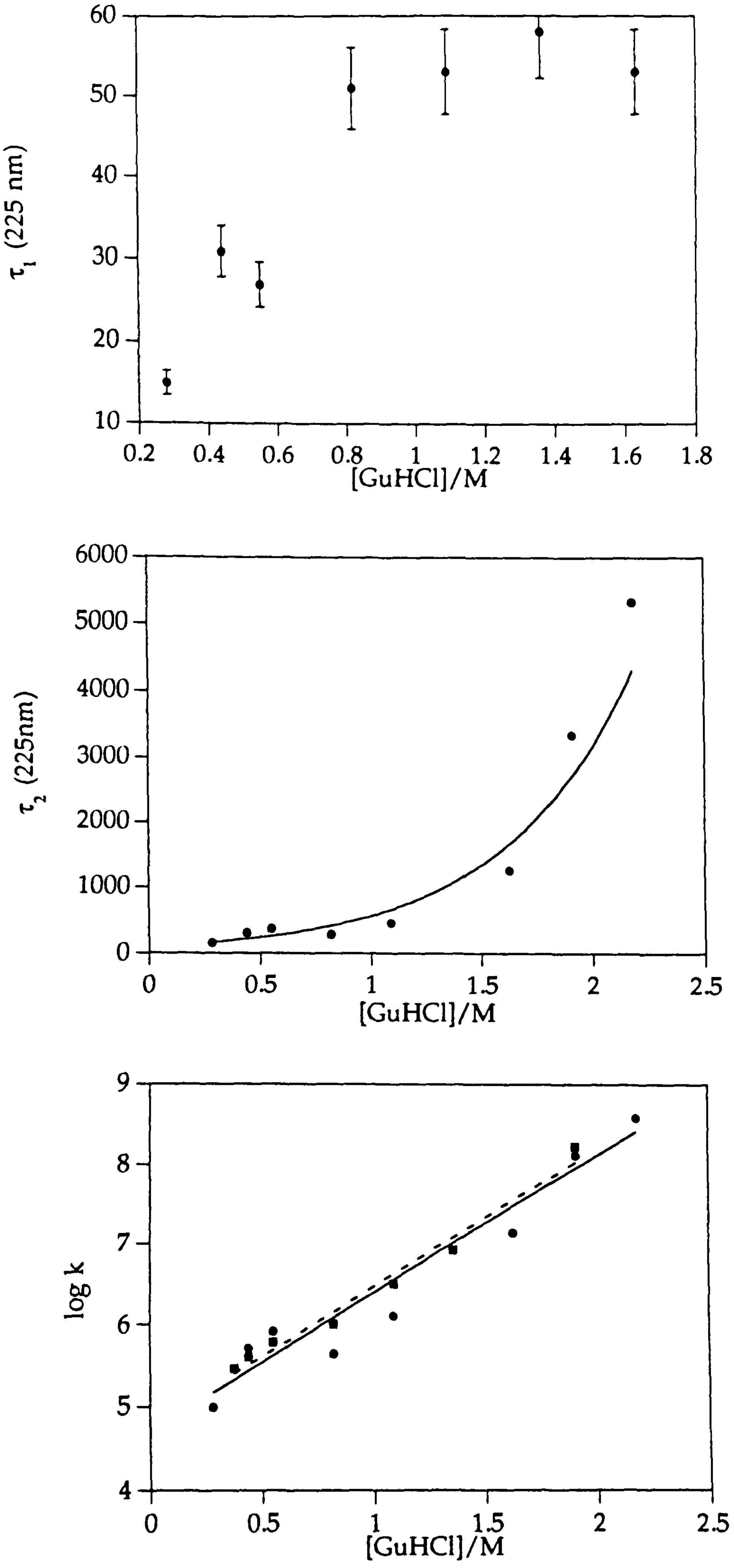


**Figure C.4.** Plot of rate of refolding measured by CD at 225 nm. (a) Fast phase, (b) slow phase. (c) Plot of $\log_{10}$(rate constant) *versus* GuHCl for the near-UV (——) and slow phase of far-UV CD (- - - -) data.

These results suggest that the pathway of folding of lysozyme is altered depending on the concentration of denaturant present in the refolding mixture. At low concentrations of GuHCl, the well characterised folding properties are observed, with buildup of a detectable intermediate species clearly seen in the form of an overshoot in the CD signal at 225 nm. As the final denaturant concentration is increased, however, the population of this species decreases and folding appears to be a much more cooperative event until at 1.91 M GuHCl, this partially folded state is no longer detected. This suggests that formation of the α-domain intermediate is perhaps a non-productive step in the refolding of lysozyme, but is instead a kinetically trapped species which must rearrange in some way to form the native state of the protein. At increased denaturant concentrations, this state is destabilised to such an extent that it does not accumulate as a detectable intermediate, but instead folding proceeds *via* other pathways. This behaviour is very similar to that observed for the folding of CM$^{6,127}$-lysozyme, in which the removal of one of the four native disulphide bridges prevents formation of a detectable α-domain intermediate, as described in Chapter 4. By contrast to the folding of CM$^{6,127}$-lysozyme, however, the overall rate of folding is significantly slowed as the concentration of GuHCl is increased, reflecting the increased activation energy of the process as a result of destabilising the entire protein structure in this manner.

