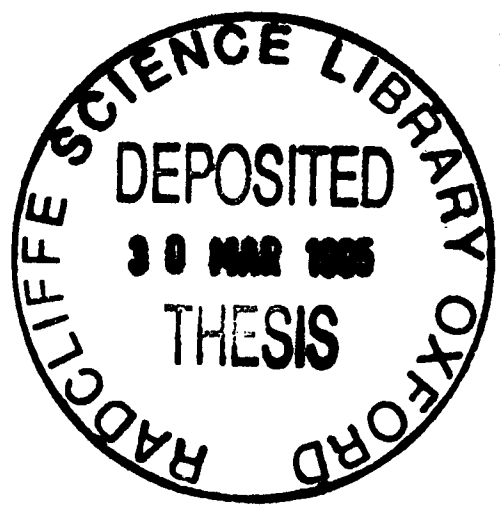
NMR Studies of the Dynamics and the Folding of Hen Lysozyme

Matthias Buck

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.

Brasenose College
Oxford

Michaelmas Term
1994



Abstract

NMR Studies of the Dynamics and the Folding of Hen Lysozyme

Matthias Buck, Brasenose College, Oxford

D. Phil.Thesis, Michaelmas Term, 1994

This thesis describes an investigation into the folding behaviour of hen lysozyme by characterisation of the structure and the dynamics of different conformational states of the protein. The native state, a partially structured state generated by the addition of a cosolvent, trifluoroethanol (TFE), and a highly denatured state of the protein in presence of urea at low pH, have been studied at equilibrium by circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy. The principal methods utilised in this thesis are the measurement and interpretation of amide hydrogen exchange and ^{15}N relaxation data.

Amide hydrogen exchange rates from a highly denatured state of hen lysozyme in 8 M urea, at pH 2.0, are well approximated by recent estimates for intrinsic exchange rates from unstructured polypeptides, implying that residual structures cannot be persistent in this state. In the native protein, by contrast, protection factors are measured to be of the order of 10^4-10^8 and can be rationalised by the involvement of amides in secondary structure and burial from the protein surface. Surface accessibility has also been found to be an important determinant of the dynamics of main and sidechain groups monitored by ^{15}N relaxation. For this purpose ^{15}N resonances of native lysozyme were assigned and order parameters derived. The dynamic behaviour was considered in the light of features of the crystal and NMR structures, suggesting that a lack of van der Waals contacts is a major determinant for mobility.

A denatured state of hen lysozyme, formed by addition of TFE cosolvent, has been developed as a model for a partially ordered conformation of the protein. Amide hydrogen exchange measurements show that sites with the highest protection factors, up to 250, are located in stable native-like helices. Near complete assignment of ^{15}N-edited NMR spectra allowed a detailed description of secondary structure in the TFE denatured state. Non-native α-helical structure is present as extensions to the native helices and in a region of the polypeptide which forms part of the β-sheet in the native state. These structures have been shown to be in accord with the helical propensities of the primary sequence.

Preliminary structure calculations suggest that, despite the absence of extensive and persistent tertiary interactions, topological restraints exist in parts of the TFE denatured state, resulting in considerable propensities for native-like arrangements of the secondary structure. Mainchain dynamics determined by ^{15}N relaxation, although increased in magnitude, appear to be related to that of the native state and are dominated by the position of disulphide bonds and by chain hydrophobicity. Thus the structural and the dynamic behaviour of the polypeptide chain in the TFE denatured state could be rationalised, at least in part, by consideration of features of the primary sequence.

To my Parents

Contents

Acknowledgements

Note on the Format of this Thesis

Abbreviations

Chapter 1: Introduction and Overview.

Chapter 2: Hydrogen Exchange in the Native State of Hen Lysozyme.

Chapter 3: Amide Hydrogen Exchange in a Highly Denatured State: Hen Lysozyme in Urea.

Chapter 4: A Partially Folded State of Hen-Egg White Lysozyme in Trifluoroethanol: Structural Characterisation and Implications for Protein Folding.

Chapter 5: Contributions of Disulphide Bonds to Structure of a Partially Ordered State of Hen Lysozyme in Trifluoroethanol.

Chapter 6: Structural Determinants of Protein Dynamics: Analysis of ^{15}N NMR Relaxation Measurements for Mainchain and Sidechain Nuclei of Hen Lysozyme.

Chapter 7: A Heteronuclear NMR Study of a Partially Ordered State of a Protein. Structural Preferences of Hen Lysozyme in Trifluoroethanol.

Chapter 8: A Heteronuclear NMR Study of a Partially Ordered State of a Protein. ^{15}N Relaxation Behaviour and Mainchain Dynamics of Hen Lysozyme in Trifluoroethanol.

Chapter 9: Epilogue

Publications

Appendix I: Theory of ^{15}N Relaxation Behaviour and Methods of Data Analysis.

Appendix II: Preliminary Calculations of the Structure of the TFE Denatured State of Hen Lysozyme.

Appendix S: Supplementary Data.

Acknowledgements

My principal thanks go to my supervisor, Christopher Dobson, for his generous support, sound advice and faith over the years. The guidance and the environment he provided allowed me glimpses of how to tackle important problems in science; many should prove invaluable for my future.

I would also like to thank Sheena Radford, for her continuous encouragement and the confidence she placed in me and my work. It was a pleasure to work with her.

Thanks are due to those who joined me in various projects, particularly to Christina Redfield and Jonathan Boyd, for their expertise and unfailing help in NMR theory, practice and analysis. Hui Lu and Jenny Yang for their dedication to studies with that bizarre cosolvent, TFE.

Special thanks go to Harald Schwalbe for taking on the challenge to pursue projects on the urea- and TFE-denatured states, for sharing good advice and his encyclopaedic knowledge of NMR. Christina Redfield and Lorna Smith provided advice on using X-PLOR and the protocols for structure calculation.

David Archer, Donald MacKenzie, Andrew Spencer and David Jeenes (all Institute for Food Research, Norwich) supplied large quantities of ^{15}N labelled hen lysozyme, without which my work would not have been possible. Jonathan Boyd, Nick Soffe and Mark Leonnard and Alison Rodger, Sheena Radford, respectively maintained the NMR and CD spectrometers.

I thank Niccolo Taddei (Florence) who challenged me with studies on acylphosphatase over a period of 3 years, and Dimitry Dolgikh (Pushkino) for introducing me to the equally difficult art of characterising a *de novo* designed protein, albebetin, by NMR.

I look back on the highly stimulating environment of scientific and sometimes social activities in Oxford thanks to my friends in College (known as the boring graduates: Nigel Cooper, Adrian Elcock, Adrian Muholland, Neil Moran, Chan Raut and Alfred Wong) as well as the many members of the Dobson group, not all of whom I can thank by name here. Special thanks go to those mentioned above as well as Andrei Alexandrescu, Linda Ball, Kim Bolin, Steve Eyles, Don Haynie, Shaun Hook, Andrew Miranker, Charles Morgan...and finally Caroline Bevan, who keeps the entire place running with her bizzazz.

Note on the Format of this Thesis:

For the convenience of both the readers and the author, this thesis is divided into seven largely independent sections, the format of each following closely that of a scientific paper. The following features result from this organisation:

Each section is comprised by a summary, introduction, materials and methods, the results and discussion sections as well as a list of references. In order to keep sections largely independent, some repetition of material was necessary.

Some of the sections have been published as papers; data from other contributors has been excised in Chapter 2. It should be noted that, the pulse sequences for the measurement of ^{15}N relaxation in NH_2 goups (Chapter 6) were developed and tested by Jonathan Boyd. He also recorded spectra of the TFE denatured state at 750 MHz ^1H frequency (Chapter 7 & 8). Harald Schwalbe recorded and analysed the data for $^3J(N^H,H_\alpha)$ coupling constants of the TFE denatured state (Chapter 7). These data are included here for completeness.

Several sections have been updated with recent data. For a historical development, especially with respect to the revision of correction parameters for the calculation of intrinsic hydrogen exchange rates, which occurred during the period of this research, attention is drawn to the published papers. Cross-references are indicated where appropriate, however, and an outline of the overall structure and objective of the research is given in the Introductory chapter. A number of outstanding questions are discussed with reference to this work in the Epilogue, Chapter 9.

Figures and Tables are presented together with their captions at the end of Chapters and Appendices.

Abbreviations

2D, 3D,	two and three dimensional;
$^3J(H^N,H_\alpha)$,	coupling between H^N and H_α;
CD,	circular dichroism;
COSY,	J-correlated spectroscopy;
CPMG,	Carr-Purcell-Meiboom-Gill;
CSA,	chemical shift anisotropy;
CSI,	chemical shift index;
$d_{i,i+1}(\alpha,N)$,	distance between e.g. H_α and H^N of residue i and i+1;
HEWL,	hen egg white lysozyme (E.C. 3.2.1.17);
HMQC,	heteronuclear multiple-quantum coherence;
HSQC,	heteronuclear single-quantum coherence;
$J(\omega)$,	spectral density function sampled at frequency ω (rad s^{-1});
mc,	mainchain;
NMR,	nuclear magnetic resonance;
NOE,	nuclear Overhauser enhancement;
NOESY,	nuclear Overhauser and exchange spectroscopy;
R_1,	longitudinal relaxation rate ($=1/T_1$);
R_2,	transverse relaxation rate ($=1/T_2$);
rmsd,	root-mean-squared deviation;
S^2,	order parameter;
sc,	sidechain;
Std,	standard deviation;
TFE,	2,2,2,-trifluoroethanol;
TOCSY,	total correlation spectroscopy;

Chapter 1

Introduction

Proteins are macromolecules of varied character and universal importance in living organisms, functioning in amongst other roles as structural scaffolds, catalysts and in signal transduction (Creighton, 1993). Their functionality depends on the details of the exquisitely well defined three dimensional structure, that the polypeptide chain, composed of a linear sequence of amino acid residues, adopts in space. While this sequence of amino acids is specified by the genetic code which is translated in a linear manner on ribosomes, the information for folding of the chain into complex three dimensional structures, is not encoded entirely locally but is distributed over significant regions of the protein sequence. The elucidation of the rules for the higher order interactions, which convert the disordered polypeptide chain into a compact well defined structure, has become known as the protein folding problem and remains as one of the major exacting challenges of biophysical science (Kim & Baldwin, 1982, 1990; Creighton, 1990; Matthews, 1993). Understanding the protein folding code would not only be a landmark in the study of complex systems, but would also have wide reaching practical applications in protein structure prediction and design, bridging the gap between genetic information and functional purpose for these molecules.

In recent years it has become apparent that protein synthesis in the cell is accompanied by a multitude of events, involving the association of the incompletely folded protein with a series of auxiliary proteins, including so called molecular chaperones, processes such as translocation through membranes and post-translational modifications (Hendrick & Hartl, 1993; Creighton, 1993). Classic experiments demonstrated, however, that small monomeric globular proteins generally contain all the necessary information for efficient folding in isolation of such components (Anfinsen, 1973). High resolution techniques have been developed to probe protein structure *in vitro* and have helped to suggest some of the basic principles of protein folding (e.g. Matthews, 1993). It is likely that these rules are universal and also apply to the folding processes *in vivo*.

For example, according to Levinthal a random search of the total number of possible conformations of even a moderately sized polypeptide chain would take an unrealistic length of time, estimated to be in excees of the age of the universe, while the folding process is accomplished in nature on a time-scale of typically milliseconds to minutes (Levinthal, 1968). It follows that folding must proceed as a directed process, involving folding pathways. Because such a reduction in the accessible conformational space is necessarily step-wise, folding intermediates must exist also. Indeed, some of these intermediates are sufficiently stable that they accumulate transiently during the folding of many proteins. Mechanisms by which efficient folding is achieved have been suggested from observations made in kinetic refolding studies of small proteins. Several of these are described below, using the folding pathway of hen lysozyme from a highly denatured state in 6 M guanidine, as an example (Figure 1).

The folding of this protein may be described from the experimentally available data as a three stage process (after Radford & Dobson, 1994). Details of the experimental methods and results are summarised elsewhere (Dobson et al., 1994; Evans & Radford, 1994).

Secondary Structure Formation and Chain Condensation

Extensive, but labile, secondary structure is formed extremely rapidly (milliseconds or less) in the refolding process of hen lysozyme and of other proteins (Kuwajima et al., 1993). However, it is known that the population of secondary structure in short isolated peptides is usually marginal and therefore additional longer range interactions must contribute to the stability of such structures. Interactions between non-polar sidechains are also likely to form on a short time-scale (Chan & Dill, 1991). Such clustering may occur locally, for example in concert with secondary structure formation, and act as nucleation sites from which further structures may propagate. Clusters may also interact with regions of the polypeptide chain more distant in sequence through a diffusion-collision process (Karplus & Weaver, 1994), probably at first in a non-specific manner, and this may generate an environment, which stabilises secondary structure formation. It is not clear at present whether such a chain condensation, or partial "hydrophobic collapse", pre- or succeeds, or is indeed concurrent with the secondary structure formation detected by circular dichroism.

Stabilisation of Structural Units

A second stage of folding is thought to involve the stabilisation of secondary structure to the extent that significant protection of amides against solvent exchange is detected (e.g. on the simplest of models this requires that secondary structure is populated for >90-99% of the time). This persistency may be achieved, for example by a partial docking of the secondary structural units against each other through complementary hydrophobic surfaces and by the formation of a rudimentary hydrophobic core. Depending on the extent and complementarity of the interacting surfaces, any number of secondary structural units may be involved, ranging from docking of two strands or helices to the cooperative stabilisation of entire structural domains. In the latter, it is likely that by this time the molecule has assumed a topology which is overall distinct and often native-like. The observation of parallel pathways for the refolding of hen lysozyme suggests that this second, and probably also the first, steps of folding may not always result in the correct coalescence of secondary structural units. It is possible that some of the molecules explore alternative topologies which have to be unfolded before folding to the native state can proceed. This should be possible because the majority of sidechains are not yet locked into tight packing interactions as indicated by the absence of near uv CD.

Acquisition of the Native Structure

In case of hen lysozyme and many other proteins, the transition state of folding is known to be close to the native state (Segawa & Sugihara, 1984) and therefore the last steps of folding are the slowest kinetically. These are thought to involve several events, such as the expulsion of water molecules remaining in the hydrophobic core, the locking of sidechains into tight packing configurations, the organisation of the conformations of loops and turns at the protein surface and, in the case of multi-subdomain proteins the docking of these structural units. Little is currently known experimentally about the details of these last steps of folding. Analysis of transition states, characterised by the effect of site-directed mutagenesis on structural stability and on folding or unfolding kinetics (Fersht, 1993), as well as the study of structural intermediates in unfolding reactions by hydrogen exchange and fast NMR techniques (Mayo & Baldwin, 1993), are, however, beginning to provide insights into these processes.

While it is likely that the processes which occur during folding, here roughly outlined in three steps, are universal to all globular proteins, the requirements

as to the magnitude and the temporal arrangement of these steps will be determined by the type of fold and the nature of interactions formed in the native structure. Thus, the study of the folding behaviour of a number of different proteins in conjunction with examination of characteristics of their native states has begun to provide insights into the rules that govern folding processes (e.g. Lu & Dahlquist, 1992; Varley et al., 1993; Otzen et al., 1994).

Folding Studies at Equilibrium

Complete understanding of the folding pathway of a protein requires detailed knowledge of the structures of the intermediates which must be sampled and of their energetics, as well as of the barriers associated with their interconversion. Such characterisation is complicated in kinetic folding experiments by the low population of folding intermediates, their heterogeneity, their short-lived nature, the apparent cooperativity of many rate determining events, the existence of parallel folding pathways and the often limited availability of structural probes (Schmid, 1992). Furthermore, the characteristics of the starting and end states of the folding reaction are likely to provide critical insights into the folding processes. Studies of conformational states of proteins under equilibrium conditions can be carried out with an increasingly wide range of high resolution and site specific methods and is beginning to provide this information.

The central aim of the research presented in this thesis has, therefore, been to study one protein, hen lysozyme, in different conformational states at equilibrium, obtaining information about the structure and the dynamic behaviour of the polypeptide chain. In the following sections an overview is given of our knowledge of selected features of denatured states of proteins and protein fragments, of states with intermediate order, and of native states. References are made to reviews for more extensive discussions and to the main body of the thesis for a review of studies pertaining to hen lysozyme and related proteins.

Denatured states (U).

"Denatured" has come to refer to conformations of proteins which are at least in part disordered and non-native. Highly denatured states are a critical reference state in kinetic as well as thermodynamic investigations of protein folding (Shortle, 1993). Its physiological relevance arises because proteins emerge from the ribosome as they are translated from the genetic material, at least initially, in unstructured conformations. They must also be substantially unfolded for translocation processes across membranes and for their

proteolytic degradation (Dill & Shortle, 1991). The notion that the ploypeptide adopts a truely random conformation in highly denatured proteins may be dismissed by considering the physico-chemical differences between residue sidechains. Rotations about mainchain bonds are severely restricted in Ramachandran space by the presence of a β-substituent. Thus repulsive van der Waals interactions limit the accessible conformational space, with glycine and proline presenting special cases. Similar arguments apply to the conformations of sidechains, which, in theoretical and experimental studies (Dunbrack & Karplus, 1994), are found to be largely restricted to a few rotamers around each bond.

No atom in the heteropolymer exists in a environment neutral to all the other substituents, whether covalently linked or not. The balance between the multitude of competing interactions of the polypeptide with itself and with the solvent, will depend on the precise solution conditions, such as pH, ionic strength, and the presence of other cosolvents as well as the temperature. Furthermore, it is clear, that the extent of residual non-covalent interactions of the polypeptide chain with itself are highly non-uniform depending on the amino acids of the underlying sequence (Shakhnovich & Finkelstein, 1989; Dill & Shortle, 1991). The details of the direct or indirect interactions of solvent with the polypeptide chain are still incompletely understood, especially in the case of multicomponent systems, e.g. denaturation in urea, alcohol or sugars.

Experiments on denatured states of intact proteins have spanned many decades and have used methods ranging from optical spectroscopic techniques to gel filtration, low angle X-ray scattering and NMR (Tanford, 1968; Privalov et al., 1989; Dill & Shortle, 1991; Evans et al, 1991). Recently, the site-specific characterisation of denatured proteins has become possible by the use of heteronuclear and multidimensional NMR methods (e.g. Neri et al., 1992; Logan et al., 1994). The results on local structures are similar, at least in outline, to those obtained for isolated peptides in aqueous solution, although the population of well defined structures is generally reduced when chemical denaturants are present, such as urea or guanidine. It is possible, however, that a denaturant also may stabilise certain structures indirectly, by effects on water structure for example or directly by binding to the polypeptide (Makhatadze & Privalov, 1992; Neri et al., 1992; Liepinsh & Otting, 1994).

The latter possibility and the observation of significant residual structure in denatured states of hen lysozyme by a number of biophysical techniques (reviewed in the main part of the thesis, particularly Chapters 3 and 4) has

prompted an investigation of the dynamic behaviour of the mainchain of lysozyme in the urea denatured state of the protein at low pH. This study, carried out by the measurement of amide hydrogen exchange rates, is reported in Chapter 3 and results are compared with similar data on other denatured states of hen lysozyme and the related protein α-lactalbumin. The significance of amide protection for the characterisation of residual structures in denatured states is discussed.

Short Peptides and Protein Fragments

Study of residual structure in denatured states of intact proteins often involves extreme conditions of temperature, pH or chemical denaturant. In small peptides and protein fragments long range interactions are by definition removed, providing an opportunity to study local structural propensities. For a long period, the α-helix formed in the S-peptide of ribonuclease A was the only example of a peptide in aqueous solution populating a significantly ordered conformation (Bierzynski et al., 1982). This introduced the notion that such structures are generally unstable in short peptides (< 20 amino acids), but that they could be detected if structure stabilising cosolvents, such as TFE were added.

Over recent years, this picture has changed, as stable turn conformations were identified in very short peptides by NMR (Dyson & Wright, 1991), as the number of examples of peptides with an appreciable content of α-helix structure has increased (Scholtz & Baldwin, 1992) and recently, as isolated β-hairpin structures have been characterised in aqueous solution (Blanco et al., 1994). However, in general, structures in peptides, just as the majority of regions with residual structures in denatured states of intact proteins, exist in rapid equilibrium with unfolded states, thus making their characterisation difficult. It is clear, however, that ordered conformations, such as a series of turns, referred to as a nascent helix, are sampled by many peptides for at least a fraction of time (Dyson & Wright, 1993). Several investigators use peptide systems to model particularly stable structures in order to elucidate the interactions required for their formation and stability (Staley & Kim, 1990; Scholtz & Baldwin, 1992; Munoz & Serrano, 1994). It is generally envisaged that an understanding of the propensities for the formation of local structures is an important step towards solving the folding problem. In addition it is presumed that structures which are stable in the presence of only short regions of polypeptide chain may be related to those formed in the earliest initiation events of the folding process (e.g. Wright et al., 1988; Moult & Unger, 1991).