Interestingly, in all cases, both folding of lysozyme into varying concentrations of GuHCl and refolding of the three-disulphide derivative, the burst phase is apparently unaffected by these changes in conditions. This clearly indicates that although the formation of metastable intermediates and of the native state itself is significantly affected by the changes in stability, initial hydrophobic collapse to what is thought to be a molten globule like state with significant secondary structure but no fixed tertiary interactions, is almost completely independent of these changes.

## *C.5. Refolding monitored by hydrogen exchange pulse labelling*

The use of pulsed hydrogen exchange labelling experiments to monitor refolding of lysozyme enables determination of the order of formation of specific regions of secondary structure by measuring the rate at which specific backbone amides become protective against hydrogen exchange as a result of hydrogen bond formation or burial from solvent (Pedersen *et al.*, 1991). By this method it is possible to construct detailed kinetic pathways for the folding of proteins. In the case of lysozyme, the pathway of refolding under 'standard conditions' has been investigated in great depth by optical methods and more specfically with pulse labelling (Radford *et al.*, 1991; Dobson *et al.*, 1994).

Under the 'standard conditions' for refolding of lysozyme, pulse labelling experiments have revealed a number of very interesting features. Firstly the α-domain of the protein, comprising the four α-helices, forms protective structure significantly faster than the remainder of the protein. There is also a very rapid phase which, in a small proportion of the population of molecules, affords complete protection to all of the amides simultaneously. This suggests alternative folding pathways, a hypothesis which has been confirmed by the use of pulse intensity experiments and hydrogen exchange pulse labelling in conjunction with ESI-MS (see Chapters 4 and 6).

Pulse labelling experiments were carried out refolding lysozyme into the following final concentrations of denaturant: 0.27 M (control for refolding of CM$^{6,127}$-lysozyme), 0.55 M ('standard conditions'), 1.09 M, 1.45 M and 1.91 M GuHCl (Figure C.5). In the case of the former three concentrations of GuHCl, no significant differences in the refolding characteristics of the protein were observed. Apart from slight differences in the rate constants for the steps, the overall pathway of folding seems to be unchanged at these relatively low concentrations of denaturant. By contrast to this, as the final denaturant concentration in the refolding buffer is increased further, several changes in the folding properties are observed. Refolding into 1.45 M GuHCl, although there is still evidence of the independent nature of the folding domains, this distinction is far less clear than at lower concentrations of GuHCl. From the evidence of this and the evolution of CD signal in the far-UV region, this would be consistent with the α-domain intermediate becoming unstable in higher GuHCl concentrations and hence a decreased proportion of molecules folding via this pathway. This is consistent with the fact that the overshoot in the far-UV signal decreases until it is completely absent refolding into high concentrations of GuHCl. At 1.9 M GuHCl, refolding appears to be completely cooperative, with no distinction between the rates of protection of amides in the α- and β-domains.

These results clearly demonstrate that as the final concentration of GuHCl in the refolding buffer is increased, significant changes in the folding properties of lysozyme occur. The transient kinetic intermediate in which the α-domain of the molecule becomes protective against exchange in the absence of persistent structure in the remainder of the molecule ceases to be observed at higher concentrations of denaturant. The overall folding of the protein becomes increasingly cooperative in a manner similar to that observed during refolding of the three-disulphide derivative $CM^{6,127}$-lysozyme (Chapter 4), although the overall refolding rate of refolding is significantly reduced. This suggests that removal of the Cys6–Cys127 disulphide bridge specifically destabilises the α-domain of the protein during refolding but does not significantly alter the activation energy of the rate determining step. By contrast, refolding into high concentrations of denaturant destabilises the entire protein structure and hence causes refolding of lysozyme to occur much more cooperatively and more slowly. Further investigation of these effects and other means of changing the stability of proteins, such as increased temperature, is required to draw any further conclusions.

# Bibliography

Acharya, K. R., Ren, J., Stuart, D. I., Phillips, D. C. & Fenna, R. E. (1991) *J. Mol. Biol.* **221**, 571-581.

Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M. & Phillips, D. C. (1989) *J. Mol. Biol.* **208**, 99-127.

Alexandrescu, A. T., Abeygunawardana, C. & Shortle, D. (1994) *Biochemistry* **33**, 1063-1072.

Alexandrescu, A. T., Broadhurst, R. W., Wormald, C., Chyan, C.-L., Baum, J. & Dobson, C. M. (1992) *Eur. J. Biochem.* **210**, 699-709.

Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J. & Dobson, C. M. (1993) *Biochemistry* **32**, 1707-1718.