Intermediate States of Proteins

Intensive attention has focused on species intermediate between the highly denatured and the native states of proteins. Many of these conformations are characterised by a significant content of native-like secondary structure. Persistent and specific tertiary interactions are largely absent in so called "molten globule states", which are compact intermediates, generally observed to bind the hydrophobic dye 1-anilinonaphthalnene-8-sulfonate, or ANS (Ptitsyn, 1992). This requires that hydrophobic sidechains are significantly clustered while remaining at least partially accessible to solvent. Early investigations concluded that, in the case of α-lactalbumin, such partially folded states at equilibrium may be indentical to states populated transiently on the folding pathway (Ikeguchi et al., 1986) (see Epilogue). While such a state may represent a stable third thermodynamic state at equilibrium for some proteins (Friere et al., 1993; Ptitsyn, 1994) it is also proposed as a universal kinetic refolding intermediate (Ptitsyn, 1992).

Experimental characterisation of partially folded states at equilibrium is complicated by the high cooperativity of the folding process in most proteins; these states may be unstable relative to the native and highly denatured ones. Thus a wide range of solution conditions has been used for the stabilisation of partially folded states, including the use of extremes of pH, temperature and salt concentration as well as the addition of chemical denaturants. Furthermore, structural analysis of partially folded states, especially classical "molten globule" states, such as that of α-lactalbumin, by high resolution NMR is often complicated by unfavourable dynamic properties (Alexandrescu et al., 1993). Motions on the micro- to millisecond timescale lead to severe resonance broadening, making extensive assignment extremely difficult. Successful characterisation has been possible for a number of states which are either more unfolded or more highly ordered (Wüthrich, 1994; Dobson, 1994)

The "molten globule" concept has gradually been widened to include these states. For example, significant parts of apocytochrome b562 (Feng et al., 1994), Interleukin-4 at low pH (Redfield et al., 1994), as well as of disulphide intermediates of BPTI (van Mierlo et al., 1993) are as structured in the partially folded state as in their native states. A tightly packed core is present, while other regions of these proteins, such as loops and termini of some of the helices or sheets, are either highly unstructured or have "molten globule" type character. By contrast to these "ordered molten globules", other states are overall less structured (Dobson, 1994). Only a few regions of secondary structure are clustered through non-specific contacts between their interfaces in

the acid denatured state of cytochrome c for example (Jeng & Englander, 1992). The existence of pre- or so called "wet molten globule" states has been proposed from theoretical considerations and experimental data (Shakhnovich & Finkelstein, 1989; Ptitsyn, 1994).

Given the different characteristics of the partially folded and the "molten globule" states which have been studied, it is indeed questionable whether they are universal intermediates with largely common structural features (Kuwajima, 1989; Ptitsyn, 1992; Fink et al., 1994). For example, mutations in the hydrophobic core of apomyoglobin do not seem to affect the structure of its molten globule state significantly, in agreement with the notion that tertiary contacts are largely non-specific in such states (Barrick & Baldwin, 1993). By contrast, mutations of residues known to be involved in non-native and relatively fixed tertiary contacts in α-lactalbumin (Alexandrescu, et al., 1993) have dramatic consequences on the folding behviour of the protein (Brew, personal communication). A large number of questions remain unanswered regarding the nature of the structures which exist in "molten globule" type-states and their role in the folding process.

The characterisation of a partially folded state of hen lysozyme, generated by the addition of TFE cosolvent is described in chapters 4 and 5. Due to its favourable spectral characteristics, direct NMR structure determination was possible using ^{15}N labelled protein and is reported in chapter 7. The structure in this state has been assessed in the light of the predicted propensities of the polypeptide chain for structure formation. Furthermore, the dynamic behaviour of the mainchain in this partly structured state has been probed by ^{15}N relaxation measurements, as described in chapter 8.

Native States

The ultimate aim of investigations into the protein folding problem is to explain how the specificity and stability of structures in the native state of proteins arise from the primary sequence of amino acids. Many features of native states themselves provide important clues to this end. These include, amongst others, the high density of packing of sidechains in the protein core (Richards, 1994), the partitioning of sidechains according to their chemical character into the protein core or to the surface or protein exterior as well as the optimisation of main and sidechain hydrogen bonding (Baker & Hubbard, 1984; McDonald & Thornton, 1994). In recent years, the number of protein structures solved to high resolution by X-ray or NMR spectroscopy has increased dramatically. Although, increasingly proteins are found to have

structures homologous to folds already represented in the database of protein structures, the number of unique folds has more than doubled in the last three years (Holm & Sander, 1994). This wealth of information has lead to several "knowledge-based" approaches towards predicting the structure of a protein (Wodak & Rooman, 1993). Indeed, it is likely that homology modelling and sequence-structure comparison methods will shortly be able to predict structures to considerable accuracy of those proteins which have folds already represented in the database.

Site directed mutagenesis allows, in favourable cases, the specificity and stability of interactions which give rise to structures to be probed in the context of the folded protein (eg. Horovitz & Fersht, 1992). Studies on the λ-repressor and T4 lysozyme, for example, demonstrated that the folding code is largely redundant and that a large number of mutations may be tolerated (e.g. Bowie et al., 1990; Heinz et al., 1992). It has become apparent that even the tightly packed core of proteins is somewhat malleable, as both sidechains and the mainchain may move to fill holes or avoid bad steric contacts which are created by some of the mutations. These findings are in agreement with the notion that the information content of a complex system, such as the fold of a protein, is generally largely distributed and that there can be considerable redundancy.

Although the naive state of a protein is often seen as a unique state at its global thermodynamic minimum, this does not mean that structures are entirely static (Karplus & McCammon, 1983). A wide range of motions occur on time scales ranging over typically 12 orders of magnitude. The great majority of motions involve small perturbations from an average structure. This equilibrium between conformational substates arises because the multidimensional energy landscapes for proteins are very complicated in detail and contain many minima (Frauenfelder et al., 1988). Larger motions, such as hinge bending or sliding motions between modular domains have been inferred for several proteins (Faber & Matthews, 1990; Gerstein et al., 1994). Just as with the smaller motions within domains, it is not clear whether they are essential for protein function or simply a necessary consequence of the protein architecture (Dobson, 1993).

Studies of the native state of hen lysozyme were, therefore, undertaken to investgate possible relationships between structure and dyamics. Dynamics were studied by two different techniques reflecting motions on two largely disparate timescales. Firstly by the study of protection of amide hydrogens

against solvent exchange and secondly by measurements of ^{15}N relaxation are described in Chapters 2 and 6 respectively.

References

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

Alexandrescu, A.T. & Shortle, D. (1994) J.Mol.Biol. 242, 527-546.

Anfinsen, C. (1973) Science 181, 223-230.

Baker, E.N. & Hubbard, R.E. (1984) Prog.Biophys.Mol.Biol. 44, 97-179.

Barrick, D. & Baldwin, R.L. (1993) Protein Science 2, 869-76.

Bierzynski, A., Kim, P.S. & Baldwin, R.L. (1982) Proc.Natl.Acad.Sci. (USA) 79, 2470-2474.

Blanco, F.J., Rivas, G. & Serrano, L. (1994) Nature Struct.Biol. 1, 584-590.

Bowie, J.U., Reidhaar-Olson, J.F., Lim, W.A. & Sauer, R.T. (1990) Science 247, 1306-1310.

Chan, H. & Dill, K. (1991) Ann.Rev.Biophys.Biophs.Chem. 22, 447-490.

Creighton, T.E. (1990) Biochem.J. 270, 1-16.

Creighton, T.E. (1993) in "Proteins: Structure Function and Molecular Properties", Freeman & Company, N.Y.

Dill, K.A. & Shortle, D. (1991) Ann.Rev.Biochem. 60, 795-825.

Dobson, C.M. (1993) Curr.Biol. 3, 530-532.

Dobson, C.M. (1994) Curr.Biol. 4, 636-640.

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

Dunbrack, R.L. & Karplus, M. (1994) Nature Struct.Biol. 1, 334-40.

Dyson, H.J. & Wright, P.E. (1991) Ann.Rev.Biophy.Biophy.Chem 20, 519-538.

Dyson, H.J. & Wright, P.E. (1993) Curr.Opin.Struct.Biol. 3, 60-65.

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

Evans, P.A., Topping, K.D., Woolfson, D.N. & Dobson, C.M. (1991) Proteins: Struct.Funct. & Genetics 9, 248-266.

Faber, H.R. & Mathews, B.W. (1990) Nature 348, 263-266.

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

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

Fink, A.L., Calciano, L.J., Goto, Y., Kurotsu, T. & Palleros, D.R. (1994) Biochem. 33, 1204-1211.

Frauenfelder, H., Parak, F. & Young, R.D. (1988) Ann.Rev.Biophys.Biophys.Chem. 17, 451-479.

Friere, E., Haynie, D.T. & Xie, D. (1993) Proteins: Struct.Funct. & Genetics 17, 111-123.

Gerstein, M., Lesk, A.M. & Chothia, C. (1994) Biochem. 33, 6739-6749.

Heinz, D.W., Baase, W.A. & Matthews, B.W. (1992) Proc.Natl.Acad.Sci. (USA) 89, 3751-3755.

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

Holm, L. & Sander, C. (1994) Proteins: Struct.Funct. & Genetics 19, 165-173.

Horovitz, A. & Fersht, A.R. (1992) J.Mol.Biol. 224, 733-740.

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

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

Karplus, M., & McCammon, J.A. (1983) Ann.Rev.Biochem. 53, 263-300.

Karplus, M. & Weaver, D.L. (1994) Protein Sceince 3, 650-668.

Kemmink, J. & Creighton,T.E. (1993) J.Mol.Biol. 234, 861-878.

Kim, P.S. & Baldwin, R.L. (1982) Ann.Rev.Biochem. 51, 459-489.

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

Kuwajima, K. (1989) Proteins: Struct.Funct. & Genetics 6, 87-103.

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

Levinthal, C. (1968) J.Chimie Phys. 65, 44-45.

Liepinsh, E. & Otting, G. (1994) J.Am.Chem.Soc. 116, 9670-9674.

Logan, T.M., Theriault, Y., Fesik, S.W. (1994) J.Mol.Biol. 236, 637-658.

Lu, J. & Dahlquist, F.W. (1992) Biochem. 31, 4749-4756.

Makhatadze, G.I. & Privalov, P.L. (1992) J.Biol.Mol. 226, 491-505.

Matthews, C.R. (1993) Ann.Rev.Biochem. 62, 653-683.

Mayo, S.L. & Baldwin, R.L. (1993) Science 262, 873-876.

McDonald, I.K. & Thornton, J.M. (1994) J.Mol.Biol. 238, 777-793.

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

Moult, J. & Unger, R. (1991) Biochem. 30, 3816-3824.

Munoz, V. & Serrano, L. (1994) Nature: Struct.Biol. 1, 399-409.

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

Otzen, D.E., Itzhaki, L.S., Elmasry, N.F., Jackson, S.E. & Fersht, A.R. (1994) Proc.Natl.Acad.Sci. (USA) 91, 10422-10425.

Privalov, P.L., Tiktopulo, E.I., Venyaminov, S.Y., Griko, Y.V., Makhatadze, G.I. & Khechinashvili (1989) J.Mol.Biol. 205, 737-750.

Ptitsyn, O.B. (1992) in "Protein Folding", T.E. Creighton Ed, Freeman & Company, N.Y.

Ptitsyn, O.B. (1994) Protein Eng. 7, 593-596.

Radford, S.E. & Dobson, C.M. (1994) Proc.Roy.Soc.B. in press.

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

Richards, F.M. (1992) in "Protein Folding", T.E. Creighton Ed. Freeman, N.Y.

Rizo, J., Blanco, B.K., Bruch, M.D. & Gierasch, L.M. (1993) Biochem. 32, 4881-94.

Schmid, F-X. (1992) in "Protein Folding", T.E. Creighton Ed. Freeman, N.Y.

Scholtz, J.M., & Baldwin, R.L. (1992) Ann.Rev.Biophy.Biomol.Struc. 21, 95-118.

Segawa, S-I. & Sugihara, M. (1984) Biopolymers 23, 2473-2488.

Shakhnovich, E. & Finkelstein, A. (1989) Biopolymers 28, 1667-1680.

Shortle, D. (1993) Curr.Opin.Struct.Biol. 3, 66-74.

Staley, J.P & Kim, P.S. (1990) Nature 344, 683-688.

Tanford, C. (1968) Adv.Prot.Chem. 23, 121-282.

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

Wodak, S.J. & Rooman, M.J. (1993) Curr.Opin.Struct.Biol. 3, 247-259.

Wright, P.E., Dyson, H.J. & Lerner, R.A. (1988) Biochem. 27, 7167-7175.

Wüthrich, K. (1994) Curr.Opin.Struct.Biol. 4, 93-99.

Outline of Thesis

Part I. Chapters 2 to 5 describe hydrogen exchange, homonuclear NMR and CD studies carried out on the native state, a partially folded state, and a highly denatured state of hen lysozyme. These techniques allowed a description of the dynamics of approximately half of the mainchain sites in these states, while providing a characterisation of global structural features.

Part II. Chapters 6 to 8 describe heteronuclear NMR studies on the native state of hen lysozyme as well as on the partially ordered state in TFE. Resonance assignments of ^{15}N edited spectra are presented together with measurements of ^{15}N relaxation behaviour for both the mainchain and for the majority of ^{15}N labelled sidechains in the native state. Several structural features are considered as determinants for mobility in the native state and characteristics of the amino acid sequence are assessed in the light of the dynamics observed in the denatured state. A description of the structure in the partially ordered state is presented and compared with intrinsic propensities of the polypeptide chain. An introduction to the theory of ^{15}N relaxation and the data analysis is given in Appendix I. A preliminary structure calculation for the TFE state is reported in Appendix II.

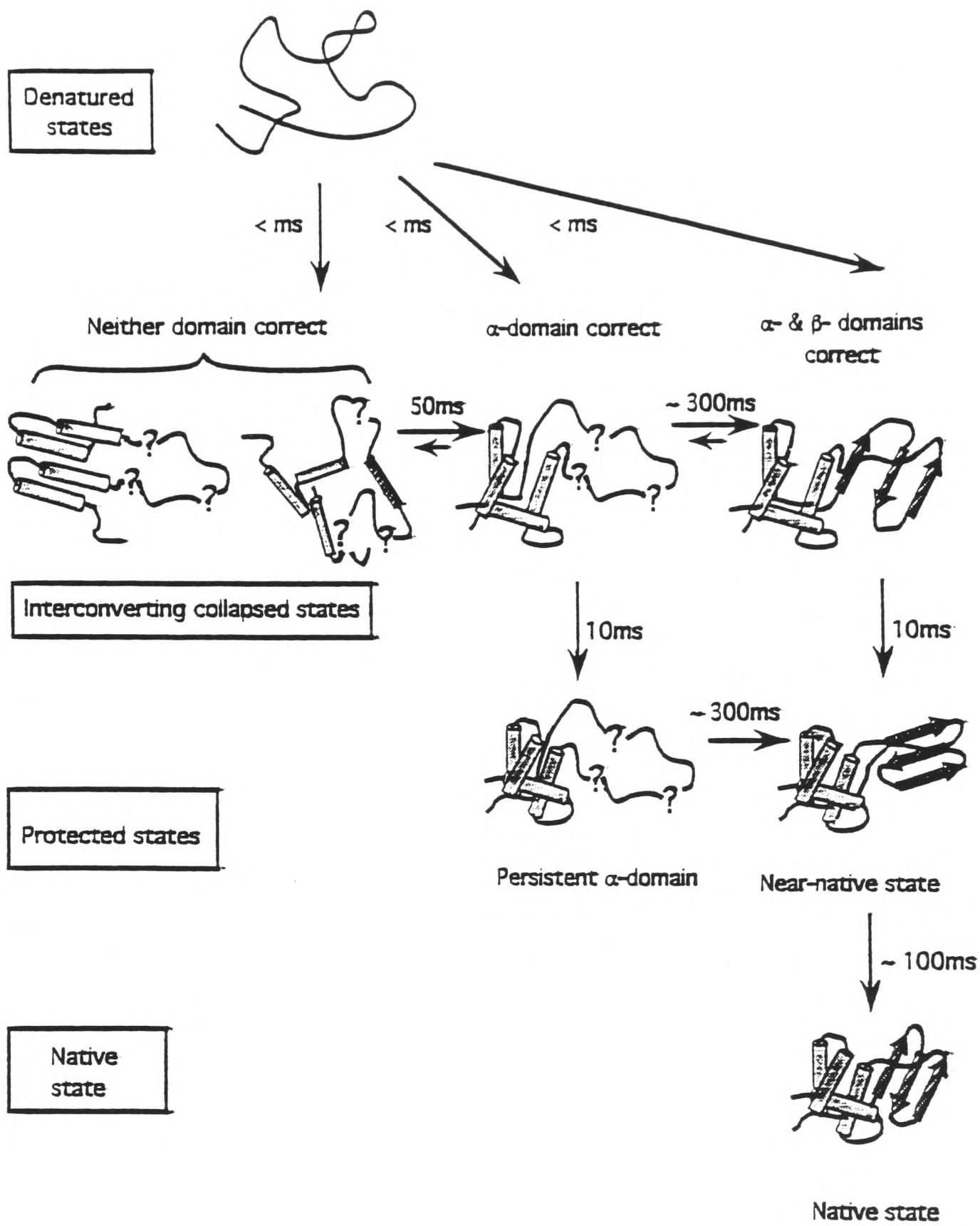


Figure 1
The folding pathway of hen lysozyme (adapted from Radford & Dobson, 1994), showing folding as a process in three stages together with the parallel nature of the pathway.

Chapter 2

Hydrogen Exchange in the Native State of Hen Egg-White Lysozyme.

Summary

The hydrogen exchange kinetics of 67 individual amide protons in the native state of hen lysozyme have been measured at pH 7.5 and 30°C by 2D ^{1}H NMR methods. These constitute the most protected subset of amides, with exchange half lives some 10^4-10^8 times longer than anticipated from studies of small model peptides. The observed distribution of rates under these conditions can be rationalised to a large extent in terms of the hydrogen bonding of individual amides and their burial from bulk solvent.

Introduction

The kinetics of hydrogen exchange have become increasingly important as a means of characterising the structure and dynamics of proteins. In native globular proteins there has long been an interest in the dynamic processes by which labile hydrogens, ordinarily buried in the structural core become accessible to solvent water and permit hydrogen exchange to occur (Hvidt et al., 1966; Woodward & Simon, 1982; Englander & Kallenbach, 1984). More recently, the utility of hydrogen exchange as a probe for residual, specifically stabilised structure in non-native states, including transient folding intermediates, has become established (Roder, 1989; Kim & Baldwin, 1990; Creighton, 1991).

The mechanisms of exchange in native proteins are still not well understood. In the case of a number of proteins, including hen egg-white lysozyme, it has been observed that the dominant exchange mechanism for many of the more highly protected hydrogens changes with temperature (Woodward & Hilton, 1980; Wedin et al., 1982). Under conditions remote from those of denaturation, exchange is often mediated by low activation energy processes and it is supposed that these reflect localised fluctuations of the structure. The exchange of the most protected amides in the structural core of the protein, however, may be dominated by larger scale fluctuations of much higher activation energy which are of considerable interest in relation to protein folding (Woodward & Hilton, 1980; Wedin et al., 1982; Delepierre, et al. 1983; Roder et al., 1985; Woodward, 1994).

In the present paper the pattern of amide protection was examined under strongly native conditions, similar to those under which the pattern of protection during the refolding process was previously investigated (Miranker et al., 1991). The rates of exchange of over half of the 126 amide hydrogens were determined. This extends considerably the previous qualitative examination of the exchange properties of hen lysozyme (Pedersen et al., 1991). Although under a single set of conditions a full interpretation of the hydrogen exchange kinetics cannot be deduced, the data are of particular importance since the pattern of exchange kinetics can be interpreted in terms of structural protection afforded by the native conformation.

Materials & Methods

Hen egg-white lysozyme was obtained from Sigma and further purified by extensive dialysis against distilled water at pH 3.0 before use.

Amide NH Exchange in the Native State at pH 7.5, 30°C.

Exchange was carried out at 30°C in a D_2O incubation mixture containing 0.1 M NaCl, 3 mM sodium azide and 20 mM sodium phosphate buffer, pH 7.5. pH meter readings, calibrated against standard buffers (Aldrich), were not corrected for isotope effects and remained constant to within ±0.1 pH units throughout the experiment. Hydrogen exchange was initiated by dissolution of lyophilized protein in the exchange buffer to give a protein concentration of 2.5 mg/ml (0.17 mM). This protein concentration was chosen so that aggregation of lysozyme was negligible (<5%) under these conditions (Sophianopoulos & Van Holde, 1964; Deonier & Williams, 1970). This was confirmed by the observation that exchange rates were identical when selected time points were repeated at a lower protein concentration of 0.25 mg/ml. Sixteen samples were withdrawn from the incubation after time intervals ranging from 40 seconds to 70 days and the exchange quenched by addition of 2.3 ml of 240 mM deuterated sodium acetate buffer, pH 3.6, giving a final pH of 3.8. Samples were concentrated 30 fold by ultrafiltration at 4°C to give a final protein concentration of 5 mM. Samples were stored at either 4°C (for short intervals) or -20°C (for longer intervals) before acquisition of the NMR spectrum.

NMR Spectroscopy.

NMR experiments were performed using the 500 MHz Bruker spectrometer belonging to the Oxford Centre for Molecular Sciences. Phase sensitive COSY spectra were acquired at 35°C using TPPI (Marion & Wüthrich, 1983). Data sets comprised 256 x 2048 data points in the frequency domain. The intensity of each NH - CαH cross peak was taken to be the average of the heights (neglecting sign) of the four components of the cross peak, and were normalised by comparison with the intensities of correlations between the non-exchangeable protons of Tyr23 C3,5H-C2,6H and Tyr53 C3,5H-C2,6H. These intensities were then normalised with respect to those measured for a control sample in which exchange of protected amide protons was negligible. This sample was obtained by dissolving fresh lysozyme in D_2O at pH 3.8 and immediately acquiring a COSY spectrum.

Results

The rates of exchange of 67 amides were measured at pH 7.5, 30°C. The half lives for exchange are shown versus the protein sequence in Figure 1a (and exchange rates are listed in Appendix S-I). Rates of exchange for all of the amides measured here are extremely slow; values for the half-life vary from 1 to 10^5 min, in comparison with a theoretical half-life of about 10^{-4} min, calculated for small model peptides under these conditions (Bai et al., 1993). Indeed, for two amides (Met12 and Val29) only upper limits for the exchange rates could be estimated since exchange of the amides was negligible, even after 70 days at pH 7.5 and 30°C. The remaining amides exchanged too quickly to be measurable in this experiment.

In order to compare the exchange properties of individual amides in the native protein protection factors were calculated. These are defined as $P = k_{int}/k_{ex}$, where k_{ex} is the measured exchange rate and k_{int} is the predicted intrinsic exchange rate for each amide hydrogen in small unstructured peptides under the same conditions, based on data for the pH, temperature and sequence dependence of amide exchange (Bai et al., 1993). The protection factors measured for the native enzyme are shown as a function of the protein sequence in Figure 1b.

Amide hydrogens in lysozyme have previously been classified into four categories (I-IV) based upon their exchange behaviour in solution at pH 4.2 and 7.5 (Pedersen et al., 1991). Amides in class I and two members of class II (Arg14 and Ser85) correspond to those which exchanged too fast to be measured under the conditions of the present study. Protection factors for the majority of amides for which rates were obtained ranged from 10^4 to 10^8, consistent with their previous classification into the exchange categories II-IV. They are listed in Table 1a of Chapter 3.

Discussion

The protection factors of between 10^4 and 10^8, shown in Figure 2, emphasise both that the exchange rates of protected amide hydrogens in lysozyme at pH 7.5, 30°C are very slow and that they vary widely. This implies that the mechanism by which these amides become transiently exposed to solvent, permitting exchange to occur, is not a cooperative unfolding of the structure under these conditions but rather that localised fluctuations, of widely differing frequency, are involved. Only for the few slowest exchanging amides of all, for

which rates were too slow to measure under the present conditions, does the possibility remain that exchange occurs by a global unfolding mechanism. It is thus of considerable interest to look for relationships between the distribution of exchange rates and features of the folded structure of the protein. Lysozyme is an α+β protein, comprising two structural lobes (Blake et al., 1965). One is constructed largely from helical backbone segments enclosing a large hydrophobic core; the other includes a rather distorted three stranded β-sheet and an extended loop of irregular structure.

It is clear from Figure 1 that there is a close correlation, under these conditions, between slow amide exchange and participation of the particular amide in a region of secondary structure. Thus, the majority of amides in helical or β-sheet structures have protection factors exceeding 10^5. In general, exchange of amides in short, irregular helical regions is faster than that for amides which occur in more regular and extensive segments of secondary structure, see Table II; for example the short D helix (encompassing residues 108-115) is much more weakly protected than the three longer α-helices (residues 4-15, 24-36 and 88-99, respectively); protection factors in helix D were some 100-fold smaller than those seen for residues in helices A, B and C. In addition, the two short segments of 3_{10}-helix which are located in opposite lobes of the lysozyme structure have similar protection factors to each other and to the D-helix.

It is interesting to note that there are pronounced differences in exchange rate within individual helices; protection at the extremities of the helices is, in general, smaller than that at the centre. Thus, for example, the four slowest exchanging amides of all (Met12, Val29, Ala95 and Lys96) lie in the centre of their respective helices. In addition, amides which line the exposed side of amphipathic helices exchange much more rapidly than those which abut the hydrophobic core of the enzyme. Thus residues Val92, Ala95 and Ile98 in helix C exchange much more slowly than those along the opposite face (Asn93, Cys94 and Lys97). Such a pattern is not observed for helix B which is the most deeply buried and hydrophobic helix of all; in this case all amides exchange relatively slowly. It is apparent, therefore, that individual helices do not unfold as cooperative elements to permit exchange under these conditions; rather the dominant exchange mechanism seems to involve much smaller scale structural fluctuations, the rates of which vary widely depending on the details of the structural context of the individual hydrogen bonded amide. The nature of the fluctuations, however, is not known (Hvidt et al., 1966; Woodward & Simon, 1982; Englander & Kallenbach, 1984).

The highest protection factors for amides in the triple stranded β-sheet (residues 41-60) are 10^7, up to a magnitude smaller than the highest protection factors in the α-helical subdomain. This is interesting, since these lobes of the lysozyme molecule have been shown to be stabilised independently on the folding pathway (Miranker et al., 1991). As discussed above, the widely disparate rates observed for different amides indicate that their exchange is mediated by independent, small scale fluctuations. The implication is, therefore, that the energy required to bring about these structural distortions may be slightly lower in the most stable regions of the β-subdomain, than it is in the α-subdomain. A second, much shorter double stranded β-sheet is formed between residues 1-3 and 38-40 in native lysozyme. Though this is stabilised by only two hydrogen bonds, the amides of Phe38, Asn39 and Thr40 have protection factors exceeding $10^{4.5}$, whilst that of Phe3 is about 10^3. Clearly participation in hydrogen bonding is of fundamental importance in determining the rate of hydrogen exchange. The majority of amides not involved in elements of secondary structure exchange rapidly under these conditions and an upper limit for their protection must be about 10^2. Significant exceptions to this relationship are, however, found to exist. Four amides (Cys64, Leu75, Cys76 and Ile78) are involved in a region of the protein which is irregular in structure yet are also highly protected from exchange (by factors greater than $10^{4.5}$). Two of these amides (Leu75 and Cys76) form main chain hydrogen bonds in the native enzyme. The exchange properties of the remaining two amides cannot, however, be rationalised in this manner, Ile78 forms only a long and weak hydrogen bond to the sidechain of Asn74 and Cys64 is not identified in the X-ray structure as forming a hydrogen bond at all.

Proximity of the amide nitrogen to surface water has also been used to rationalise the exchange properties of amides in the native states of several proteins (Kossiakoff, 1982). The overall relationship between the rate of exchange and surface proximity, estimated as the distance to the nearest accessible water molecule (Pedersen et al., 1991) is shown in Figure 2. The six amides (Trp28-Ala32, inclusive and Cys64) most deeply buried in the core of the enzyme are amongst those with the slowest rates of exchange. On the other hand several amides (e.g. Ala11 and Lys97) lie in regions of secondary structure quite close to the surface of the enzyme and yet exchange quite slowly, while a number of others (Thr40, Gln41, Ser50, Thr51 and Arg112) lie at intermediate distance from the nearest accessible water molecule yet have some of the fastest rates of exchange. The hydrophobic core of lysozyme is formed essentially from the side chains of Leu17, Tyr20, Tyr23, Trp28, Ile98, Met105, Trp108 and Trp111. Interestingly, the amides of only two of these residues (Trp28 and Ile98) show

protection factors exceeding 10^5 and both of these lie in helical elements of structure and are quite deeply buried from the solvent. Two of the remaining six amides (Tyr20 and Met105) exchange at a rate too fast to measure under the conditions employed. Involvement of the side chain in a hydrophobic environment, apparently, is not itself sufficient to retard significantly the rate of hydrogen exchange of the backbone. The pattern of amide protection also shows no obvious correlation with the distribution of hydrogen bond energies, the B-factors estimated from the crystal structure or the temperature dependence of the amide hydrogen chemical shifts in the NMR spectrum (Itzhaki, Redfield & Dobson, unpublished data). The results indicate, therefore, that the rate of exchange of the majority of amides cannot be rationalised merely by either consideration of the pattern of hydrogen bonding in the native structure or distance from the surface of the molecule alone. However, consideration of both the distribution of hydrogen bonds and burial of the amide from solvent can account for the exchange behaviour of virtually all of the amides studied here.

References

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

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

Creighton, T.E. (1991) Curr. Biology 1, 8-10.

Delepierre, M., Dobson, C.M., Selvarajah, S., Wedin, R.E. & Poulsen, F.M. (1983) J.Mol.Biol. 168, 687-692.

Deonier, R.C. & Williams, J.W. (1970) Biochem. 9, 4261-4267.

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

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

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

Kossiakoff, A.A. (1982) Nature 296, 713-721.

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

Miranker, A., Radford, S.E., Karplus, M. & Dobson, C.M. (1991) Nature 349, 633-636.

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.

Roder, H. (1989) Meth.Enzym. 176, 446-473.

Roder, H., Wagner, G. & Wüthrich, K. (1985) Biochem. 24, 7396-7407.

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

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

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

Woodward, C.K. & Hilton, B.D. (1980) Biophys.J. 32, 561-575.

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

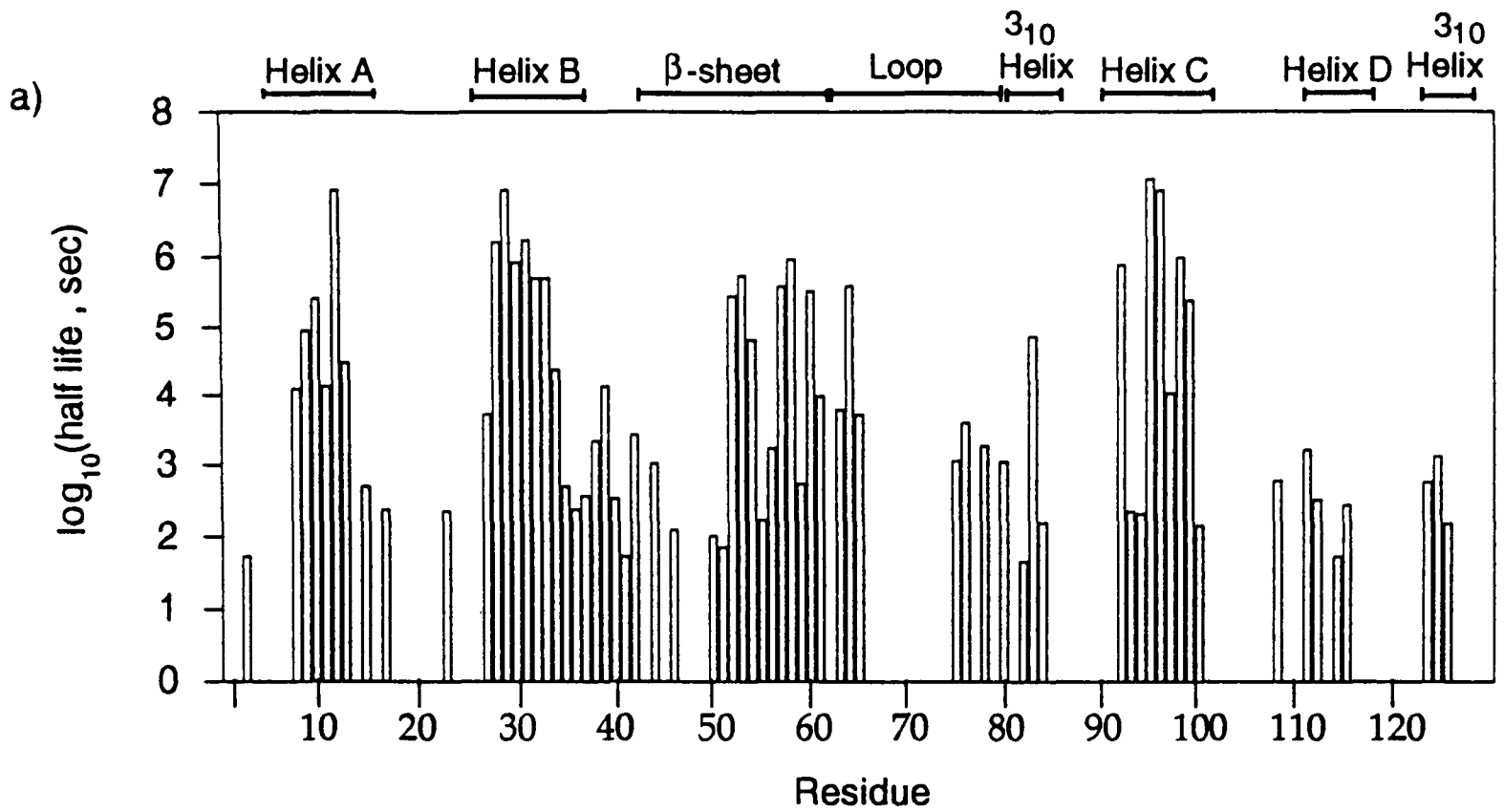


Figure 1
Profile of a) the half life for hydrogen exchange versus the protein sequence.
b) the protection factors versus the protein sequence; each at pH 7.5, 30°C. The location of elements of structure in the native protein are shown. Note that the scale is logarithmic.

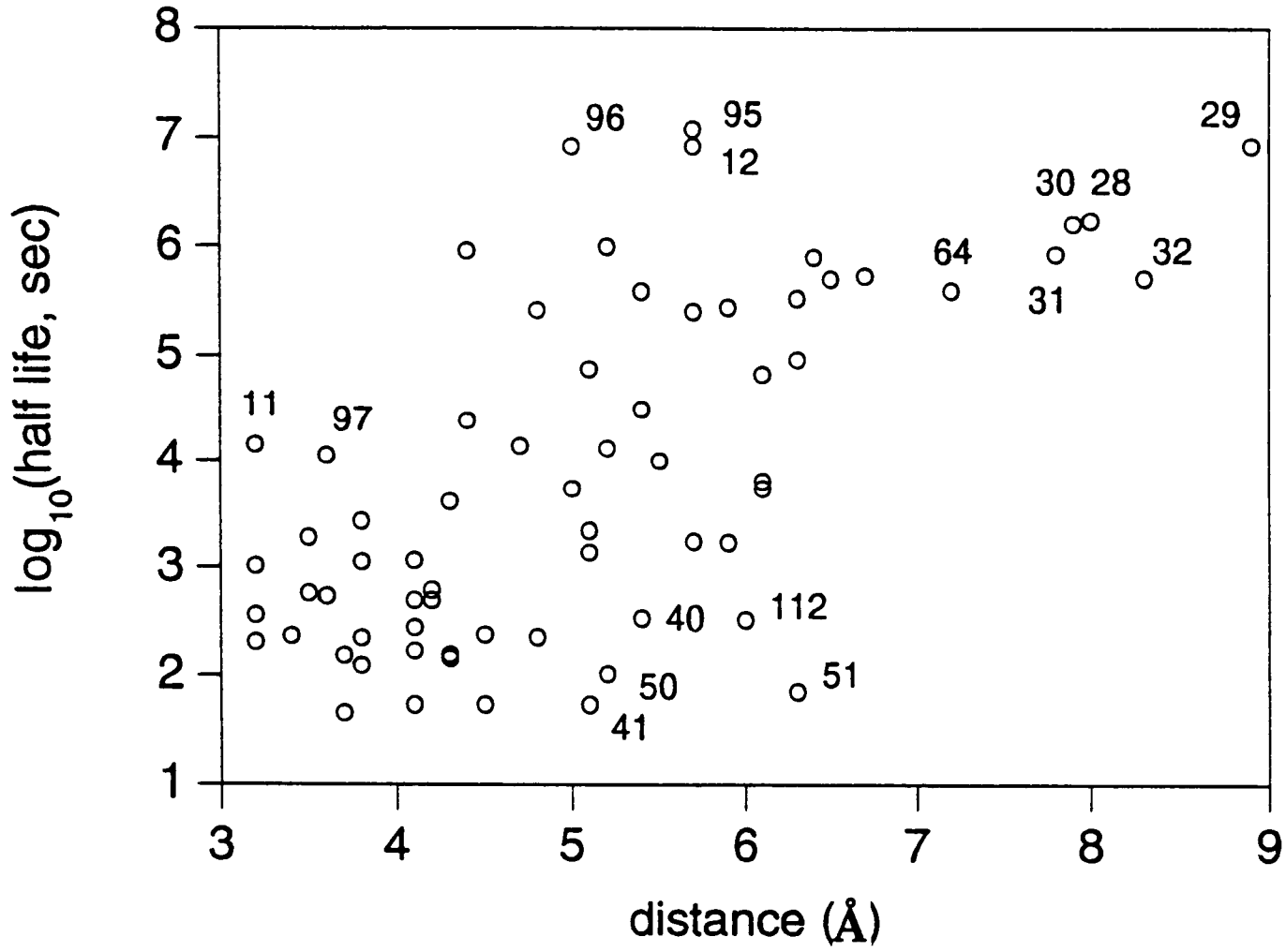


Figure 2
Relationship between the half life for amide exchange in the native state of lysozyme (pH 7.5, 30°C) and surface proximity, estimated as the distance of the individual amide nitrogen to the nearest accessible water molecule (Pedersen et al., 1991). Amides of particular interest are labelled.

Chapter 3

Amide Hydrogen Exchange in a Highly Denatured State: Hen Egg-White Lysozyme in Urea.

Summary

The amide hydrogen exchange behaviour of hen egg-white lysozyme denatured in 8 M urea has been studied at pH 2.0, 20°C. The observed exchange rates have been compared with those predicted for the same residues in a random coil conformation using recently derived parameters for sidechain inductive and temperature effects on exchange catalysis. The protection factors for exchange obtained in this way were found to be close to unity, with 41 of the 61 residues that could be followed having protection factors less than 2. No protection factor was greater than 5. In addition, previous data for hen lysozyme denatured thermally and for a three disulphide derivative, CM^{6-127} lysozyme, denatured at pH 2.0 have been reanalysed using the new reference parameters, and the protection factors were found to be similar to those of hen lysozyme denatured in 8 M urea. Thus, although ^1H NMR and far uv CD spectroscopy suggest that considerable deviations from random coil behaviour occur in these denatured states, such residual structure is insufficient to protect amides significantly against hydrogen exchange. This behaviour contrasts with that of a partly folded state of hen lysozyme denatured in trifluoroethanol (Chapter 4) and with that of the molten globule state of a homologous protein, guinea pig α-lactalbumin. Here protection factors for many amides exceed 30 and are located in continuous regions of the amino acid sequence, indicating the presence of persistent structure. The study of hydrogen exchange in substantially denatured states of a protein, therefore, provides a basis for the interpretation of protection factors in partially folded states.

Introduction

The relationship between the amino acid sequence, the structure and the dynamic properties of the native conformations of proteins has long been subject of intensive study. Recently, the study of denatured and partially folded states has been of increasing interest, since determination of their structure and stability may provide critical insight into the mechanisms of protein folding (Kim & Baldwin, 1990; Creighton, 1991; Dobson, 1991, 1992). Investigations have focused on the use of denaturants, such as temperature, urea, guanidinium hydrochloride or extremes of pH, which in certain instances, disrupt the overall cooperative fold of the protein but allow substructures to persist (Dill & Shortle, 1991; Ptitsyn, 1992). It has been postulated that such residual structures may resemble folding intermediates or serve as initiation sites for protein folding (Wright et al., 1988; Moult & Unger, 1992). The importance of states that are substantially denatured is now being recognised as a reference for thermodynamic (Privalov et al., 1989) and kinetic studies (Lu & Dahlquist, 1992) of protein folding. Specifically, in the case of urea, it has been suggested that unfolding, even at low pH, is far from complete and dependent on the amino acid sequence (Pace et al., 1990, 1992; Evans et al., 1991; Neri et al., 1992b). For these reasons, the mechanism and effectiveness of urea denaturation of protein structures is a topical issue of widespread importance (Cheek & Lilley, 1988; Makhatadze & Privalov, 1992; Sjipkes et al., 1993).

Evidence for non-random structure in denatured states of proteins has usually been inferred from their macroscopic properties (Dubin et al., 1973; Privalov et al., 1988), as the heterogeneity of these states at equilibrium makes their detailed characterisation extremely difficult. Recently, however, progress in NMR techniques has allowed partial site specific characterisation of several denatured states by magnetisation transfer techniques, hydrogen exchange measurements and heteronuclear multidimensional NMR (Dobson & Evans, 1984; Baum et al., 1989; Roder, 1989; Evans et al., 1991; Robertson & Baldwin, 1991; Radford et al., 1992a; Neri et al., 1992a,b). This has resulted in direct evidence for the nature of the interactions persisting in such states and their relationship to structure in the native conformation.