Alexandrescu, A. T., Ng, Y.-L. & Dobson, C. M. (1994) *J. Mol. Biol.* **235**, 587-599.

Anderson, W. L. & Wetlaufer, D. B. (1976) *J. Biol. Chem.* **251**, 3147-3153.

Anfinsen, C. B. & Scheraga, H. A. (1975) *Adv. Protein Chem.* **29**, 205-300.

Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976) *J.Chem.Phys.* **64**, 2229-2246.

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993) *Proteins: Struct. Funct. Genet.* **17**, 75-86.

Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol.* **3**, 84-91.

Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989) *Biochemistry* **28**, 7-13.

Bax, A. & Freeman, R. (1981) *J. Magn. Reson.* **44**, 542-561.

Betz, S. F. (1993) *Protein Science* **2**, 1551-1558.

Betz, S. F., Raleigh, D. P. & DeGrado, W. F. (1993) *Curr. Opin. Struct. Biol.* **3**, 601-610.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965) *Nature* **206**, 757-761.

Bolin, K. A. (1994) D. Phil. Thesis, University of Oxford.

Bolotina, I. A. (1987) *Mol. Bio. (USSR) English Translation* **21**, 1328-1337.

Bolotina, I. A. & Lugauskas, V. Y. (1985) *Mol. Bio. (USSR) English Translation* **19**, 1154-1166.

Bradbury, J. H. & Norton, R. S. (1975) *Eur. J. Biochem.* **53**, 387-396.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994) *Nature* **371**, 578-586.

Brems, D. N. & Havel, H. A. (1989) *Proteins: Struct. Funct. Genet.* **5**, 93-95.

Brew, K., Castellino, F. J., Vanaman, T. C. & Hill, R. L. (1970) *J. Biol. Chem.* **245**, 4570-4582.

Buck, M., Radford, S. E. & Dobson, C. M. (1993) *Biochemistry* **32**, 669-678.

Bundi, A. & Wüthrich, K. (1978) *Biopolymers* **18**, 285-287.

Bychkova, V. E., Pain, R. H. & Ptitsyn, O. B. (1988) *FEBS Lett.* **238**, 231-234.

Bychkova, V. E. & Ptitsyn, O. B. (1993) *Chemtracts-Biochem. Molec. Biol.* **4**, 133-163.

Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. (1990) *Nature* **346**, 488-490.

Chaffotte, A., Guillou, Y. & Goldberg, M. E. (1992) *Biochemistry* **31**, 9694-9702.

Chelvanayagam, G., Reich, Z., Bringas, R. & Argos, P. (1992) *J. Mol. Biol.* **227**, 901-916.

Chowdhury, S. K., Katta, V. & Chait, B. T. (1990) *J. Amer. Chem. Soc.* **112**, 9012.

Christensen, H. & Pain, R. H. (1991) *Eur. Biophys. J.* **19**, 221-229.

Christensen, J. J., Hansen, L. D. & Izatt, R. M. (1976) *Handbook of Proton Ionization Heats*, Wiley, New York.

Chyan, C.-L., Wormald, C., Dobson, C. M., Evans, P. A. & Baum, J. (1993) *Biochemistry* **32**, 5681-5691.

Clarke, J. & Fersht, A. R. (1993) *Biochemistry* **32**, 4322-4329.

Clarke, J., Hounslow, A. M., Bycroft, M. & Fersht, A. R. (1994) *Proc. Natl. Acad. Sci.* **90**, 9837-9841.

Clore, G. M., Kay, L. E., Bax, A. & Gronenborn, A. M. (1991) *Biochemistry* **30**, 12-18.

Collins, K. D. & Washabaugh, M. W. (1985) *Q. Rev. Biophys.* **18**, 323-422.

Connelly, G. P., Bai, Y., Jeng, M.-F. & Englander, S. W. (1993) *Proteins: Struct. Funct. Genet.* **17**, 87-92.

Cooper, A., Eyles, S. J., Radford, S. E. & Dobson, C. M. (1992) *J. Mol. Biol.* **225**, 939-943.

Creighton, T. E. (1974) *J. Mol. Biol.* **87**, 579-602.

Creighton, T. E. (1975) *J. Mol. Biol.* **96**, 767-776.

Creighton, T. E. (1986) *Methods Enzymol.* **131**, 83-106.

Creighton, T. E. (1988) *Bioessays* **8**, 57-63.

Creighton, T. E. (1992a) *Science* **256**, 111-112.

Creighton, T. E. (1992b) *Protein Folding*, W.H. Freeman & Co., New York.

Darby, N. J., van Mierlo, C. P. M. & Creighton, T. E. (1991) *FEBS Lett.* **279,** 61-64.

Delepierre, M., Dobson, C. M., Karplus, M., Poulsen, F. M., States, D. J. & Wedin, R. E. (1987) *J. Mol. Biol.* **197,** 111-130.