Amide hydrogen exchange with solvent deuterons is strongly dependent on their involvement in stable hydrogen bonds, although non-hydrogen bonded amides can be protected from exchange as a result of their burial from solvent. Therefore, the observation of slowed amide exchange is a powerful and site

specific probe for detecting persistent structure in proteins (Roder, 1989; Creighton, 1991; Woodward, 1993). Hydrogen exchange studies have made particularly significant contributions in this respect to protein folding, since the technique can be applied to a great variety of conformational states of a protein, including partially folded states under equilibrium conditions and transient kinetic intermediates formed during the folding process itself (Baum et al., 1989; Hughson et al., 1990; Jeng & Englander, 1991; Radford et al., 1992b; Baldwin, 1993 & references cited therein). A number of hydrogen exchange studies of more highly denatured proteins have also been carried out (Wagner et al., 1985; Robertson & Baldwin, 1991; Radford et al., 1992a; Lu & Dalquist, 1992; Neri et al., 1992a). In these studies significant apparent deviations in the rates of hydrogen exchange from the random coil behaviour anticipated from data on small model peptides (Molday et al., 1972) have been detected. It is not clear, however, whether these are the result of residual structure in these denatured states or whether apparent protection from exchange arises from inaccuracies in the parameters used for the calculation of protection factors (Jeng et al., 1991; Radford et al., 1992a; Robertson & Baldwin, 1992). These parameters have recently been revised and extended (Bai et al., 1993), providing the impetus to examine more closely hydrogen exchange protection data for a variety of protein states, under different experimental conditions.

Here, we report hydrogen exchange data for the highly denatured state of hen lysozyme in 8 M urea at low pH. These data are compared with the intrinsic exchange rates calculated with the revised parameters (Bai et al., 1993), taking into account the effect of urea on the intrinsic exchange rates (Loftus et al., 1986). Our previously published data for protection in other denatured states of lysozyme are also recalculated and compared with the extent of protection observed in the urea denatured state. The significance of these results for the characterisation of structure and dynamics in denatured states of proteins is discussed.

Materials and Methods

Hen egg-white lysozyme was obtained from the Sigma Chemical Company and was extensively dialysed at pH 2.0, lyophilised and stored at -20°C. CM^{6-127} lysozyme was prepared according to Radford et al., (1991). Ultrapure urea (BRL) was deuterated by repeated lyophilisation from 99.8% D_2O. CD measurements were made on a Jasco J-720 spectropolarimeter. A 1 mm thermostatted cell was used throughout. The concentration of lysozyme samples was typically 10 μM. Molar residue ellipticities (Θ) (deg.cm^2.dmol^{-1}res^{-1}) are reported. Spectra in urea could only be acquired down to 214 nm due to its high absorbance below this wavelength.

Amide exchange from 1.5 mM lysozyme in 8 M urea-d_4/D_2O was initiated by dilution from a 10 mM stock solution of the enzyme which had been denatured in 8 M urea-h_4/H_2O for 45 min at 20°C and was allowed to proceed for various lengths of time (1.5 to 50 min) at pH 2.0, 20°C. The protein was then rapidly refolded by 8-fold dilution into 20 mM deuterated sodium acetate buffer in D_2O, final pH 3.8 at 4°C. Protein samples were washed with the sodium acetate-d_3/D_2O buffer and concentrated using centricon C-10 microfilters. Phase sensitive COSY spectra were acquired at 35°C and 256 increments over 2k complex points and 16 transients were collected on a GE/omega 500 spectrometer belonging to the Oxford Centre for Molecular Sciences. The total acquisition time was 3.5 hr. Amide hydrogen intensities were taken as the sum of the absolute values of the four phase sensitive components, scaled to the cross peaks of the nonlabile ring hydrogens of Tyr23 and Tyr53 and were normalised to the intensities measured for a freshly dissolved sample of hen lysozyme in D_2O at pH 3.8 and 35°C acquired under identical conditions. Intrinsic exchange rates (k_{int}) predicted for amides in an unstructured polypeptide chain were calculated taking into account near neighbour inductive effects on acid and water catalysis (Bai et al., 1993) and were corrected for urea effects on the intrinsic exchange rates, using the factors proposed by Loftus et al., (1986).

Results and Discussion

Traditionally, far uv circular dichroism (CD) has been used to measure the overall secondary structural content of peptides and of native and denatured proteins. Figure 1 shows a comparison of the far uv CD spectrum of native lysozyme at pH 2.0 with three denatured states of the enzyme. The small magnitude of the ellipticity of hen lysozyme denatured in 8 M urea at pH 2.0, 20°C suggests that little, if any, regular secondary structure is present in this state. The far uv CD spectrum of the protein denatured thermally (pH 2.0, 64°C), however, deviates significantly from the spectrum expected for a random coil polypeptide and nearly 60% of the native signal intensity persists at 222 nm. Reduction of one of the four disulphide bridges results in a further denatured state of hen lysozyme at pH 2.0 (Radford et al., 1991) with significant signal intensity at 222 nm. The results indicate, therefore, that the far uv CD spectra of the three different denatured states of lysozyme studied here are distinct, despite the fact that all these states are substantially denatured as judged by the lack of an appreciable near uv CD signal (Radford et al., 1991; Chapter 4), but that the urea denatured state is closest to an ideal random coil as judged by CD.

These conclusions are consistent with the results of calorimetry on some of these denatured states, which show that at least the majority of the global structure is disrupted (Privalov et al., 1989; Makhatadze & Privalov, 1992). However, a significant degree of residual chemical shift dispersion is evident in the ^1H NMR spectra of these denatured states (Evans et al., 1991; Radford et al., 1991), and reactivity to flavin dyes in photo-CIDNP experiments is substantially reduced relative to that of unstructured peptides (Broadhurst et al., 1991). Both these findings indicate that significant non-random interactions persist involving aromatic residues in hydrophobic clusters or partially collapsed states. The observed chemical shift perturbations and photo-CIDNP effects are relatively diminished in the urea denatured state, suggesting that the residual structure is reduced in the presence of urea. However, even under these conditions, the behaviour is still significantly perturbed relative to that expected for a random coil state (Broadhurst et al., 1991; Evans et al., 1991).

In order to examine further the nature of these highly denatured states, amide hydrogen exchange measurements have been carried out for lysozyme denatured in 8 M urea. After dilution (8 fold) of the protonated denatured protein in 8 M urea/H_2O into 8 M urea-d_4/D_2O, decay of amide intensity in ^1H

NMR spectra is complete within 1 hr, even at pH 2, 20°C. For this reason, and because of the difficulty in making site-specific assignment of resonances in denatured states, an urea concentration jump technique was employed to measure the exchange rates of individual amides in the urea denatured state. Thus, exchange was allowed to occur in D_2O at pH 2.0 for defined periods of time in presence of 8 M urea. Exchange was then quenched by diluting the urea to a final concentration of 1.0 M urea and raising the pH to 3.8, which induced rapid refolding to the native conformation. After concentration and washing of the protein, the extent of exchange of individual amides as a function of time in the urea denatured state was then obtained from the intensities of the well resolved and assigned cross-peaks in the ^1H NMR spectrum of the native protein.

Cross-peak intensities for 61 of the 127 amide hydrogens in hen lysozyme were monitored by this urea jump method, and were measured at nine time points ranging from 1.5 to 50 minutes. The resulting exchange profiles were fitted to single exponential decays. Half-lives of exchange determined for the urea denatured state are shown in Figure 2a (and exchange rates are listed in Appendix 2-I); values of protection factors, defined as the ratio of the experimentally determined rates to the rates predicted for the same amides under identical conditions but in a random coil conformation, are given in Figure 2b. The latter were obtained from tabulated data derived from measurements of exchange behaviour in unstructured model peptides (Bai et al., 1993) and were corrected for the effect of urea (Loftus et al., 1986).

These results indicate that the degree of protection of amides against hydrogen exchange in the urea denatured state of hen lysozyme is generally very low (Figure 2b). Protection factors are all less than 5, and indeed for the 41 of the 61 measured amides are less than 2. Although it is difficult to interpret the significance of the marginal protection factors observed in the urea denatured state, the pattern of residues with protection factors exceeding 2 is essentially discontinuous in sequence suggesting that little, if any, regular secondary structure able to protect against exchange persists in this state. This is consistent with the observation from the far uv CD and ^1H NMR data that hen lysozyme is substantially denatured in 8M Urea at low pH at 20°C. Apparent residual protection may be a consequence of experimental errors in addition to the remaining uncertainties in the parameters used to estimate intrinsic reference exchange rates especially since these corrections are of a considerable magnitude (illustrated by comparing patterns of Figure 2a with 2b).

It is particularly interesting in this regard that if protection factors are calculated with the earlier, less complete, set of model compound data (Molday et al., 1972), a much greater variation in protection factors results with values for 8 amides exceeding 10. This emphasises the improvements made to this analysis by the recent refinement of the reference parameters (Bai et al., 1993). Of particular importance in the new data set is the fact that parameters for the inductive effect on hydrogen exchange of valine, leucine, isoleucine and methionine have now been measured individually and it is no longer necessary to assume their effect is identical with that of alanine. These sidechain substituents were found to slow their own amides three to eight times in their acid catalysed exchange rate relative to alanine, suggesting that significant reinterpretation may be necessary for hydrogen exchange protection factors published for polypeptide regions rich in these hydrophobic residues. For example, in the extreme case of two sequence adjacent isoleucine residues, hydrogen exchange is retarded by a factor of up to 32 at pH 2.0 for the second, C-terminal residue relative to the case when two alanines are adjacent.

In the light of these results, and in order to compare amide protection of lysozyme in 8 M urea with hydrogen exchange measurements reported previously for the thermally denatured states of lysozyme and the acid denatured state of $CM^{6\text{-}127}$-lysozyme (Radford et al., 1992a), the latter were reanalysed using the newly calibrated parameters of Bai et al., (1993). This reanalysis results in a significant reduction in the apparent protection factors for several residues (e.g. that of Ile78 is reduced by a factor of 4 and 4.5 in the thermally and acid denatured states respectively). Apparent protection for several other residues, however, is markedly increased on recalculation (e.g. Asn65 by a factor of 2.8 and 5.4 in the thermally and acid denatured states, respectively). The average protection factors are therefore only slightly altered from previous values. However, the detailed pattern of protection has changed, most significantly hydrophobic residues are no longer those with the highest protection factors (Table Ia). The average protection factors (Table Ib) are, however, somewhat greater for thermally denatured lysozyme and acid denatured $CM^{6\text{-}127}$-lysozyme than for lysozyme in 8 M urea. This is consistent with the greater degree of residual structure in these states compared with the urea denatured state as indicated by CD and NMR data. The significance of this result, however, must be treated with caution given the different solution conditions. It is interesting to note, that a number of residues with marginally elevated protection factors are in close proximity to disulphide bonds and their

presence could be responsible at least in part for the magnitude of the protection factors observed. Most importantly, given that all three states can be regarded as substantially denatured by other criteria, is the fact that the hydrogen exchange rates measured under these diverse solution conditions are overall approximated well by the model random coil parameters.

The success of the random coil approximation for these denatured states is further evident when compared with the protection factors for the denatured state of hen lysozyme that is formed in 50% TFE at low pH (Chapter 4) (Figure 3). and with similarly, recalculated protection factors for amides in the acid denatured "molten globule" state of guinea pig α-lactalbumin (Chyan et al., 1993) (Table Ia). In accord with this, the ellipticity in the far uv CD is substantially greater for this state and for the TFE state of lysozyme than for the other denatured states of lysozyme considered here. Moreover, the highest protection factors are found in stretches corresponding to the centres of native-like α-helices in these partially folded states. Taken together, these data show that these partly folded states are clearly distinct from the substantially unfolded states of lysozyme, both in the magnitude and in the pattern of hydrogen exchange protection.

Concluding remarks

Hydrogen exchange is dominated by the most labile structures which present amides to environments where they are accessible for exchange. Persistency of structure is a stringent requirement for the observation of significant protection; for example a protection factor of 5 corresponds, on the simplest of models, to 80% occupancy of a structured state affording full protection, rapidly interconverting with one or more states providing none. In native states of proteins protection factors are frequently greater than 10^6, while in kinetic refolding intermediates they can exceed 500 (Radford et al. 1992a,b; Baldwin, 1993).

In the present study, protection factors in the acid, thermally and urea denatured states of lysozyme are found to be of very limited magnitude, generally less than 5, when appropriately corrected for sequence, temperature and cosolvent effects. Amides with protection factors exceeding this value are scattered over the amino acid sequence. This is consistent with the view that these states are substantially denatured and that any residual structure, as indicated to different extents by far uv CD and ^{1}H NMR in these denatured

states, is fluctuating rapidly. Other techniques, such as multidimensional and heteronuclear NMR may be able to detect such residual structures as small populations of locally ordered conformations in exchange with substantially unstructured states (Neri et al., 1992b; Kemmink et al., 1993).

The results of this study strongly support the value of amide hydrogen exchange for identifying the presence of persistent elements of structure and show that characterising the exchange behaviour of amides in substantially denatured states provides a baseline for the interpretation of hydrogen exchange behaviour in other states. The recently revised reference parameters for rates of exchange in random coil models have been extremely important in this regard. Data on the urea denatured state of hen lysozyme suggests that experimental protection factors less than 5 can be expected for a substantially denatured protein. Given that protection found in equilibrium molten globules, frequently used as models for folding intermediates, are often less than 100, this finding has considerable significance for the interpretation of hydrogen exchange rates in these states. It suggests that, particularly when associated with a regular pattern of protection, as in the TFE state of lysozyme, the molten globule state of α-lactalbumin and in other proteins (Hughson et al., 1990; Jeng & Englander, 1991), values of protection factors even as low as 10 may be sufficiently reliable to detect stable structural regions in these states.

References

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

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

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

Broadhurst, R.W., Dobson, C.M., Hore, P.S., Radford, S.E. & Rees, M.L. (1991) Biochem. 30, 405-412.

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

Cheek, P.J. & Lilley, T.H. (1988) J.Chem.Soc, Faraday Trans. I, 84, 1927-1940.

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

Creighton, T.E. (1991) Curr.Biology 1, 8-10.

Dill, K.A. & Shortle, D. (1991) Ann.Rev.Biochem. 60, 795-825.

Dobson, C.M., Evans, P.A. & Williamson, K.L. (1984) FEBS Lett. 168, 331-334.

Dobson, C.M. (1991) Curr.Opin.Struct.Biology 1, 22-27.

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

Dubin, S.B., Feher, G. & Benedek, G.B. (1973) Biochem. 12, 714-720.

Evans, P.A., Topping, K.D., Woolfson, D.N. & Dobson, C.M. (1991) Proteins: Struct.Funct. & Genetics 9, 248-266.

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

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

Kemmink, J., van Meirlo, C.P.M., Scheek, R.M. & Creighton, T.E. (1993) J.Mol.Biol. 157, 331-355.

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

Loftus, D., Gbenle, G.O., Kim, P.S. & Baldwin, R.L. (1986) Biochem. 25, 1428-1436.

Lu, J. & Dahlquist, F.W. (1992) Biochem. 31, 4749-4756.

Makhatadze, G.I. & Privalov, P.L. (1992) J.Biol.Mol. 226, 491-505.

Molday, R.S., Englander, S.W., Kallen, R.G. (1972) Biochem. 11, 150-158.

Moult, J. & Unger, R. (1991) Biochem. 30, 3816-3824.

Neri, D., Wider, G. & Wüthrich, K. (1992a) Proc.Natl.Acad.Sci.(USA) 89, 4397-4401.

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

Pace, C.N., Laurents, D.V. & Thomson, J.A. (1990) Biochem. 29, 2564-2572.

Pace, C.N., Laurents, D.V. & Erickson, R.E. (1992) Biochem. 31, 2728-2734.

Privalov, P.L., Tiktopulo, E.I., Venyaminov, S.Y., Griko, Y.V., Makhatadze, G.I. & Khechinashvili, N.N. (1989) J.Mol.Biol. 205, 737-750.

Ptitsyn, O.B. (1992) in "Protein Folding", T.E. Creighton Ed, Freeman & Company, N.Y..

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

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

Radford, S.E., Dobson, C.M. & Evans, P.A. (1992b) Nature 358, 302-305.

Robertson, A.D. & Baldwin, R.L. (1991) Biochem. 30, 9907-9914.

Roder, H. (1989) Meth.Enzym. 176, 446-473.

Sijpkes, A.H., v.d. Kleut, G.J. & Gill, S.J. (1993) Biophys.Chem. 46, 171-177.

Wagner, G., Roder, H. & Wüthrich, K (1985) Biochem. 24, 7407-7411.

Woodward, C. (1993) Trends Biol.Sci. 18, 359-361.

Wright, P.E., Dyson, H.J. & Lerner, R.A. (1988) Biochem. 27, 7167-7175.

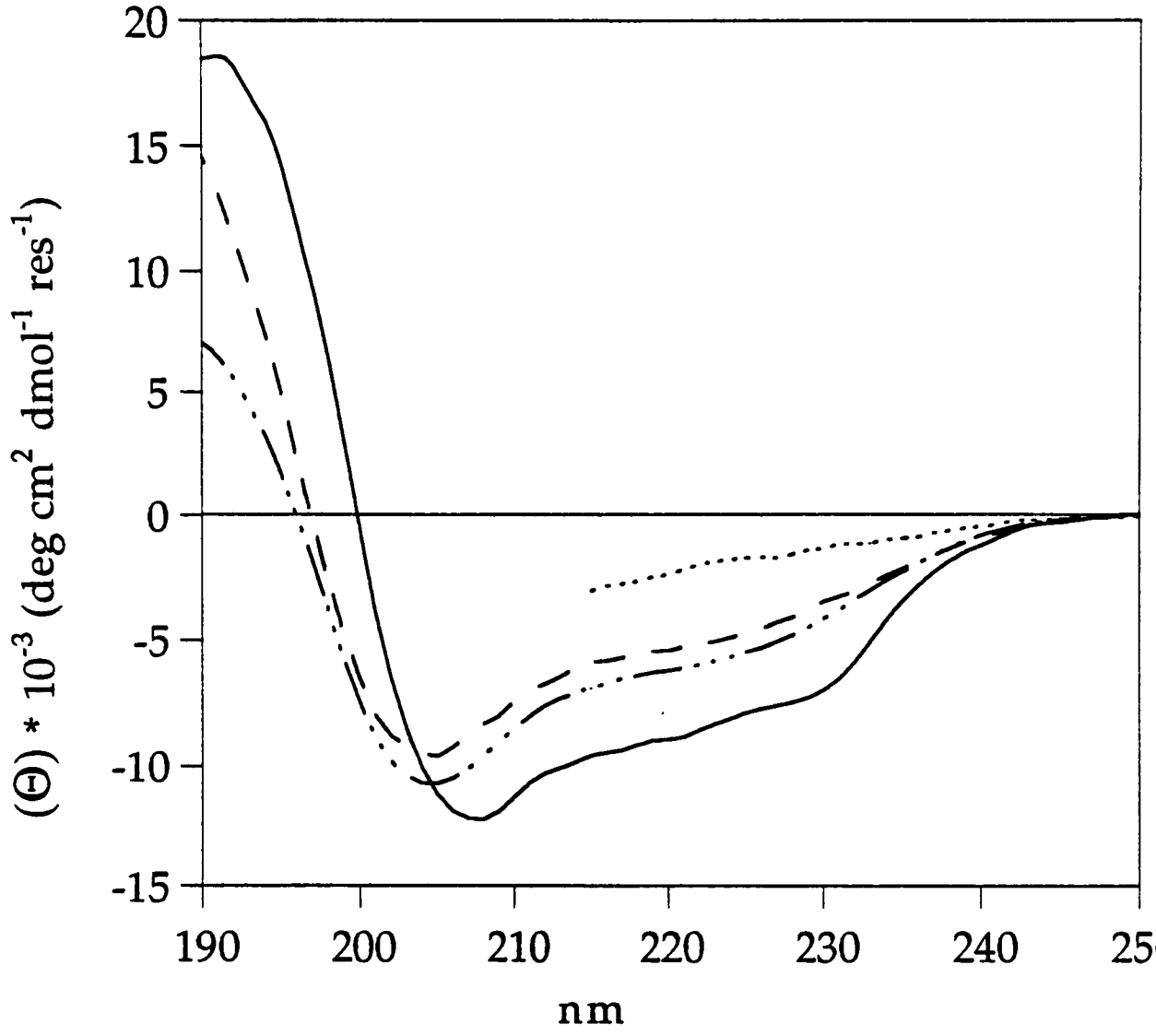
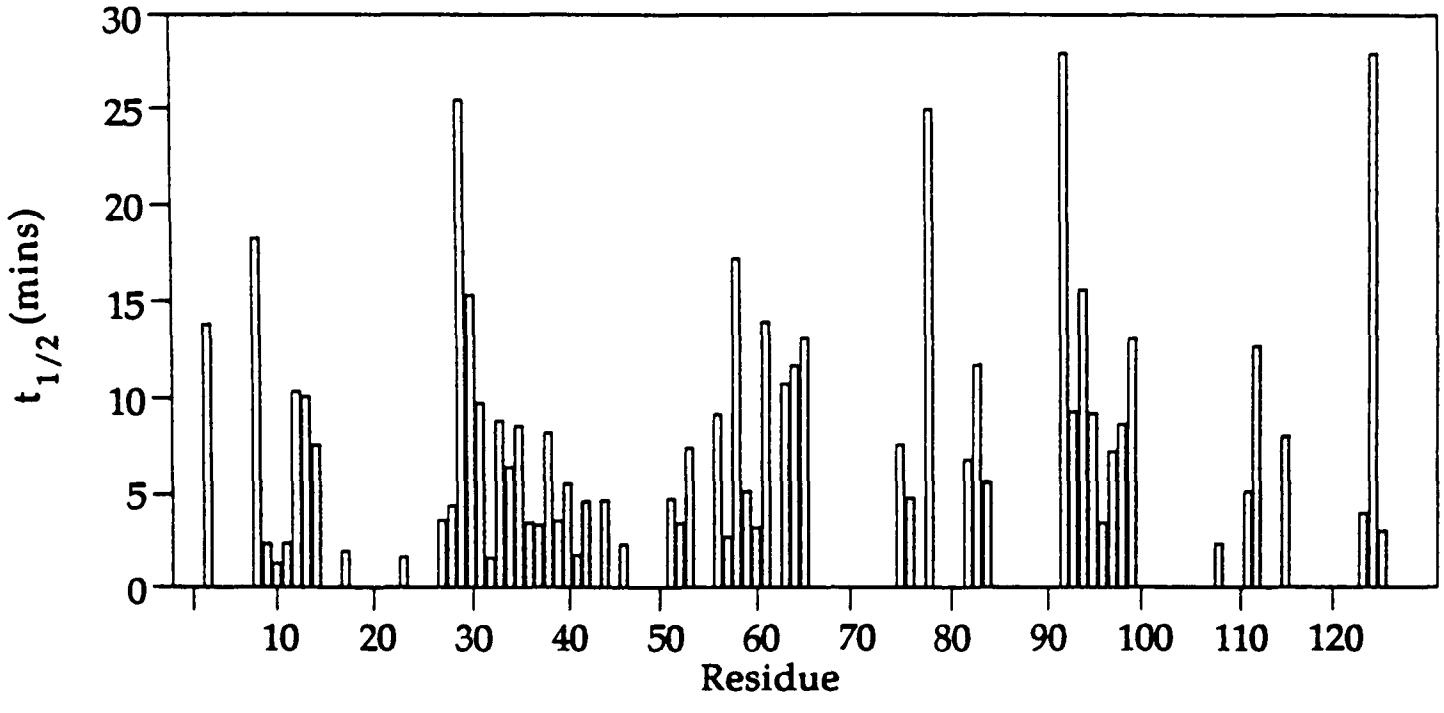


Figure 1
Far uv CD spectra of hen lysozyme in different states at pH 2.0 in D_2O, (——) the native state at 20°C, (— —) the thermally denatured state at 64°C, (- - - -) the urea denatured state at 20°C, 8M urea and (- - -——) the acid denatured state of CM^{6-127} lysozyme at 35°C.

a)



b)

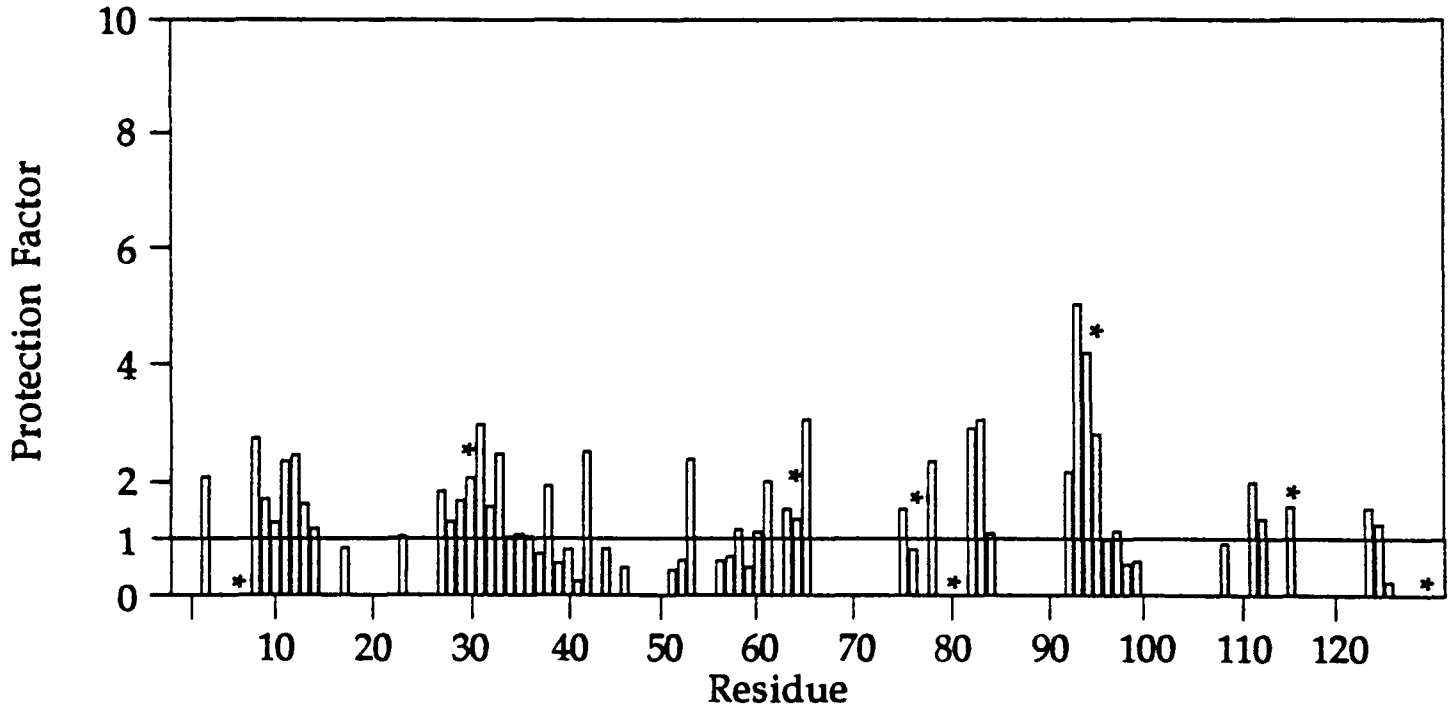


Figure 2

Profile of a) half lives and b) protection factors for the exchange of amide protons from the urea denatured state of hen lysozyme (8M urea, 0.1M NaCl, 50mM phosphate buffer at pH 2.0, 20°C) versus the protein sequence. The location of cysteine residues, forming disulphide bridges 6-127, 30-115, 64-80 and 76-94 are indicated (*).

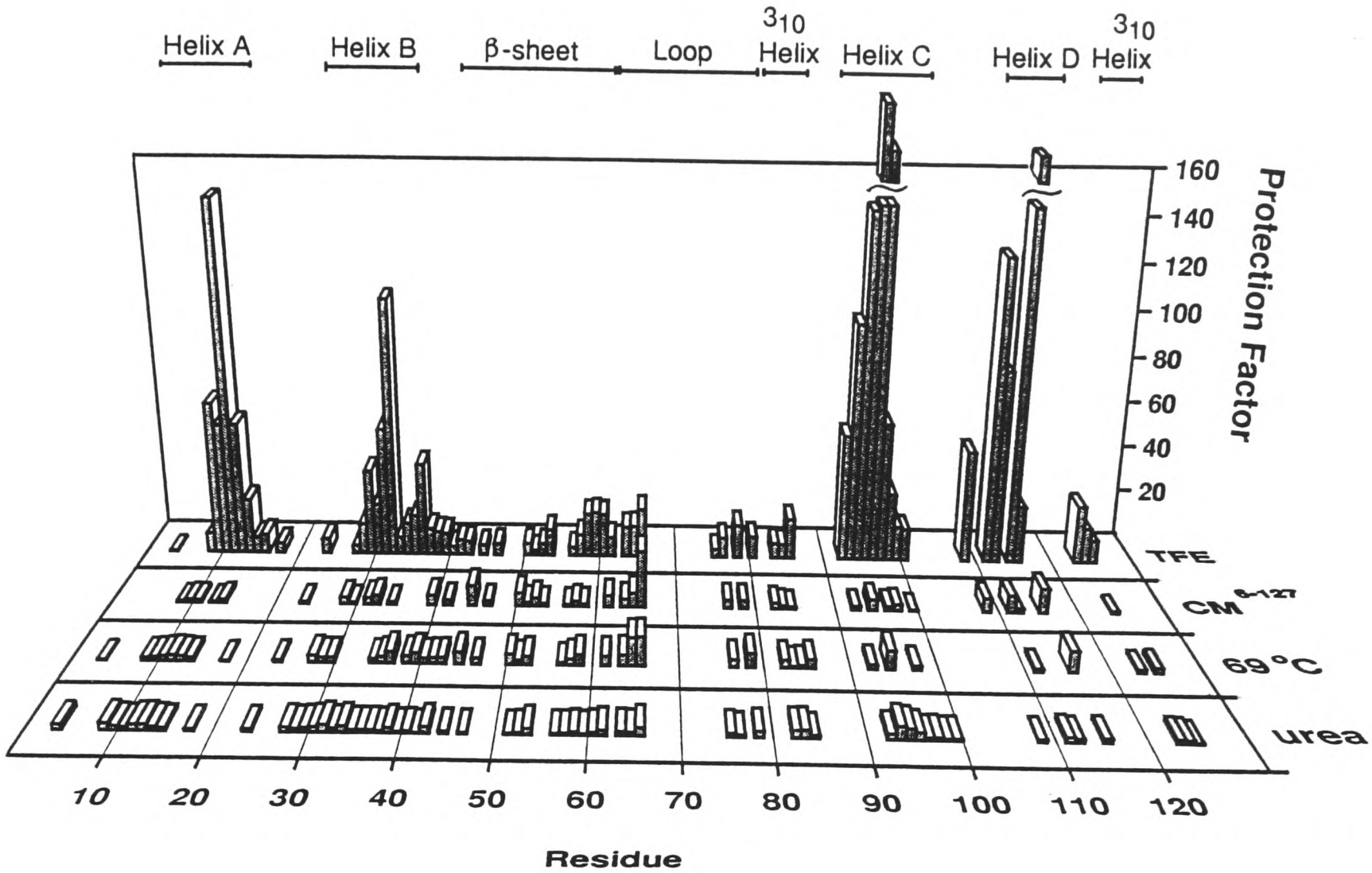


Figure 3

Protection factors for amide exchange in a partly folded state of hen lysozyme in 50% trifluoroethanol (TFE)/D_2O (v/v) at pH 2.0, 27°C (Chapter 4) are shown together with protection factors in the substantially denatured states of lysozyme, plotted on the same scale. Protection factors of residues Ala95, Lys96 and Arg114 exceed this scale with values of 450, 215 and 185 respectively. The location of elements of secondary structure in the native protein is shown. No correction for the effect of TFE on instrinsic rates were applied as preliminary examination of model peptides suggests that the effect is within a factor of 3 at low pH (Chapter 4). Protection factors of the thermally denatured state of hen lysozyme at 69°C, pH 3.8 and the acid denatured CM^{6-127} lysozyme at pH 2.0, 35°C were recalculated using the published rate constants (Radford et al., 1992a) together with the revised parameters of Bai et al., (1993).

Table I: Protection factors in different conformational states of hen lysozyme and in the acid denatured state of guinea-pig α-lactalbumin

a) Individual values.

HEWL Residue		log_{10} Native 30°C, pH7.5	8M urea 20°C, pH2.0	Thermal 69°C, pH2.0	CM^{6-127} 35°C, pH2.0	50% TFE 27°C, pH2.0	GPLA Residue		A-state 25°C, pH2.0
Phe	3	3.1	2.1			0.5			
Leu	8	5.1	2.7	1.6	0.9	8.4			
Ala	9	6.5	1.7	1.8	1.5	65			
Ala	10	7.1	1.3	2.3	1.3	56			
Ala	11	5.9	2.4	2.7		160			
Met	12	8.6	2.5	2.2	0.5	56			
Lys	13	6.3	1.6	1.6	1.6	23			
Arg	14		1.2			6.7			
His	15	4.5				7.9			
Leu	17	3.7	0.8	0.5		4.2			
Tyr	23	4.0	1.1	1.0	1.0	6.5			
Asn	27	6.1	1.8	3.0		3.4			
Trp	28	7.8	1.3	2.2	3.0	13	Trp	26	> 58
Val	29	7.8	1.7	2.3	0.8	37	Leu	27	12
Cys	30	8.1	2.1			22	Cys	28	210
Ala	31	8.4	3.0		2.9	55	Ile	29	41
Ala	32	7.4	1.6		4.0	110	Ile	30	> 8
Lys	33	7.4	2.5			7.3	Phe	31	> 21
Phe	34	6.0	1.0	1.8	1.1	14			
Glu	35	4.0	1.1	2.8		17			
Ser	36	4.3	1.0	4.8		40			
Asn	37	5.1	0.7			11			
Phe	38	5.1	1.9	3.2		9.8			
Asn	39	6.4	0.6	4.3	4.3	9.3			
Thr	40	4.5	0.8	2.2		3.4			
Gln	41	3.7	0.3	2.5	2.3	6.5			
Ala	42	5.4	2.5	2.6		5.8			
Asn	44	5.4	0.8	4.9	8.6	4.3	Val	42	2.9
Asn	46	4.5	0.5	2.7	2.1	5.0			
Ser	50	4.3		4.4	5.8	5.1			
Thr	51	3.8	0.5	1.9	1.6	2.0			
Asp	52	7.1	0.6	3.5	3.7	5.0			
Tyr	53	7.0	2.4		1.1	11			
Gly	54	6.9							
Ser	55	3.4							
Leu	56	4.2	0.6	1.1	0.9	3.6			
Gln	57	7.2	0.7	2.6	2.4	9.4			
Ile	58	7.2	1.2	4.3	1.0	19			
Asn	59	4.7	0.5			19			
Ser	60	7.9	1.1			18			
Arg	61	6.1	2.0	3.9	4.8	8.3			
Trp	63	5.0	1.5	4.4	3.4	13			
Cys	64	7.8	1.3	12	6.2	13	Cys	61	2.9
Asn	65	6.4	3.1	12	24	21	Glu	62	3.7
Leu	75	4.5	1.5			3.5			
Cys	76	5.7	0.8	3.2	3.6	8.6			
Asn	77						Asp	74	48
Ile	78	4.6	2.3	5.7	3.5	14	Ile	75	3.1
Cys	80	5.1				9.2			
Ala	82	3.7	2.9	3.2	2.6				
Leu	83	6.0	3.1	0.7	1.5	6.4			
Leu	84	3.1	1.1	1.2	1.0	6.0			
Ser	85			3.5		17			
Val	92	7.2	2.2	1.9	1.8	54	Ile	89	9.0

a) Individual values (continued)

HEWL Residue		$\log_{10}$ Native 30°C, pH7.5	8M urea 20°C, pH2.0	Thermal 69°C, pH2.0	CM^{6-127} 35°C, pH2.0	50% TFE 27°C, pH2.0	GPLA Residue		A-state 25°C, pH2.0
Asn	93	4.4	5.0			100	Met	90	4.1
Cys	94	4.9	4.2	7.7	4.7	150	Cys	91	16
Ala	95	9.3	2.8			450			
Lys	96	8.6	1.0		1.8	220	Lys	93	22
Lys	97	5.9	1.1	1.8	2.7	60	Lys	94	21
Ile	98	7.1	0.6			30	Ile	95	31
Val	99	6.2	0.6		0.4	15	Leu	96	9.8
Ser	100	4.1				14			
Asp	101						Asp	97	10
Gly	102						Ile	98	7.4
Trp	108	4.1	0.9		6.0	49			
Trp	111	4.5	2.0	1.4	7.4	130			
Arg	112	4.2	1.4		1.9	83			
Arg	114	3.9				180			
Cys	115	4.9	1.6	8.6	11...	25			
Trp	123	4.1	1.5	1.4		25			
Ile	124	4.0	1.3		1.1	15			
Arg	125	3.8	0.2	2.1		9.8			

b) Averaged by regions of secondary structure in the native state.

2° structure	Residues	Native	urea	69°C	CM^{6-127}	TFE	A-state GP α-lac
A-helix	8-15	$10^{5.6}$	1.9	2.0	1.2	380	--
B-helix	24-36	$10^{6.3}$	1.7	2.8	2.3	32	60
C-helix	88-99	$10^{6.7}$	2.2	3.8	2.3	120	12
D-helix	108-115	$10^{4.3}$	1.5	5.0	6.4	94	--
3_{10} (I)	79-84	$10^{4.5}$	2.4	2.2	1.8	10	--
3_{10} (II)	120-124	$10^{4.0}$	1.0	1.7	1.1	17	--
β-sheet	41-60	$10^{5.5}$	1.0	3.1	3.0	9	--
cluster	60-65,74-8	$10^{5.7}$	1.8	7.3	3.6	12	15

The number of amide residues averaged differ for the different states, depending of availability of data. Protection factors for amides in the acid denatured, molten globule state of guinea-pig α-lactalbumin were recalculated as described above, using the data of Chyan et al., (1993). HEWL, hen egg-white lysozyme; GPLA, guinea pig α-lactalbumin; CM^{6-127}, lysozyme carboxymetylated at Cys6 and Cys127.

Chapter 4

A Partially Folded State of Hen Egg-White Lysozyme in Trifluoroethanol: Structural Characterisation and Implications for Protein Folding

Summary

The effect of 2,2,2-trifluoroethanol (TFE) on the solution conformation of hen egg white lysozyme has been investigated using circular dichroism (CD) and ^1H nuclear magnetic resonance (NMR) spectroscopy. Addition of TFE to lysozyme at pH 2.0, 27°C up to a concentration of 15% (v/v) induces only slight changes in the NMR spectrum. However, above this concentration a cooperative transition to a new, but partially structured state of the protein is observed. This state shows no structural cooperativity against further denaturation and is characterised by an ellipticity in the far uv CD greater than that of the native protein. Near uv CD intensity is dramatically reduced compared with that of the native state, and ^1H NMR studies indicate that sidechain interactions are substantially averaged in this denatured state. Solvent proton/deuterium exchange rates for 66 amide hydrogens were measured site specifically by a combination of amide trapping experiments and 2D ^1H NMR. Significant protection from exchange occurs for about 25 backbone amides, the majority of which are located in regions of the protein that are helical in the native enzyme. By contrast, amides located in a second region of the native protein which contains a β-sheet, one 3_{10}-helix as well as a long loop show little protection. This pattern of protection resembles that found in the stable molten globule state of α-lactalbumin and in an early kinetic intermediate detected in the refolding of hen lysozyme.

Introduction

The characterisation of conformational transitions and folding intermediates is central to the study of protein folding (Creighton, 1990; Kim & Baldwin 1990), but the detailed description of partially folded proteins is often difficult due to the high cooperativity of folding. Using mild denaturing conditions, however, partially folded states have been found for a number of proteins that are stable at equilibrium. The nature and degree of structure observed in these states can vary widely; some are largely disordered whilst others exhibit substantial conformational preferences (Christensen & Pain 1991; Dill & Shortle, 1991; Dobson 1992). A major objective is, therefore, to characterise these stable partially folded states and to relate their structures to those inferred for transient folding intermediates from kinetic studies.

Studies of lysozyme and the homologous protein, α-lactalbumin, have been particularly important in this regard. A stable, compact molten globule state is observed during equilibrium denaturation of α-lactalbumin, by contrast, the unfolding transition of lysozyme is a highly cooperative event in which only the native and highly unfolded states are stable at equilibrium (Dolgikh et al., 1981; Evans & Dobson, 1984; Baum et al., 1989). Despite this stopped-flow CD experiments have suggested that similar intermediates are populated during kinetic refolding of the two proteins. In addition NMR studies on hydrogen protection have shown that the stable molten globule state of α-lactalbumin resembles an intermediate populated transiently during the refolding of hen lysozyme (Kuwajima et al., 1985; Miranker et al., 1991; Radford et al., 1992b). There is, therefore, considerable incentive to try to generate stable partially folded states of lysozyme at equilibrium using different denaturing conditions.

A second strategy used to overcome the cooperativity of protein folding is to study protein fragments (Wright et al., 1988) and peptide models (Staley & Kim, 1990). Characterisation of the conformation of such peptides in aqueous solutions is, however, often complicated by the low stability and dynamic behaviour of these molecules (Dyson & Wright, 1991). Consequently, structure stabilising cosolvents, principally trifluoroethanol, are commonly employed to stabilise peptide conformations in solution (Sanford et al., 1991; Bruch et al., 1991). Recently it has been shown in several peptide systems that TFE does not induce secondary structure indiscriminantly, but regions of the polypeptide chain that are helical in the native state are stabilised preferentially (Nelson & Kallenbach, 1989; Lehrman et al., 1990; Segawa et al., 1991; Sönnichsen et al., 1992; Dyson et

al., 1992). Little work has, however, been carried out on the effect of TFE on the conformation of intact proteins. Whether TFE stabilises native-like structures or alters or induces novel structures, therefore, is a fundamentally important question.

In the light of these issues we have studied the effect of TFE on the structure of intact hen lysozyme in its native and denatured states. Evidence from CD for the denaturation of this protein to a partially structured state in aqueous TFE solution has previously been reported (Galat, 1985). Here we report NMR and CD evidence on the nature of the equilibrium denaturation transition of hen lysozyme in TFE and a detailed characterisation of the structural features of its partially denatured state.

Materials and Methods

Materials: Hen egg white lysozyme and the tetrapeptides Glu-Gly-Tyr-Arg and Ala-Gly-Ser-Glu were obtained from the Sigma Chemical Company. Lysozyme was dialysed extensively at pH 2.0, lyophilised and stored at -20°C; 99% 2,2,2-TFE-d_3 was purchased from Cen Saclay (France). The exchangeable deuterium on TFE-d_3 was exchanged by distillation from H_2O to allow for a steady deuterium lock on TFE in 2D ^{1}H NMR experiments. TFE-h_3 was purchased from Aldrich.