Denton, M. E., Rothwarf, D. M. & Scheraga, H. A. (1994) *Biochemistry* **33,** 11225-11236.

Denton, M. E. & Scheraga, H. A. (1991) *J. Protein Chemistry* **10,** 213-232.

Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993) *Proc. Natl. Acad. Sci.* **90,** 1942-1946.

Dobson, C. M. (1992) *Curr. Opin. Struct. Biol.* **2,** 6-12.

Dobson, C. M. (1994) *Current Biology* **4,** 636-640.

Dobson, C. M. & Evans, P. A. (1984) *Biochemistry* **23,** 4267-4270.

Dobson, C. M., Evans, P. A. & Radford, S. E. (1994) *Trends Biochem. Sci.* **19,** 31-37.

Doig, A. J. & Williams, D. H. (1991) *J. Mol. Biol.* **217,** 389-398.

Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y. & Ptitsyn, O. B. (1981) *FEBS Lett.* **136,** 311-315.

Dolgikh, D. A., Kolomiets, A. P., Bolotina, I. A. & Ptitsyn, O. B. (1984) *FEBS Lett.* **165,** 88-92.

Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A. & Wright, P. E. (1992) *J. Mol. Biol.* **226,** 819-835.

Elöve, G. A., Chaffotte, A. F., Roder, H. & Goldberg, M. E. (1992) *Biochemistry* **31,** 6876-6883.

Englander, S. W., Calhoun, D. B., Englander, J. J., Kallenbach, N. R., Leim, R. H., Malin, E. L., Mandal, C. & Rogero, J. R. (1980) *Biophys. J.* **32,** 577-589.

Englander, S. W. & Kallenbach, N. R. (1984) *Q. Rev. Biophys.* **16**, 521-655.

Eriksson, A. E., Baase, W. A., Wozniak, J. A. & Matthews, B. W. (1992) *Nature* **355**, 371-373.

Evans, P. A. & Radford, S. E. (1994) *Curr. Opin. Struct. Biol.* **4**, 100-106.

Ewbank, J. J. & Creighton, T. E. (1991) *Nature* **350**, 518-520.

Ewbank, J. J. & Creighton, T. E. (1993a) *Biochemistry* **32**, 3677-3693.

Ewbank, J. J. & Creighton, T. E. (1993b) *Biochemistry* **32**, 3694-3707.

Eyles, S. J., Radford, S. E., Robinson, C. V. & Dobson, C. M. (1994) *Biochemistry* **33**, 13038-13048.

Fan, P., Bracken, C. & Baum, J. (1993) *Biochemistry* **32**, 1573-1582.

Feng, R. & Konishi, Y. (1992) *Anal. Chem.* **64**, 2090-2095.

Feng, Y., Sligar, S. G. & Wand, A. J. (1994) *Nature Struct. Biol.* **1**, 30-35.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1989) *Science* **246**, 64-71.

Fersht, A. & Winter, G. (1992) *Trends Biochem. Sci.* **17**, 292-294.

Fersht, A. R. (1993) *FEBS Lett.* **325**, 5-16.

Fersht, A. R. & Serrano, L. (1993) *Curr. Opin. Struct. Biol.* **3**, 75-83.

Flory, P. J. (1956) *J. Amer. Chem. Soc.* **78**, 5222-5235.

Flynn, G. C., Beckers, C. J. M., Baase, W. A. & Dahlquist, F. W. (1994) *Proc. Natl. Acad. Sci.* **90**, 10826-10830.

Freedman, R. B. (1992) In *Protein Folding* (Creighton, T. E., Ed.), pp 457-541, W.H. Freeman, New York.

Ganem, B., Li, Y.-T. & Henion, J. D. (1991) *J. Am. Chem. Soc.* **113**, 7818-7819.

Gething, M.-J. & Sambrook, J. (1992) *Nature* **355**, 33-45.

Goodman, E. M. & Kim, P. S. (1989) *Biochemistry* **28**, 4343-4347.

Goto, Y., Calciano, L. J. & Fink, A. L. (1990) *Proc. Natl. Acad. Sci.* **87**, 573-577.

Goux, W. J. & Hooker, T. M., Jr. (1980) *Biopolymers* **19**, 2191-2208.

Haezebrouck, P., Joniau, M., Van Dael, H., Hooke, S. D., Woodruff, N. D. & Dobson, C. M. (1995) *J. Mol. Biol.* **246**, 382-387.

Harding, M. M., Williams, D. H. & Woolfson, D. N. (1991) *Biochemistry* **30**, 3120-3128.

Hartl, F. U., Hlodan, R. & Langer, T. (1994) *Trends Biochem. Sci.* **19**, 20-25.

Harushima, Y. & Sugai, S. (1989) *Biochemistry* **28**, 8568-8576.

Hayer-Hartl, M. K., Ewbank, J. J., Creighton, T. E. & Hartl, F. U. (1994) *EMBO J.* **13**, 3192-3202.