Circular dichroism (CD): CD measurements were made on a Jasco J-720 spectropolarimeter. A 1 cm cell was used throughout. The concentration of lysozyme samples was typically 2-4 μM for far and 20-40 μM for near uv CD studies. The pH of each sample was adjusted to 2.0 after addition of TFE or urea. Molar ellipticities Θ (deg.cm^2.dmol^{-1}.res^{-1}) are reported.

^{1}H NMR spectroscopy: ^{1}H NMR spectra were acquired at 27°C on 500 and 600 MHz Bruker AM spectrometers belonging to the Oxford Centre for Molecular Sciences. One dimensional NMR spectra were acquired over 8k complex points using a sweep width of 7246.4 and 6024.2 Hz, respectively. NOESY spectra were acquired in the phase sensitive mode using a mixing time of 200 ms; 1k increments of 4k complex data points were collected. The F1 dimension was zero-filled twice and resolution enhancement with double exponential window functions was applied in both dimensions. Protein concentrations were typically 2.5 mM. The pH was adjusted for individual samples in TFE and is uncorrected for isotope and TFE effects.

Amide hydrogen/deuterium exchange.

a) Bulk exchange experiments: The lyophilised enzyme (final concentration 2.5 mM) was dissolved in 50% TFE-d_3/D_2O and the pH rapidly adjusted to 2.0. 1D NMR spectra were acquired immediately at 27°C at intervals of 4.5 min over a period of 12 hr. At chosen time points the protein was refolded by 6-fold dilution into 20 mM deuterated sodium acetate buffer in D_2O, pH 3.8. Spectra were acquired immediately and 4 hr after the dilution. Amide hydrogen intensities were integrated between 8.80 and 7.75 p.p.m. and were normalised to the area of the 59 aromatic hydrogens (7.75-6.60 p.p.m.) acquired for the same sample, but in which all of the amide hydrogens had been exchanged with deuterons by heating the sample to 58°C for 15 min.

b) Selective bulk exchange of rapidly exchanging amide hydrogens: Lysozyme was fully deuterated by dissolution at pH 2.0 in 99.9% D_2O and heating to 80°C for two periods of 20 min. The protein solution was cooled and lyophilised. Deuterated lysozyme was then dissolved at 2.5 mM in H_2O, pH 2.0, 27°C for 2 hr. During this procedure selective exchange of amide deuterons that are weakly protected in the native state occurs for solvent hydrogens. Following a second lyophilisation step to remove H_2O, protein samples were dissolved in either D_2O or 50% TFE-d_3/D_2O at pH 2.0, 27°C and exchange of amide hydrogens was monitored as described above. In this case spectra were acquired at intervals of 1.5 min.

c) Site-specific measurement of exchange rates from the TFE state: Hydrogen exchange of lysozyme in 50% (v/v) TFE-d_3/D_2O was allowed to proceed for various lengths of time (3-810 min) at pH 2.0, 27°C, as previously described for the bulk exchange experiments. After various periods of exchange, the protein was rapidly refolded by 6-fold dilution into D_2O, pH 3.0 to the native state and the protein solution was immediately lyophilised. Samples were redissolved, to a final concentration of 5 mM, in 20 mM deuterated sodium acetate buffer in D_2O, pH 3.8 and phase sensitive COSY spectra were acquired at 35°C (Redfield & Dobson, 1988). A total of 256 increments over 2k complex points and 16 transients were collected; the total acquisition time was 3.5 hr. Amide proton intensities were taken as the sum of the absolute values of the four phase sensitive components and were scaled to the C3,5H-C2,6H cross peaks of the nonlabile protons of Tyr23 and Tyr53. Amide proton occupancies were normalised to those measured for a sample of hen lysozyme freshly dissolved in D_2O at pH 3.8 and 35°C acquired under identical conditions. Hydrogen exchange rates (k_{ex}) were fitted as single exponential decays (and are listed in Appendix S-I). Intrinsic exchange rates (k_{int}) predicted for amides in an unstructured polypeptide chain were calculated taking

into account near neighbour inductive effects on acid and water catalysis (Bai et al., 1993). Protection factors (k_{int}/k_{ex}) uncorrected for TFE cosolvent effects are reported together with exchange half lives.

d) Hydrogen exchange in tetrapeptides in 50% (v/v) TFE-d_3: The effect of TFE upon intrinsic exchange rates of unprotected amides was estimated on the basis of two unstructured tetrapeptides, Gly-Gly-Tyr-Arg and Ala-Gly-Ser-Glu. Peptides were lyophilised from H_2O or 50% TFE-h_3/H_2O at pH 2.0. Each peptide was then dissolved (8 mg/ml) at pH 2.0 in D_2O or 50% TFE-d_3/D_2O and exchange was monitored by recording 1D NMR spectra at intervals of 1 min at 27°C. The decreases in amide resonance intensities were fitted to single exponential decays.

Results

Perturbations induced by TFE in the native state of hen lysozyme.

Figure 1 shows near and far uv CD spectra of lysozyme in different concentrations of TFE. In the presence of 15% (v/v) TFE the far uv CD spectrum of the native enzyme is unperturbed, indicating that this concentration of TFE apparently has little effect on the secondary structure of the protein. Small intensity changes in the near uv CD are observed under these conditions and are consistent with association of alcohol molecules with the native enzyme (Kurimatsu et al., 1972). In accordance with this, analysis of 1D and 2D ^1H NMR spectra shows that 15% TFE-d_3/H_2O (v/v) induces little change in chemical shift (< 0.08 p.p.m.) for the majority of resonances (Figure 2). For a limited set of resonances, however, significant changes in chemical shift (> 0.08 p.p.m.) are observed. Of particular interest are the resonances of the indole NH hydrogens of Trp63, Trp108 and Trp111 as well as the amide resonance of Asn66 which experience changes in chemical shift similar to those observed in the presence of $(GlcNAc)_3$ or urea (Lumb & Dobson, 1992). These shifts can be attributed to the binding of TFE molecules to the active site of the native enzyme. Other shifts are consistent with TFE molecules binding at surface sites, many of which have been previously identified as binding sites for ethanol molecules (Lehmann et al., 1986). Taken together, these results indicate that the structure of the native enzyme is not significantly affected by the change of solution conditions at low TFE concentrations.

Nature of the denaturation transition of lysozyme induced by TFE.

Figure 3a shows the ellipticity at 222 nm as a function of the concentration of TFE. At concentrations of TFE exceeding 15% (v/v) a substantial increase in ellipticity at 222 nm is seen, consistent with cooperative restructuring of the protein molecules to a new state, which is characterised by a negative ellipticity much greater than that of the native enzyme. The midpoint of this transition occurs at about 20% (v/v) TFE. A further, more gradual, increase of ellipticity is observed as the concentration of TFE is then raised further to about 80% (v/v). In the near uv CD a transition at about 20% (v/v) TFE is also observed but in this case involves a substantial loss of ellipticity.

Dramatic changes are also observed in the 1D ^{1}H NMR spectrum as the concentration of TFE is raised above 15% (v/v) (Figure 2). Resonances characteristic of denatured proteins appear with a concomitant decrease of native peak intensities. For example, two resonances for the C2H proton of the single histidine residue (His15) are observed, one at the chemical shift of the native protein and the other at a near random coil chemical shift. No appreciable line broadening is evident in the spectrum and the resonances of the native and denatured states are in slow exchange. Peak intensities of the resonances of the native and denatured protein, therefore, can be related directly to the population of these states. Figure 4 shows the change in intensity of several resolved resonances in the 1D ^{1}H NMR spectrum over the transition region of 15-25% (v/v) TFE. The resonances followed arise from residues in different parts of the protein structure, their intensity change is coincident with each other and with the changes in near and far uv CD, indicating that denaturation is, to a good approximation, a simultaneous and global event. Dilution of TFE to concentrations below 20% (v/v) in similar experiments demonstrated that the transition is fully reversible under these conditions.

Characterisation of the TFE state of hen lysozyme by CD and NMR.

A concentration of 50% (v/v) TFE-d_3 at pH 2.0 and 27°C was chosen for structural studies on the TFE denatured state, based on the observation that the conformational transition is virtually complete at this concentration of TFE as judged by CD and NMR spectroscopy. This state is referred to as the "TFE state". The significant increase of negative ellipticity in the far uv CD at 222 nm during the denaturation transition is in sharp contrast to the decrease in signal intensity observed during denaturation of lysozyme by urea (Figure 3a). The substantial increase in ellipticity in the TFE state is quite remarkable; in 50% (v/v) TFE an ellipticity more than twice that in the native enzyme is achieved. By contrast, the

intensity of the near uv CD spectrum is substantially reduced relative to that of the native enzyme (Figure 1, inset). A near-linear decrease in ellipticity is observed in the far uv CD on heating the TFE state (Figure 3b). This again contrasts with the highly cooperative transition observed for thermal denaturation of hen lysozyme at pH 2.0 in H_2O solution (Radford et al., 1991). The cooperativity of structural stabilisation in the TFE state, therefore, is substantially reduced giving rise to a broad melting curve over the range of temperatures studied. In accord with this, the 1D ^1H NMR spectrum of lysozyme in the TFE state (Figure 2) is characterised by the absence of significant chemical shift dispersion and substantial resonance overlap. Resonances for five of the six tryptophan indole hydrogens are, however, resolved in the 1D ^1H NMR spectrum and resonances arising from the ring protons of one tryptophan and a second aromatic residue are significantly perturbed from random coil values (Figure).

The 2D NOESY spectrum of the TFE state shows a significant number of NOE's, which may arise from intra- as well as inter-residue through-space interactions (Figure 5). Assignment of these resonances to specific residues could not be made by conventional methods, however , because of the poor chemical shift dispersion in the spectrum. It is, nonetheless, informative to compare this spectrum with those obtained for the native state (Redfield & Dobson, 1988) and the denatured protein in 8 M urea (Figure 5). By comparison with spectra of the native state, NOEs in the spectrum of the TFE state are fewer in number, much more poorly dispersed and of lower intensity. A substantial number of NOEs are also seen in the spectrum of the enzyme denatured in 8 M urea. In this case the resonances are even less well dispersed and are weaker in intensity than seen in the spectrum of the TFE state. The chemical shifts and linewidths of resonances in the spectrum of the TFE state are sensitive to temperature and to TFE concentrations exceeding 50% (v/v). There is no significant effect of protein concentration on these parameters. Linewidths, particularly those arising from indole and amide protons increase with decreasing temperature or increasing TFE concentration. This suggests that temperature and TFE concentration may effect a change in the dynamics or the compactness of the TFE state.

Characterisation of the TFE state of hen lysozyme by hydrogen exchange methods.

The exchange time course of all amides from the TFE state of hen lysozyme is shown in Figure 6a. The profile is clearly multiphasic suggesting that some amides exchange significantly more slowly than others. Integration of the amide intensity in the 1D ^1H NMR spectra gives an estimate of the sum total of amide hydrogens that remain with full or partial intensity. After 9 min an intensity

corresponding to approximately 84 of the original 127 amides remains. Under the conditions of our experiments (pH 2.0, 27°C) the half life for exchange of amides in unstructured model peptides is predicted to be about 3-4 min (Bai et al., 1993) and, calculated for a random coil of the length and sequence of hen lysozyme, an amide intensity corresponding to 32 protons would be expected to remain after this time. Exchange of some or all amides, therefore, is much slower than expected. A similar exchange profile was obtained when the concentration of lysozyme was diluted five-fold, confirming that the slowly exchanging amides in the TFE state do not arise as a result of intermolecular association or aggregation.

To examine the effect of TFE on the intrinsic rate of hydrogen exchange, exchange in two unstructured tetrapeptides (Gly-Gly-Tyr-Arg and Ala-Gy-Ser-Glu) in D_2O and in 50% (v/v) TFE-d_3/D_2O were measured at pH 2.0, 27°C. For these two peptides the rate of exchange of individual amides in 50% TFE (v/v) is within a factor of three of that found in D_2O. Change in the intrinsic exchange rates is, therefore, an unlikely explanation for the considerable retardation of exchange rates observed in the TFE state.

To examine whether amides protected in the TFE state are also protected in the native state of hen lysozyme, the TFE denatured protein was rapidly refolded after different periods of exchange by 6-fold dilution into 20 mM deuterated sodium acetate buffer in D_2O, pH 3.8. Figure 7 shows the 1D ^1H NMR spectrum before and after this solvent jump after 2 hr of exchange. It is apparent that the great majority (> 80%) of the amide intensity is transferred into the native state. Little exchange occurred subsequently over a period of 4 hr from the native conformation indicating that the majority of the amides concerned are also strongly protected from exchange in the native protein. It is, therefore, possible to identify the amides which are protected from exchange in the TFE state by jumping the conditions to those favouring rapid refolding to the native conformation and examining the amide cross peak intensities in phase-sensitive COSY spectra. In this manner, the exchange of individual amide hydrogens in the TFE state can be measured using the known assignments of the native protein (Redfield & Dobson, 1988).

Amide intensities for 66 of the 126 amide hydrogens in hen lysozyme were measured at 9 time points ranging from 3 to 810 min of exchange in the TFE state. The resulting exchange profiles were fitted to single exponential decays and the half-lives of exchange are shown and compared with those predicted for amides in unstructured model peptides in Figure 8. Substantial protection from exchange

(up to 200 fold) is observed for amides in the TFE state. Such large protection factors cannot be accounted for by the effect of TFE on the intrinsic rates of hydrogen exchange. In accord with this, 11 amides that are highly protected in the native state exchange at rates close to their intrinsic exchange rates in the TFE state. In addition, while the exchange rates for amides of 8 asparagine residues could be measured only 1 of these were significantly retarded in the TFE state. Similarly, rates of exchange for 7 of the 12 alanine residues could be measured but only 6 were significantly protected in the TFE state. Most significant in this regard, however, is the observation that amides located in several continuous stretches of the polypeptide sequence are protected from exchange. This suggests that these form regions with persistent structure in the TFE state.

The protection factors for amides in the TFE state fall qualitatively into two groups (Figure 9). Significantly protected sites (protection factor exceeding approximately 35) comprise amides in the four α-helices, the short 3_{10}-helix at the C-terminus of the protein, all located in the α-domain in the native protein (Figure 9). By contrast, the majority of amides which are protected only weakly lie in the central portion of the polypeptide chain which comprises a triple stranded β-sheet, a 3_{10}-helix and a long loop, or are located in loop regions which link helical elements in the α-domain. Exchange rates calculated for the 66 individual amide hydrogens may be used to examine whether there is protection of amides in the TFE state in regions not protected in the native enzyme. A bulk exchange profile may be constructed by summation of the exchange rates of the individual amides followed. This profile considers only the exchange of the 66 amides which could be measured in the TFE state because they are protected the native state. At times greater than 100 min this curve is essentially coincident with that obtained for all 126 amides measured from the bulk exchange experiment in Figure 6a. Thus, amides most highly protected in the TFE state are also protected in the native protein. However, at shorter periods of time (< 100 min) there is a significant divergence of the two curves. This could arise from amides that are protected in the TFE state and are only weakly protected or are unprotected in the native protein. The exchange of these amides was monitored selectively in a lysozyme sample in which all of the amides which are significantly protected from exchange in the native state had been selectively deuterated. The exchange profile of the remaining amide hydrogens is shown in Figure 6b, and is compared with the exchange profile calculated for an equivalent number of amides in a random polypeptide of length and sequence composition of hen lysozyme. While several amides which are weakly protected or are unprotected in the native state are also weakly or moderately protected in the TFE state, an upper limit of 2-5 can be

placed on the sum total of amides that are significantly protected only in the TFE state. Addition of this profile to that calculated for amides protected in both the TFE state and the native protein results in a profile essentially identical to the measured bulk exchange.

The rate of exchange of tryptophan indole hydrogens in the TFE state is also rapid relative to the rates in the native enzyme (Wedin et al., 1982). Even at 10°C in 50% TFE (v/v) these protons are fully exchanged within 3 min. The rate of exchange for tryptophan indole protons in unstructured peptides under these conditions is about 200 min^{-1} (Nakanishi et al., 1978). Ignoring possible effects of TFE on intrinsic rates, the protection factor of these residues in the TFE state cannot exceed about 300.

Discussion

Denaturation in TFE.

Far and near uv CD and analysis of 2D ^1H NMR spectra has shown that the conformation of native lysozyme is not significantly perturbed by concentrations of TFE up to about 15% (v/v) at pH 2.0, 27°C (Figures 1, 2 and 3a). Small changes in the near uv CD and in ^1H NMR chemical shift may be attributed to TFE molecules associating with the protein surface and the active site (Kuramitsu et al., 1972; Lehrman et al., 1986; Lumb & Dobson, 1992). Predenaturation of the protein surface and local multistate unfolding in TFE proposed by Galat (1985) were not seen under the conditions of our experiments. A major unfolding transition, however, occurs between 15 and 30% (v/v) TFE. This transition is cooperative, but increasing the concentration of TFE above 30% (v/v) causes a further increase in ellipticity in the far uv CD. Disruption of the secondary structure in the TFE state of hen lysozyme on increasing the temperature is characterised by a broad approximately linear transition suggesting that the TFE state may represent an ensemble of related conformational substates in dynamic equilibrium with each other.

Secondary structure in the TFE state.

The far uv CD spectrum of the TFE state of hen lysozyme is characterised by a negative ellipticity at 222 nm of more than twice that of the native enzyme (Figure 1). Ellipticity at this wavelength is thought to reflect predominantly α-helical structural content of proteins although aromatic residues and disulphide bonds could also contribute significantly to the far uv CD spectrum (Manning & Woody, 1989; Siligardi et al., 1992). Several possibilities for this increase of negative

ellipticity may be considered. The first possibility is that there is a significant positive contribution to the circular dichroism at 222 nm from aromatic residues or disulphide bonds in the native state of hen lysozyme; loss of such tertiary structure in the TFE state thus could cause an apparent increase in ellipticity at his wavelength. Alternatively, the α-helical content of the protein may increase in the presence of TFE either by extension of existing α-helices, by induction of new α-helices, or an increase in the regularity of the helices in the presence of reduced tertiary interactions (Manning et al., 1988). In the native enzyme approximately 44 of the 129 residues are involved in α-helical structures as judged by the magnitude of the ellipticity at 222 nm (Greenfield & Fasman, 1969). Indeed, 44 residues are part of α-helical conformations in the crystal structure of hen lysozyme (Handoll, 1985). CD studies on 3_{10}-helices suggest that these are commonly too short to contribute significantly to the ellipticity at 222 nm (Gans et al., 1991). To account for the large increase in negative ellipticity in TFE a further 36 residues of the polypeptide chain must be accommodated in α-helical structure. It is not clear how this could occur, although TFE could induce α-helical structure in the two native 3_{10}-helices (5 residues each), the β-sheet or the loop regions, particularly the long loop including residues 61-78. Interestingly, in human α-lactalbumin, the molten globule state also shows an increase (of about 40%) in negative ellipticity relative to that of the native state (Ewbank & Creighton, 1991), although in the molten globule states of several other proteins an ellipticity equal to or less than the native state is observed (Dolgikh et al., 1981; Hughson et al., 1990; Jeng & Englander, 1991).

Results similar to those described here for the TFE state of hen lysozyme have also been observed for other proteins in organic solvents (Harding et al., 1991; Fan et al., 1993; Cox et al., 1994). Indeed, several all β-proteins have been shown to adopt α-helical conformations in alcohols (Dufour & Haertle, 1991; Jackson & Mantsch, 1992; Lui et al., 1994) and it has been suggested that a peptide corresponding to the β-sheet region of lysozyme may adopt α-helical structure in TFE solution (Segawa et al., 1991). This would support the possibility that at least some additional α-helical structure may be induced in lysozyme in presence of TFE.

Through space information from nuclear Overhauser effects (NOEs) is extremely useful for identifying the presence of secondary structure in native proteins. In the case of the TFE state of lysozyme, however, the substantial chemical shift overlap in the ^{1}H NMR spectrum precluded ready assignment of resonances to individual residues. In addition, assignments could not be made using magnetisation transfer techniques (Evans et al., 1991) since conditions under which the native

and TFE states are interconverting on an appropriate timescale could not be found. Despite this, it is interesting to investigate the number and type of the NOEs observed in the NOESY spectrum of the TFE state and to compare these with spectra of other conformational states of lysozyme obtained under similar conditions (Figure 5). About 30 strong NH-NH NOEs are observed in the spectrum of the TFE state. This compares with 66 NOEs of this type observed in the spectrum of the native protein. Considering the extent of helix inferred by CD, this number is much less than expected. However, NOEs may overlap, or be weak or missing from the spectrum of the TFE state, for example, because of its high degree of conformational averaging. Furthermore, the presence of a substantial number of NH-NH NOEs cannot, in isolation, be taken as evidence for α-helical structure (Dyson & Wright, 1991). Indeed, a similar number of NH-NH NOEs are also observed in the spectrum of the protein denatured in 8 M urea (Figure 5), a state that appears substantially unfolded when examined by far and near uv CD (Figure 1). In the latter case, however, the NOEs are less well dispersed and of much lower intensity than in the spectrum of the TFE state and presumably arise from very localised conformational preferences in the denatured protein. This is supported by the reduced number and intensity of NH-sidechain NOEs observed in the spectrum of the urea denatured protein (Figure 5). The present NMR studies cannot provide a detailed description of structure in the TFE state; they are, however, consistent with the presence of at least some residual structure in this denatured state of lysozyme.