Heinrikson, R. L. & Meredith, S. C. (1984) *Anal. Biochem.* **136**, 65-74.

Hendrick, J. P. & Hartl, F. U. (1993) *Annu. Rev. Biochem.* **62**, 349-384.

Hill, C. P., Johnston, N. L. & Cohen, R. E. (1993) *Proc. Natl. Acad. Sci.* **90**, 4136-4140.

Hill, R. L. & Brew, K. (1975) *Adv. Enzymol.* **43**, 411-490.

Hooke, S. D., Eyles, S. J., Robinson, C. V., Radford, S. E. & Dobson, C. M. (1995) *J. Amer. Chem. Soc.* in press.

Hooke, S. D., Radford, S. E. & Dobson, C. M. (1994) *Biochemistry* **33**, 5867-5876.

Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990) *Science* **249**, 1544-1548.

Hutchinson, E. G., Tichelaar, W., Hofhaus, G., Weiss, H. & Leonard, K. R. (1989) *EMBO J.* **8**, 1485-1490.

Hvidt, A. & Nielsen, S. O. (1966) *Adv. Protein Chem.* **21**, 287-386.

Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. (1986a) *Biochemistry* **25**, 6965-6972.

Ikeguchi, M., Kuwajima, K. & Sugai, S. (1986b) *J. Biochem.* **99**, 1191-1201.

Ikeguchi, M., Sugai, S., Fujino, M., Sugawara, T. & Kuwajima, K. (1992) *Biochemistry* **31**, 12695-12700.

Imoto, T., Forster, L. S., Rupley, J. A. & Tanaka, F. (1971) *Proc. Natl. Acad. Sci.* **69**, 1151-1155.

Imoto, T., Johnson, L. N., North, A. C. T., Philips, D. C. & Rupley, J. A. (1972) In *The Enzymes* (Boyer, P. D., Ed.), vol. 7, pp 665-864, Academic Press, New York.

Imoto, T. & Rupley, J. A. (1973) *J. Mol. Biol.* **80**, 657-667.

Inaka, K., Taniyama, Y., Kikuchi, M., Morikawa, K. & Matsushima, M. (1991) *J. Biol. Chem.* **266**, 12599-12603.

Itzhaki, L. S., Evans, P. A., Dobson, C. M. & Radford, S. E. (1994) *Biochemistry* **33**, 5212-5220.

Iyer, K. S. & Klee, W. A. (1973) *J. Biol. Chem.* **248**, 707-710.

Jaenicke, R. (1993) *Curr. Opin. Struct. Biol.* **3**, 104-112.

Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979) *J. Chem. Phys.* **71**, 4546.

Jencks, W. P. & Regenstein, J. (1975) *Handb. Biochem. Mol. Biol.* **1**, 305-351.

Jeng, M.-F. & Englander, S. W. (1991) *J. Mol. Biol.* **221**, 1045-1061.

Jeng, M.-F., Englander, S. W., Elöve, G. A., Wand, J. A. & Roder, H. (1990) *Biochemistry* **29**, 10433-10437.

Jennings, P. A. & Wright, P. E. (1993) *Science* **262**, 892-895.

Johnson, R. E., Adams, P. & Rupley, J. A. (1978) *Biochemistry* **17**, 1479-1484.

Johnson, W. C., Jr. (1990) *Proteins: Struct. Funct. Genet.* **7**, 205-214.

Kato, S., Okamura, M., Shimamoto, N. & Utiyama, H. (1981) *Biochemistry* **20**, 1080-1085.

Kato, S., Shimamoto, N. & Utiyama, H. (1982) *Biochemistry* **21**, 38-43.

Katta, V. & Chait, B. T. (1991) *Rapid. Commun. Mass Spectrom.* **5**, 214-217.

Katta, V. & Chait, B. T. (1993) *J. Amer. Chem. Soc.* **115**, 6317-6321.

Khechinashvili, N. N., Privalov, P. L. & Tiktopulo, E. L. (1973) *FEBS Lett.* **30**, 57-60.

Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993) *Biochemistry* **32**, 7054-7063.

Kim, K.-S. & Woodward, C. (1993) *Biochemistry* **32**, 9609-9613.

Kim, P. S. & Baldwin, R. L. (1990) *Ann. Rev. Biochem.* **59**, 631-660.

Koga, K. & Berliner, L. J. (1985) *Biochemistry* **24**, 7257-7262.

Kotik, M., Radford, S. E. & Dobson, C. M. (1995) *Biochemistry* **34**, 1714-1724.

Kraulis, P. J. (1991) *J. Appl. Cryst.* **24**, 946-950.

Kuroda, Y., Kidokoro, S. & Wada, A. (1992) *J. Mol. Biol.* **223**, 1139-1153.

Kuroki, R., Inaka, K., Taniyama, Y., Kidokoro, S., Matsushima, M., Kikuchi, M. & Yutani, K. (1992) *Biochemistry* **31**, 8323-8328.

Kuwajima, K. (1989) *Proteins: Struct. Funct. Genet.* **6**, 87-103.

Kuwajima, K. (1992) *Curr. Opin. Biotechnol.* **3**, 462-467.

Kuwajima, K., Hiraoka, Y., Ikeguchi, M. & Sugai, S. (1985) *Biochemistry* **24**, 874-881.

Kuwajima, K., Ikeguchi, M., Sugawara, T., Hiraoka, Y. & Sugai, S. (1990) *Biochemistry* **29**, 8240-8249.

Kuwajima, K., Nitta, K., Yoneyama, M. & Sugai, S. (1976) *J. Mol. Biol.* **106**, 359-373.

Kuwajima, K., Semisotnov, G. V., Finkelstein, A. V., Sugai, S. & Ptitsyn, O. B. (1993) *FEBS Lett.* **334**, 265-268.

Lala, A. K. & Kaul, P. (1992) *J. Biol. Chem.* **28**, 19914-19918.

Landry, S. J. & Gierasch, L. M. (1991) *Biochemistry* **30**, 7359-7362.

Landry, S. J., Jordan, R., McMacken, R. & Gierasch, L. M. (1992) *Nature* **355**, 455-457.

Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379-400.

Levinthal, C. (1968) *Chim. Phys.* **65**, 44-45.

Linderstrøm-Lang, K. (1955) *Chem. Soc. Spec. Publ.* **2**, 1-20.

Lissin, N. M. & Hemmingsen, S. M. (1993) *FEBS Lett.* **324**, 41-44.

Logan, T. L., Thériault, Y. & Fesik, S. W. (1994) *J. Mol. Biol.* **236**, 637-648.

Loo, J. A., Loo, R. R. O., Udseth, H. R., Edmonds, C. G. & Smith, R. D. (1991) *Rapid Commun. Mas Spectrom.* **5**, 101-105.

Lumb, K. J., Aplin, R. T., Radford, S. E., Archer, D. B., Jeenes, D. J., Lambert, N., MacKenzie, D. A., Dobson, C. M. & Lowe, G. (1992) *FEBS Lett.* **296**, 153-157.

Lumry, R. & Rosenberg, A. (1975) *Col. Int. C.N.R.S. L'Eau Syst. Biol.* **246**, 55-63.

Macura, C., Huang, Y., Suter, D. & Ernst, R. R. (1981) *J. Magn. Reson.* **43**, 259-281.

Marion, D. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113**, 967-974.

Martin, J. (1991) *Nature* **352**, 36-42.

Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M. & Fersht, A. R. (1990) *Nature* **346**, 440-445.

Matsumura, M., Signor, G. & Matthews, B. W. (1989) *Nature* **342**, 291-293.

Matthews, B. W. (1993) *Curr. Opin. Struct. Biol.* **3**, 589-593.

McCammon, J. A. & Karplus, M. (1983) *Acc. Chem. Res.* **16**, 187-193.

McKenzie, H. A. & White, F. H., Jr. (1991) *Adv. Protein. Chem.* **41**, 173-315.

Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T. & Dobson, C. M. (1993) *Science* **262**, 896-899.

Mirza, U. A. & Chait, B. D. (1994) *Anal. Chem.* **66**, 2898-2904.

Moore, P. B., Traut, R. R., Noller, H., Pearson, P. & Delius, H. (1968) *J. Mol. Biol.* **31**, 441-461.

Morozova, L., Haezebrouck, P. & Van Cauwelaert, F. (1991) *Biophys. Chem.* **41**, 185-191.

Mücke, M. & Schmid, F. X. (1994a) *Biochemistry* **33**, 14608-14619.

Mücke, M. & Schmid, F. X. (1994b) *J. Mol. Biol.* **239**, 713-725.

Neri, D., Billeter, M., Wider, G. & Wüthrich, K. (1992a) *Science* **257**, 1559-1563.

Neri, D., Wider, G. & Wüthrich, K. (1992b) *FEBS Lett.* **303**, 129-135.

Oas, T. G. & Kim, P. S. (1988) *Nature* **336**, 42-47.

Ohgushi, M. & Wada, A. (1983) *FEBS Lett.* **164**, 21-24.

Okazaki, A., Ikura, T., Nikaido, K. & Kuwajima, K. (1994) *Nature: Struct. Biol.* **1**, 439-446.

Pace, C. N. (1986) *Methods Enzymol.* **131**, 266-280.

Pace, C. N. (1990) *Trends Biochem. Sci.* **15**, 14-17.

Pace, C. N., Grimsley, G. R., Thomson, J. A. & Barnett, B. J. (1988) *J. Biol. Chem.* **263**, 11820-11825.

Pedersen, T. G., Sigurskjold, B. W., Andersen, K. V., Kjaer, M., Poulsen, F. M., Dobson, C. M. & Redfield, C. (1991) *J. Mol. Biol.* **218**, 413-426.