A third technique widely used to detect persistent secondary structure in peptides and partially folded or native proteins involves the measurement of the rate of amide hydrogen exchange (Roder, 1989). Amide hydrogens involved in secondary structure or deeply buried within the protein core show relatively slow exchange with solvent deuterons (Woodward, 1982; Englander & Kallenbach, 1984). The profile of bulk exchange measured for all of the amides in the TFE state resembles closely that calculated from the measured exchange rates of the 66 amide hydrogens for which site-specific rates of exchange could be determined (Figure 6a). This clearly demonstrates that the large majority of amides strongly protected from exchange in the TFE state are also protected from exchange in the native enzyme. It follows that any novel structure in the TFE state (as viewed by far uv CD) does not significantly protect amides from exchange. By contrast, about 25 amides are significantly protected in both the native and the TFE states. Interestingly, the β-sheet of lysozyme shows only marginal protection from exchange in the TFE state despite the fact that some structure is stabilised by TFE in a peptide fragment corresponding to this region in the native enzyme (Segawa

et al., 1991). Clearly, in the TFE-state of the protein any non-native structure in the β-region of the protein in the TFE state is not as persistent as structure in the α-helical subdomain.

The largest protection factors of amides in the TFE state are about 150. These values are much smaller than those for the same amides in the native protein, which range from 10^4 to 10^8 (Radford et al., 1992a; Chapter 2). They also contrast with the exchange behaviour observed for amides in the acid denatured state of $CM^{6\text{-}127}$lysozyme and for lysozyme denatured thermally or by 8 M urea; in these cases the exchange rates for the majority of amides resemble closely those expected for an unstructured polypeptide (Radford et al., 1992a; Chapter 3, Table 1a). The magnitude of protection factors in the TFE state are similar to those reported for the molten globule states of several proteins (Baum et al., 1989; Hughson et al., 1990; Jeng et al., 1991; Chyan et al., 1993; & Chapter 3). The pattern of protection in the TFE state resembles, at least in outline, that of the molten globule state of α-lactalbumin (Baum et al., 1989; Chyan et al., 1993) and a transient intermediate detected in the refolding of hen lysozyme (Miranker et al., 1991; Radford et al., 1992b). A comparison of the protected sites in the TFE state and the major refolding intermediate is shown in Figure 9. Notably, the β-sheet region of lysozyme and the 3_{10}-helix involving residues 80-85 are not appreciably protected in any of these three states. Several similarities and differences between these states may also be noted. The four α-helices and the C-terminal 3_{10}-helix are all highly protected in the transient intermediate but the B-helix is more weakly protected in the TFE state. Interestingly, little α-helical structure is found in a peptide corresponding to the hydrophobic B-helix of lysozyme in TFE (Segawa et al., 1991). This may suggest that TFE stabilises α-helical structure only weakly in hydrophobic amino acid sequences (Lehrman et al., 1990; Segawa et al., 1991). In accord with this, a small cluster of residues involving Trp63, Cys64 and Ile78 which form a hydrophobic patch on the surface of the native lysozyme molecule is protected in the refolding intermediate but is not protected significantly in the TFE state.

Side chain ordering in the TFE state and its role in stabilising secondary structure.
The presence of significant residual chemical shift dispersion in the ^{1}H NMR spectrum of the TFE state suggests that a number of sidechain interactions persist in this state. Specifically, resonances of one tryptophan and possibly one other aromatic residue as well as resonances of several aliphatic sidechains are significantly perturbed from random-coil chemical shift values. Dispersion of several resonances has been observed in spectra of lysozyme denatured thermally

and by urea (Evans et al., 1991) although the number of resonances with significant chemical shift perturbation and the magnitude of the induced shifts are smaller in the latter two states. The presence of a large number of intense NH-sidechain and NH-aromatic NOEs in the spectrum of the TFE state suggests that there are a number of long range interactions involving sidechains in this state. However, at least the majority of the structured regions in the TFE state must lack specific and persistent tertiary interactions involving aromatic sidechains which would give rise to an intense near uv CD spectrum. The non-persistent nature of sidechain interactions in the TFE state is supported by the observation that the protection factor of tryptophan indole protons and of amides located in regular elements of secondary structure is less than about 300. This is in contrast to the more substantial protection of these protons observed in the native protein and for one tryptophan and several amides residues in an early kinetic refolding intermediate of lysozyme (Radford et al., 1992b). This suggests that sidechain interactions in the TFE state are fewer in number and are less persistent than those in the native protein and the kinetic folding intermediate.

The role of sidechain interactions in stabilising structure against hydrogen exchange is difficult to define. In the native protein, sidechain interactions may influence the rate of amide exchange by burial of amides in the protein interior, by distortion of secondary structure in hydrophobic regions or by local tertiary restraints on dynamics (Radford et al., 1992a; Goodmann & Kim, 1991). In linear peptide systems favourable sidechain interactions are presumed to slow hydrogen exchange by stabilising secondary structure directly (Rohl et al., 1992). The situation in partially folded conformations of intact proteins is likely to be complicated and may be intermediate between these two extremes.

The four α-helices in the TFE state are not protected to equal extents; the C- and D-helices are protected more strongly than the A- and B-helices in this denatured state. By contrast, in the native protein, amides in A- and B-helices have some of the highest protection factors observed (Radford et al., 1992a). Interestingly, little structure is induced by TFE in a 32-residue peptide fragment of lysozyme corresponding to the C-, D- and C-terminal 3_{10}-helices in the native enzyme (Segawa et al., 1991). This suggests that long range interactions may contribute to the stability of this region of the polypeptide chain in the TFE state. Taken together, these observations suggest that at least a few sidechain interactions persist in the TFE state but neither their extent, location nor dynamic behaviour can be inferred from the present experiments.

Implications for protein folding

It is now clear that during the refolding of proteins from denaturing media secondary structure, particularly helical structure, is formed rapidly, often within the first few milliseconds of folding (Kuwajima, 1989; Roder, 1989; Dill & Shortle, 1991; Radford et al., 1992b). This has been found to be the case with lysozyme where, interestingly in the light of the results in the present paper, the far uv CD signal overshoots the native ellipticity before recovering more slowly (Radford et al., 1992b). During refolding, collapse of the protein structure may be concomitant with, or may precede, such secondary structure formation. The environment produced by this collapse is likely to be hydrophobic in nature and may help to generate or stabilise secondary structure. If the interactions generated during the collapse are relatively non-specific, it is possible that the secondary structure formed in mixed organic/aqueous solvents may be related to that formed during the early stages of folding. If this is the case, the present finding that the most persistent regions of structure in the TFE state are a subset of those of the native structure could well be of significance in understanding protein folding. In the case of lysozyme, the majority of amides protected in TFE are in segments of the protein which are helical in the native state. This is consistent with studies of peptide fragments of proteins which show that helical structure is stabilised preferentially in regions helical in the native protein (Segawa et al., 1991; Dyson & Wright, 1992). The experimental strategy used here, however, does not enable us to determine directly whether the persistent structure in the TFE state is also helical but, particularly in the light of the far uv CD data, it is certainly consistent with this supposition.

One intriguing question raised in this work is whether non-native helices are generated in TFE. In the case of lysozyme, if new helices are formed they are clearly relatively unstable and do not protect the amides concerned from hydrogen exchange. This observation could be of relevance for folding because it suggests how pathways of folding might originate. For example, if native-like secondary structure is preferentially stabilised in non-aqueous environments, then there will be a higher probability that following hydrophobic collapse native-like higher structural elements or domains rather than non-native structures will form. In this manner, folding could be directed towards native rather than non-native structures. This does not exclude the possibility, however, that in some proteins formation of non-native structure could be significant in folding; it is possible, for example, that helical structure formed initially might subsequently be transformed into β-sheet structures later in the folding process. It is interesting in this regard that the denatured state of several predominantly β-proteins in organic solvents

contain significant α-helical structure (Dufour & Haertle, 1991; Jackson & Mantsch, 1992; Fan et al., 1993; Lui et al., 1994).

Support for the notion that denatured states such as that described here may be of significance in understanding protein folding comes from the relationship of the pattern of hydrogen exchange protection observed in the TFE state to that observed for an intermediate sampled during kinetic refolding of lysozyme from GuHCl (Figure 9). Furthermore, the pattern of exchange is also similar to that found in the stable partly folded (molten globule) state of the homologous protein α-lactalbumin formed at low pH, that has been related to transient intermediates of both proteins (Kuwajima et al.; 1985, Radford et al., 1992b). All of these states have highly protected amides centred in regions that form α-helices in the native state, but none in regions that form the β-sheet, an observation that has led to the idea of "folding domains" within the protein structure (Miranker et al., 1991; Radford et al., 1992b). The degree of protection from exchange (> 500 fold) is significantly greater in the kinetic intermediate of lysozyme than in the TFE state of this protein or in the acid denatured state of α-lactalbumin (Baum et al., 1989; Chyan et al., 1993; see Chapter 3). We believe, however, that this may well be due to the short lifetime of the former, leaving insufficient time for major conformational fluctuations that can expose amides and give rise to exchange prior to trapping in the folded state of the protein.

Although much remains to be learnt about proteins denatured in TFE, the finding in the present study that addition of this cosolvent stabilises predominantly native-like secondary structure in a partially folded state of an intact protein is of considerable significance. It provides support, for example, for the idea that studies of peptide fragments of proteins in TFE (Segawa et al., 1991; Dyson et al., 1992) can have direct relevance to understanding the structure and stability of protein folding intermediates. It also gives rise to the possibility that TFE might be of practical use in providing a favourable environment for the renaturation of proteins in high yields from their aggregated states; this is presently under investigation in our laboratory. Finally, in a more general context, the present results support the proposition that an increasingly wide range of stable denatured states can be generated, and that their detailed study and comparison of their properties with those of peptide fragments and transient kinetic intermediates should be of considerable value in efforts to understand mechanisms of protein folding on a structural level.

References

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

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

Cox, J.P.L., Evans, P.A., Packman, L.C., Williams, D.H. & Woolfson, D.N. (1994) J.Mol.Biol. 234, 483-492.

Dill, K.A. & Shortle, D. (1991) Ann.Rev.Biochem. 60, 795-825.

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.

Dufour, E. & Haertle,T. (1990) Protein Eng. 4, 185-190.

Dyson, H.J. & Wright, P.E. (1991) Ann.Rev.Biophy.Biophy.Chem. 20, 519-538.

Dyson, H.J. , Merutka, G., Waltho, J.P., Lerner, R.A. & Wright, P.E. (1992) J.Mol.Biol. 226, 795-817.

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

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

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

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

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

Kuramitsu, S., Ikeda, K., Hamaguchi, K., Miwa, S. & Nishina, T. (1972) J.Biochem. 72, 1109-1115.

Kuwajima, K. (1991) Proteins: Struct.Funct. & Genetics 6, 87-103.

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

Bruch, M.D., Dhingra, M.M. & Gierasch, L.M. (1991) Proteins: Struct.Funct. & Genetics 10, 130-139.

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

Creighton, T.E. (1990) Biochem.J. 270, 1-15.

Evans, P.A., Topping, K.D., Woolfson, D.N. & Dobson, C.M. (1991) Proteins: Struct.Funct. & Genetics 9, 248-266.

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

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

Galat, A. (1985) Biochem.Biophys.Acta 827, 221-227.

Gans, P.J., Lyu, P.C., Manning, M.C., Woody, R.W. & Kallenbach, N.R. (1991) Biopolymers 31, 1605-1614.

Goodman, E.M. & Kim, P.S. (1991) Biochem. 30, 11615-11620.

Greenfield, N. & Fasman, G.D. (1969) Biochem. 10, 4108-4116.

Handoll, H.H.D. (1985) D.Phil.Thesis, University of Oxford.

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

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

Jackson, M. & Mantsch, H.H. (1992) Biochem.Biophys.Acta 1118, 139-143.

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

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

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

Lehrman, M.S., Mason, S.A. & McIntyre, G.J. (1986) Biochem. 24, 5862-5869.

Lehrman, S.R., Tuls, J.L. & Lund, M. (1990) Biochem. 29, 5590-5596.

Lui, Z-P., Rizo, J. & Gierasch, L.M. (1994) Biochem. 33, 134-142.

Lumb, K.J. & Dobson, C.M. (1992) J.Mol.Biol. 227, 9-14.

Manning, M.C. & Woody, R.W. (1989) Biochem. 28, 8609-8612.

Manning, M.C., Illangasekare, M. & Woody, R.W. (1988) Biophys.Chem. 31, 77-86.

Miranker, A., Radford, S.E., Karplus, M. & Dobson, C.M. (1991) Nature 349, 633-636.

Nakanishi, M., Nakamura, H., Hirakawa, A.Y., Tsuboi, M., Nagamura, T. & Sajio, Y. (1978) J.Am.Chem.Soc. 100, 272-276.

Nelson, J.W. & Kallenbach, N.R. (1989) Biochem. 28, 5256-5261.

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

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

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

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

Robertson, A.B. & Baldwin, R.L. (1991) Biochem. 30, 9907-9914.

Roder, H. (1989) Methods Enzymol. 176, 446-473.

Rohl, C.A., Scholtz, M., York, E.J., Steward, J .M. & Baldwin, R.L. (1992) Biochem. 31, 1263-1269.

Sanford, D.G., Kanagy, C., Sudmeier,J .L., Furie, B.C., Furie, B. & Bachovchin, W.W. (1991) Biochem. 30, 9835-9841.

Segawa, S-S., Fukundo, T., Fujiwara, K. & Noda, Y. (1991) Biopolymers 31, 497-509.

Sönnichsen, F.D., Van Eyk, J.E., Hodges, R.S. & Sykes, B.D. (1992) Biochem. 31, 8790-8798.

Siligardi, G, Campbell, M.M., Gibbons, W.A. & Drake, A.F. (1992) Eur.J.Biochem. 206, 23-29.

Staley, J.P. & Kim, P.S. (1990) Nature 344, 683-688.

Wedin, R.E., Delepierre, M., Dobson, C.M. & Poulsen, F.M. (1991) Biochem. 21, 1098-1103.

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

Wright, P.E., Dyson, H.J. & Lerner, R.A. (1988) Biochem. 27, 7167-7175.

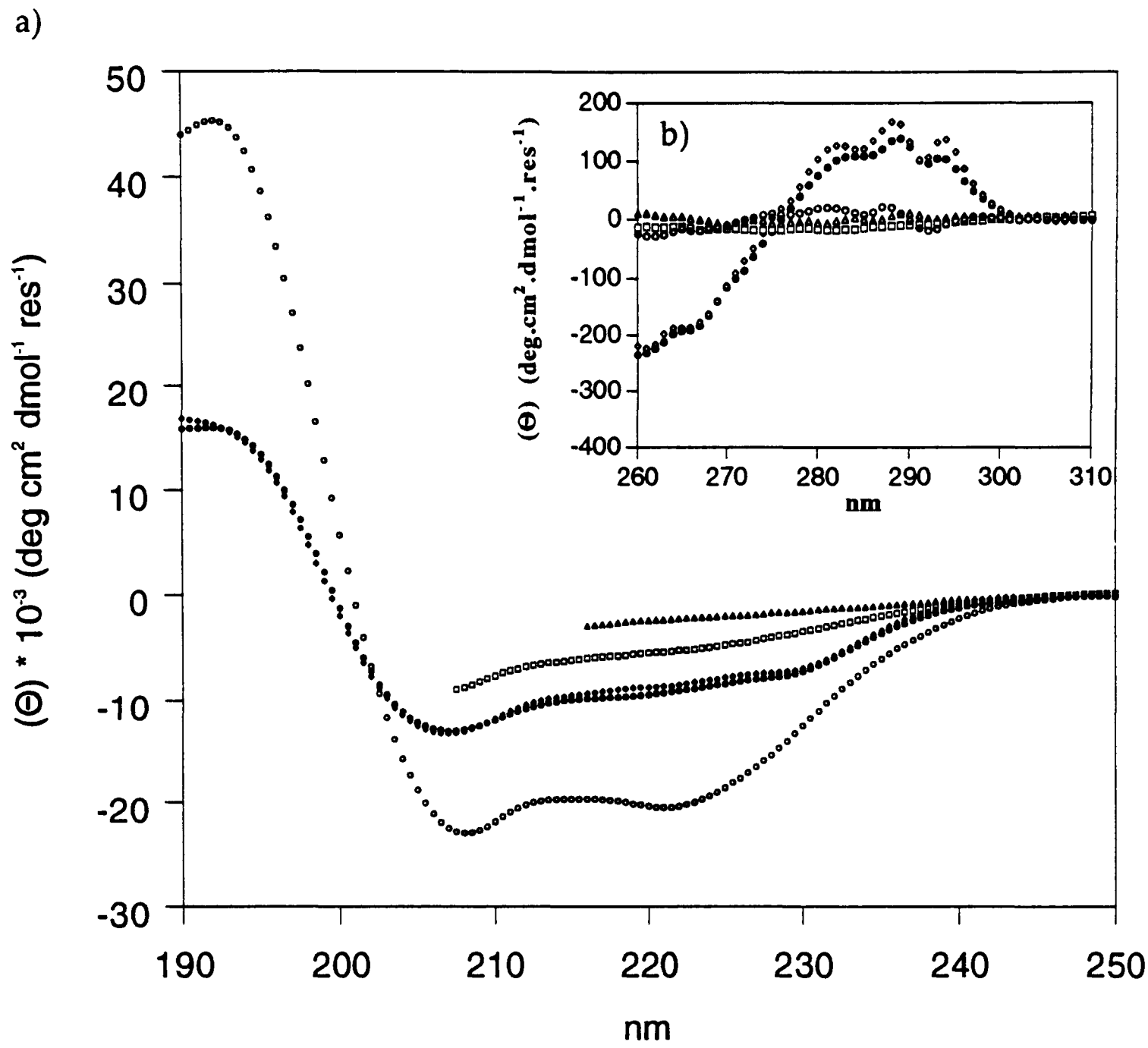


Figure 1

a) Far and b) near uv CD spectra of lysozyme in the native state in aqueous solution (●), the native state in 12.5% TFE (v/v) (◇), the TFE state (50% v/v) (○) and the thermally (65°C) (□) and urea (8M) (△) denatured states.

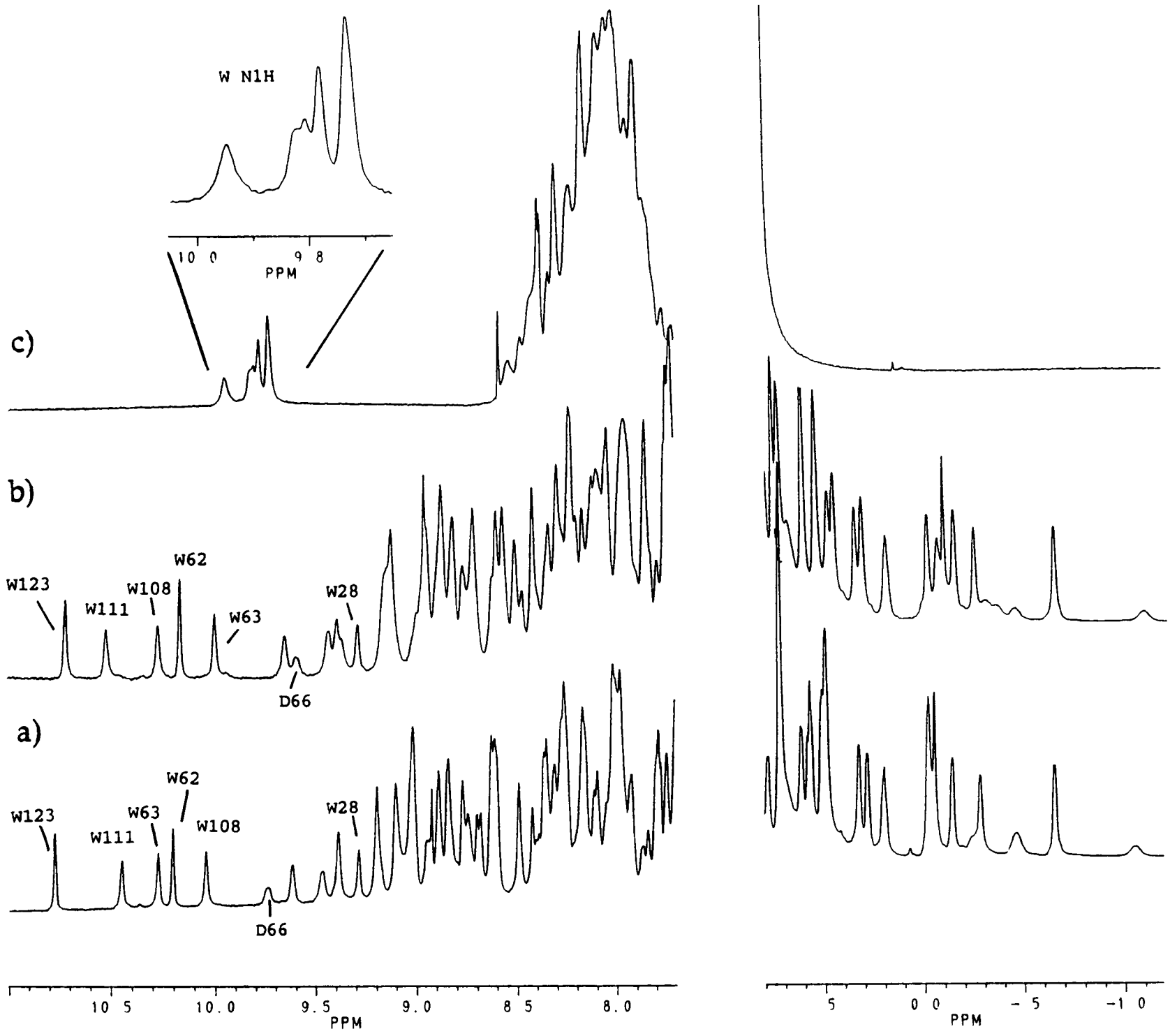


Figure 2

Amide and methyl regions of 1D ^1H NMR spectra of hen lysozyme at pH 2.0, 27°C. a) native state, 90% H_2O/10% D_2O. b) TFE bound state, 12.5% TFE-d_3/H_2O (v/v). c) 50% TFE-d_3/H_2O (v/v). The small peak at 0 p.p.m. in (c) arises from an impurity. The tryptophan indole NH resonances are shown for the TFE state as an expanded inset.