Peng, Z.-Y. & Kim, P. S. (1994) *Biochemistry* **33**, 2136-2141.

Peng, Z.-Y., Wu, L. C. & Kim, P. S. (1995) *Biochemistry* **34**, 3248-3252.

Pepys, M. B., Hawkins, P. N., Booth, D. R., Vigushin, D. M., Tennent, G. A., Soutar, A. K., Totty, N., Nguyen, O., Blake, C. C. F., Terry, C. J., Feest, T. G., Zalin, A. M. & Hsuan, J. J. (1993) *Nature* **362**, 553-557.

Perraudin, J.-P., Torchia, T. E. & Wetlaufer, D. B. (1983) *J. Biol. Chem.* **258**, 11834-11839.

Perry, L. J. & Wetzel, R. (1991) *Science* **226**, 555-557.

Pfeil, W., Bychkova, V. E. & Ptitsyn, O. B. (1986) *FEBS Lett.* **198**, 287-291.

Pfeil, W. & Privalov, P. L. (1976) *Biophys. Chem.* **4**, 23-32.

Poland, D. C. & Scheraga, H. A. (1965) *Biopolymers* **3**, 379-399.

Privalov, P. L. (1979) *Adv. Protein Chem.* **33**, 167-241.

Privalov, P. L. & Gill, S. J. (1988) *Adv. Protein Chemistry* **39**, 191-234.

Privalov, P. L. & Khechinashvili, N. N. (1974) *J. Mol. Biol.* **86**, 665-684.

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, O. I. (1990) *FEBS Lett.* **262**, 20-24.

Quinn, T. P., Tweedy, N. B., Williams, R. W. & Richardson, J. S. (1994) *Proc. Natl. Acad. Sci.* **91**, 8747-8751.

Radford, S. E., Buck, M., Topping, K. D., Evans, P. A. & Dobson, C. M. (1992a) *Proteins: Struct. Funct. Genet.* **14**, 237-248.

Radford, S. E., Dobson, C. M. & Evans, P. A. (1992b) *Nature* **358**, 302-307.

Radford, S. E., Woolfson, D. N., Martin, S. R., Lowe, G. & Dobson, C. M. (1991) *Biochem. J.* **273**, 211-217.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **117**, 479-485.

Redfield, C. & Dobson, C. M. (1988) *Biochemistry* **27**, 122-136.

Redfield, C., Smith, R. A. G. & Dobson, C. M. (1994) *Nature Struct. Biol.* **1**, 23-29.

Roder, H., Elöve, G. A. & Englander, S. W. (1988) *Nature* **335**, 700-704.

Rooman, M. J. & Wodak, S. J. (1992) *Biochemistry* **31**, 10239-10249.

Schellman, J. A. (1955) *C.R. Trav. Lab. Carlsberg Ser. Chim.* **29**, 230-259.

Scherf, T., Hiller, R. & Anglister, J. (1995) *FASEB J.* **9**, 120-126.

Schmidt, M. & Buchner, J. (1992) *J. Biol. Chem.* **267**, 16829-16833.

Segawa, T. & Sugai, S. (1983) *J.Biochem.* **93**, 1321-1328.

Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J. & Ptitsyn, O. B. (1987) *FEBS Lett.* **224**, 9-13.

Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. L. (1991) *Biopolymers* **31**, 119-128.

Senko, M. W. & McLafferty, F. W. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 763-785.

Shewale, J. G., Sinha, S. K. & Brew, K. (1984) *J. Biol. Chem.* **259**, 4947-4956.

Shin, H.-C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J. & Wright, P. E. (1993a) *Biochemistry* **32**, 6348-6355.

Shin, H.-C., Merutka, G., Waltho, J. P., Wright, P. E. & Dyson, H. J. (1993b) *Biochemistry* **32**, 6356-6364.

Shortle, D. & Abeygunawardana, C. (1993) *Structure* **1**, 121-134.

Smith, L., Sutcliffe, M. J., Redfield, C. & Dobson, C. M. (1993) *J. Mol. Biol.* **229**, 930-944.

Smith, L. J., Alexandrescu, A. T., Pitkeathly, M. & Dobson, C. M. (1994) *Structure* **2**, 703-712.

Sophianopoulos, A. J. & van Holde, K. E. (1964) *J. Biol. Chem.* **239**, 2516-2524.

Spitzfaden, C., Braun, W., Wider, G., Widmer, H. & Wüthrich, K. (1994) *J. Biomolec. NMR* **4**, 463-482.

Staley, J. P. & Kim, P. S. (1990) *Nature* **344**, 685-688.

Steudel, R. (1975) *Angew. Chem. Internat. Edn.* **14**, 655-664.

Strausberg, S., Alexander, P., Wang, L., Gallagher, T., Gilliland, G. & Bryan, P. (1993) *Biochemistry* **32**, 10371-10377.

Tanford, C. (1970) *Adv. Protein Chem.* **24**, 1-95.

Taniyama, Y., Yamamoto, Y., Kuroki, R. & Kikuchi, M. (1990) *J. Biol. Chem.* **265**, 7570-7575.

Taniyama, Y., Yamamoto, Y., Nakao, M., Kikuchi, M. & Ikehura, M. (1988) *Biochem. Biophys. Res. Commun.* **152**, 962-967.

Thomas, P. J., Ko, Y. H. & Pedersen, P. L. (1992) *FEBS Lett.* **312**, 7-9.

Thornton, J. M. (1981) *J. Mol. Biol.* **151**, 261-287.

Udgaonkar, J. B. & Baldwin, R. L. (1988) *Nature* **335**, 694-699.

van Dael, H., Haezebrouck, P., Morozova, L., Arico-Muendel, C. & Dobson, C. M. (1993) *Biochemistry* **32**, 11886-11894.

van Mierlo, C. P. M., Darby, N. J. & Creighton, T. E. (1992) *Proc. Natl. Acad. Sci.* **89**, 6775-6779.

van Mierlo, C. P. M., Darby, N. J., Keeler, J., Neuhaus, D. & Creighton, T. E. (1993) *J. Mol. Biol.* **229**, 1125-1146.

van Mierlo, C. P. M., Darby, N. J., Neuhaus, D. & Creighton, T. E. (1991a) *J. Mol. Biol.* **222**, 353-371.

van Mierlo, C. P. M., Darby, N. J., Neuhaus, D. & Creighton, T. E. (1991b) *J. Mol. Biol.* **222**, 373-390.

van Mierlo, C. P. M., Kemmink, J., Neuhaus, D., Darby, N. J. & Creighton, T. E. (1994) *J. Mol. Biol.* **235**, 1044-1061.

Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H. & Clore, G. M. (1993) *Science* **260**, 1110-1113.

Wagner, G. & Wüthrich, K. (1982) *J. Mol. Biol.* **160**, 343-361.

Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J. & Wright, P. E. (1993) *Biochemistry* **32**, 6337-6347.

Wand, A. J., Roder, H. & Englander, S. W. (1986) *Biochemistry* **25**, 1107-1114.

Wedin, R. E., Delepierre, M., Dobson, C. M. & Poulsen, F. M. (1982) *Biochemistry* **21**, 1098-1103.

Weissman, J. S. & Kim, P. S. (1991) *Science* **253**, 1386-1393.

Weissman, J. S. & Kim, P. S. (1992) *Science* **256**, 112-114.

Wells, J. A. & Powers, D. B. (1986) *J. Biol. Chem.* **261**, 6564-6570.

Wetzel, R. (1987) *Trends Biochem. Sci.* **12**, 478-482.

Wetzel, R., Perry, L. J., Baase, W. A. & Becktel, W. J. (1988) *Proc. Natl. Acad. Sci.* **85**, 401-405.

White, F. H., Jr. (1982) *Biochemistry* **21**, 967-977.

Williams, R. J. P. (1989) *Eur. J. Biochem.* **183**, 479-497.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995) *J. Biomolec. NMR* **5**, 67-81.

Wong, K.-P. & Tanford, C. (1973) *J. Biol. Chem.* **248**, 8518-8523.

Woodward, C. K. (1994) *Curr. Opin. Struct. Biol.* **4**, 112-116.

Woodward, C. K., Simon, I. & Tüchsen, E. (1982) *Mol. Cell Biochem.* **48**, 135-160.

Woody, R. W. (1985) In *The Peptides* , vol. 7, p 15, Academic Press, New York.

Wormald, C. (1991) D. Phil. Thesis, University of Oxford.

Wüthrich, K. (1986) *NMR of proteins and nucleic acids.*, J. Wiley & Sons, New York.

Wüthrich, K. (1994) *Curr. Opin. Struct. Biol.* **257**, 1559-1563.

Wyman, J. (1964) *Adv. Protein Chem.* **19**, 223-286.

Yamada, H., Uozumi, F., Ishikawa, A. & Imoto, T. (1984) *J. Biochem.* **95**, 503-510.

Yang, J. J., Pitkeathly, M. & Radford, S. E. (1994) *Biochemistry* **33**, 7345-7353.

Zahn, R., Spitzfaden, C., Ottiger, M., Wüthrich, K. & Plückthun, A. (1994) *Nature* **368**, 261-265.

Zhang, T., Bertelsen, E. & Alber, T. (1994) *Nature: Structural Biology.* **1**, 434-438.

Zhang, Z. & Smith, D. L. (1993) *Protein Science* **2**, 522-531.

Zhou, N. E., Kay, C. M. & Hodges, R. S. (1993) *Biochemistry* **32**, 3178-3187.

Zwickl, P., Pfeifer, G., Lottspeich, F., Kopp, F., Dahlmann, B. & Baumeister, W. (1990) *J. Struct. Biol.* **103**, 197-203.