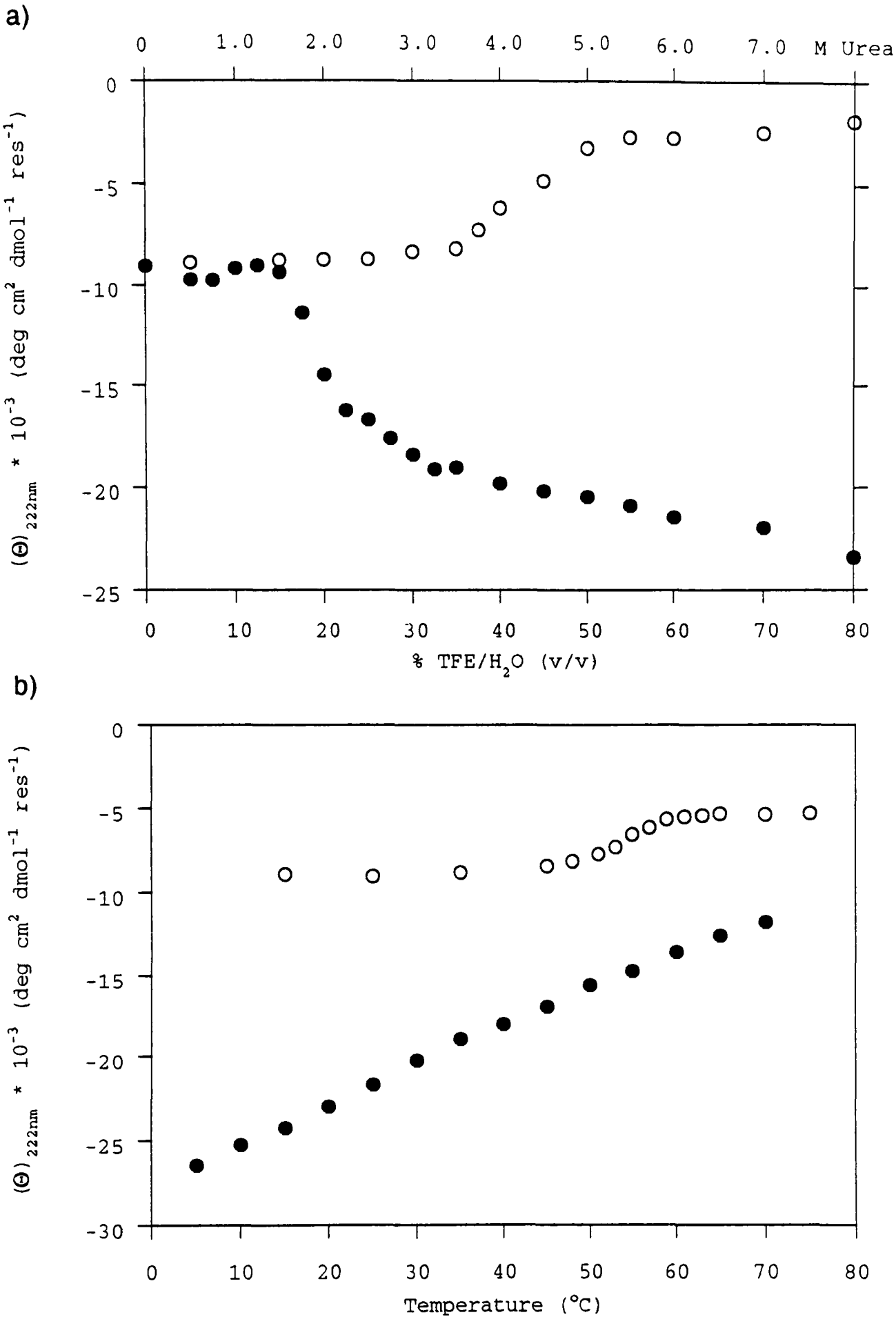


Figure 3

Molar residue ellipticity at 222nm monitoring denaturation of hen lysozyme induced a) by urea (O) and TFE (●) at pH 2.0, 27°C and b) by temperature in the presence (●) and absence (O) of 50% TFE (v/v), pH 2.0.

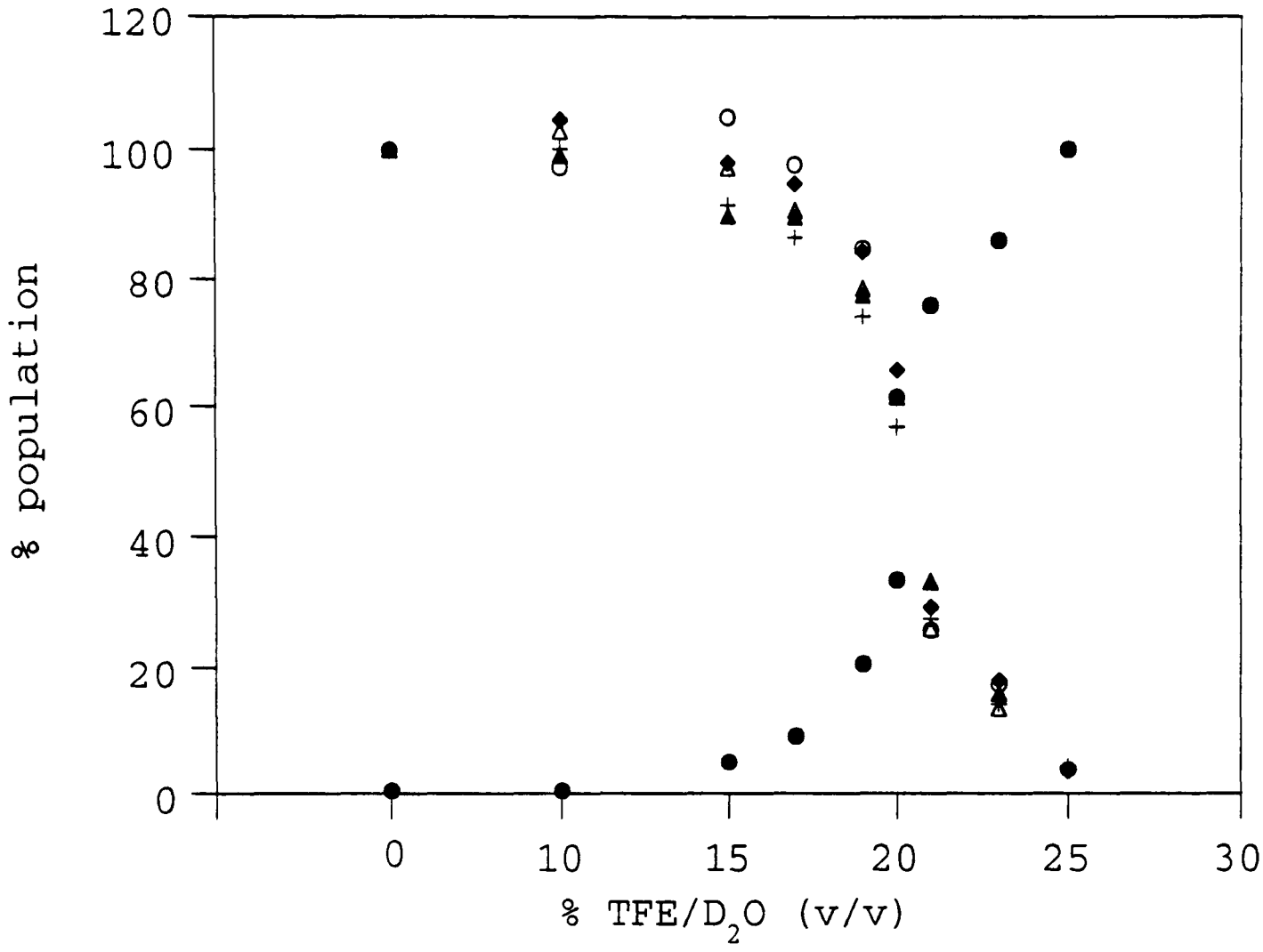


Figure 4

Intensity of selected resonances in the 1D ^1H NMR spectrum of lysozyme as a function of the concentration of TFE: (Δ) γ-CH_3 of Leu17 (A-helix); (◆) C5H of Trp28 (B-helix); (+) C5H of Trp108 (D-helix); (▲), εCH_3 of Met105 (hydrophobic box). The location of these residues in the native structure is given in parenthesis. There is no significant change in linewidth of these resonances over this range of TFE concentration. Data are also given for the C2H resonance of His15 in the native (O) and denatured (●) states as population of the respective state. These values are expressed relative to the peak intensity of the same resonance in the native protein in 100% D_2O or in case of His15 in the denatured state to the intensity of this resonance when lysozyme is fully TFE denatured (30% TFE v/v). Errors in estimation of peak intensities are ±5%.

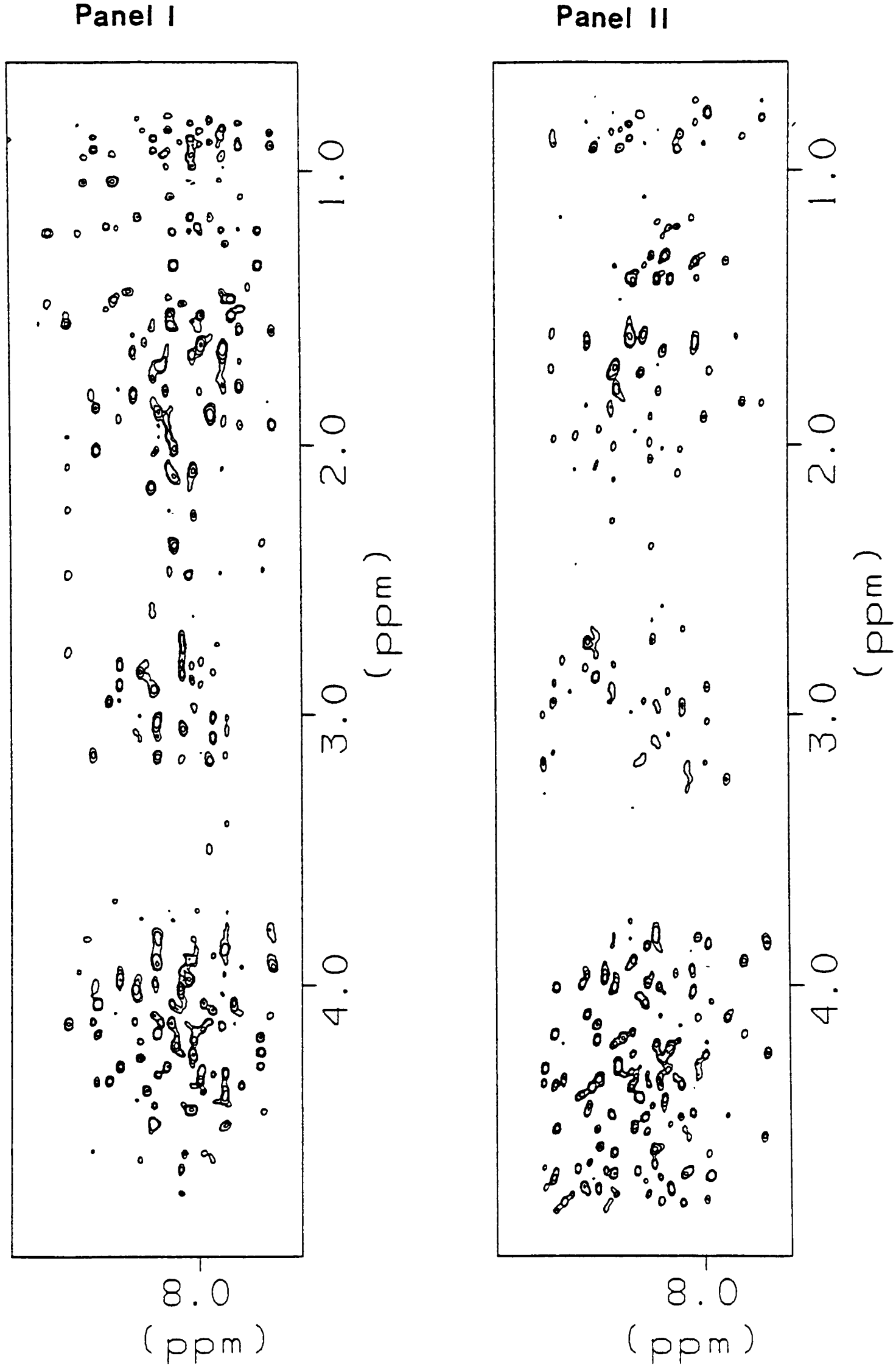


Figure 5

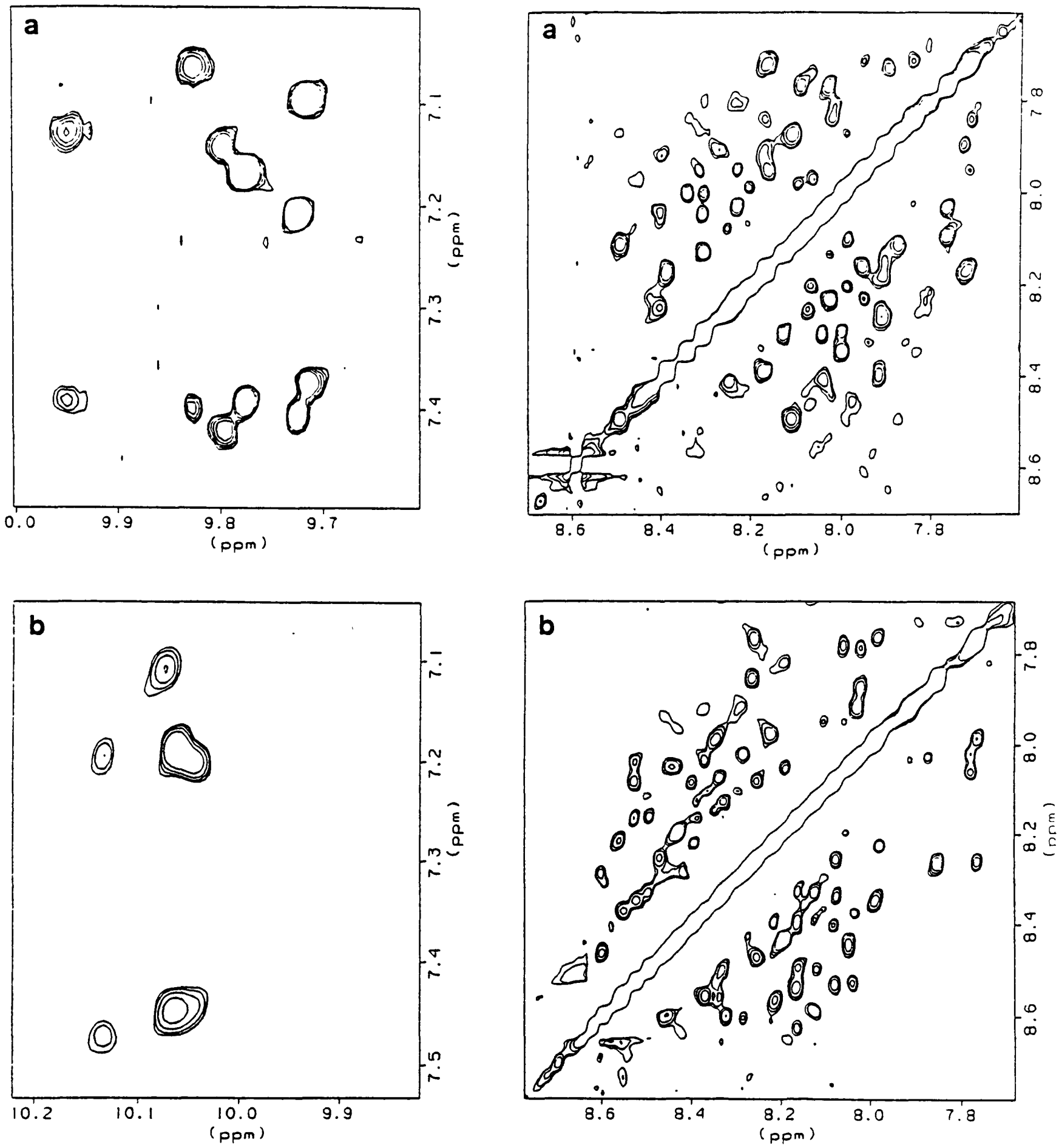
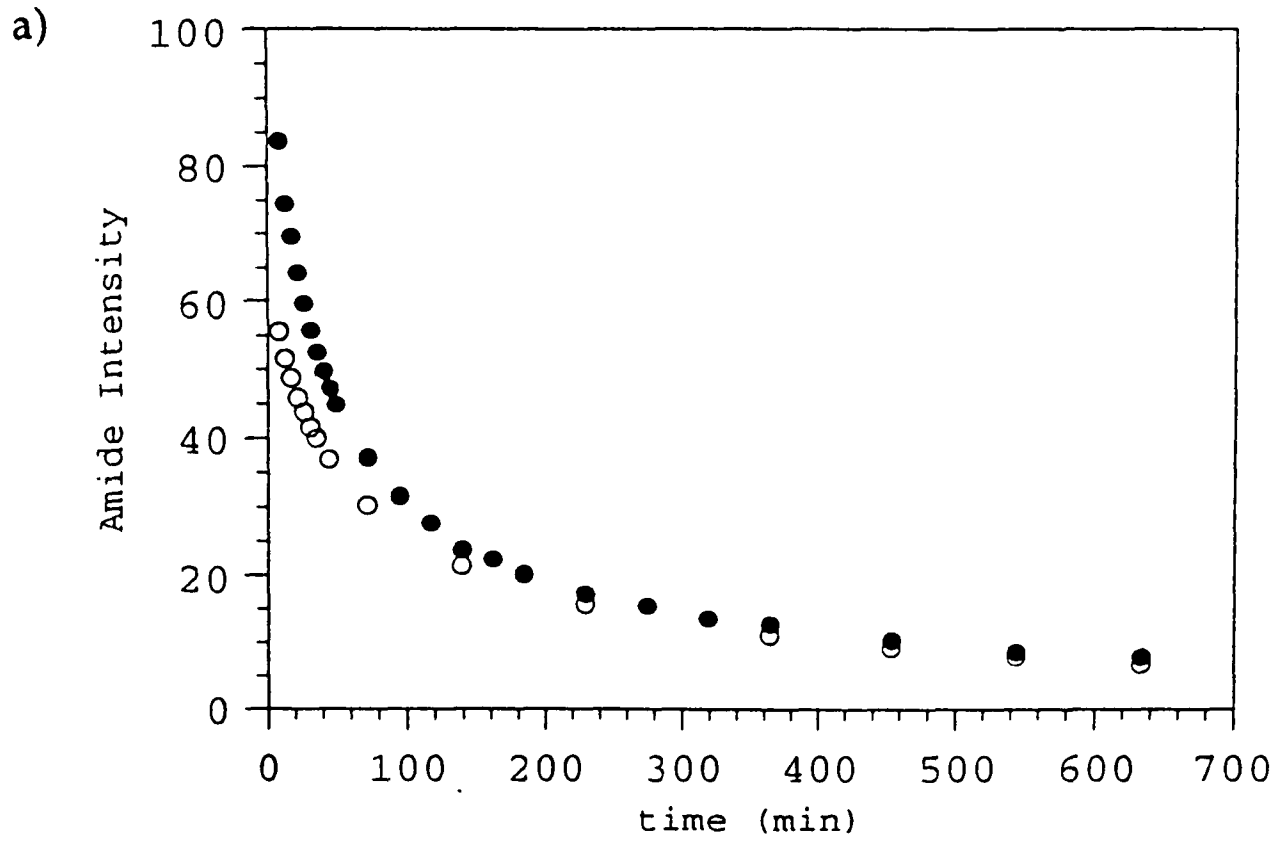


Figure 5

NH-αH and NH-sidechain region of the 2D NOESY spectrum (mixing time 200ms) of hen lysozyme in 50% TFEd_2-OH/H_2O (v/v) (Panel I) and 8M urea (Panel II) at 27°C, pH 2.0. The tryptophan indole to Trp C2H/C7H and NH-NH regions of the spectra in 50% (v/v) TFE (a) and 8M urea (b) are shown in Panels III and IV, respectively.



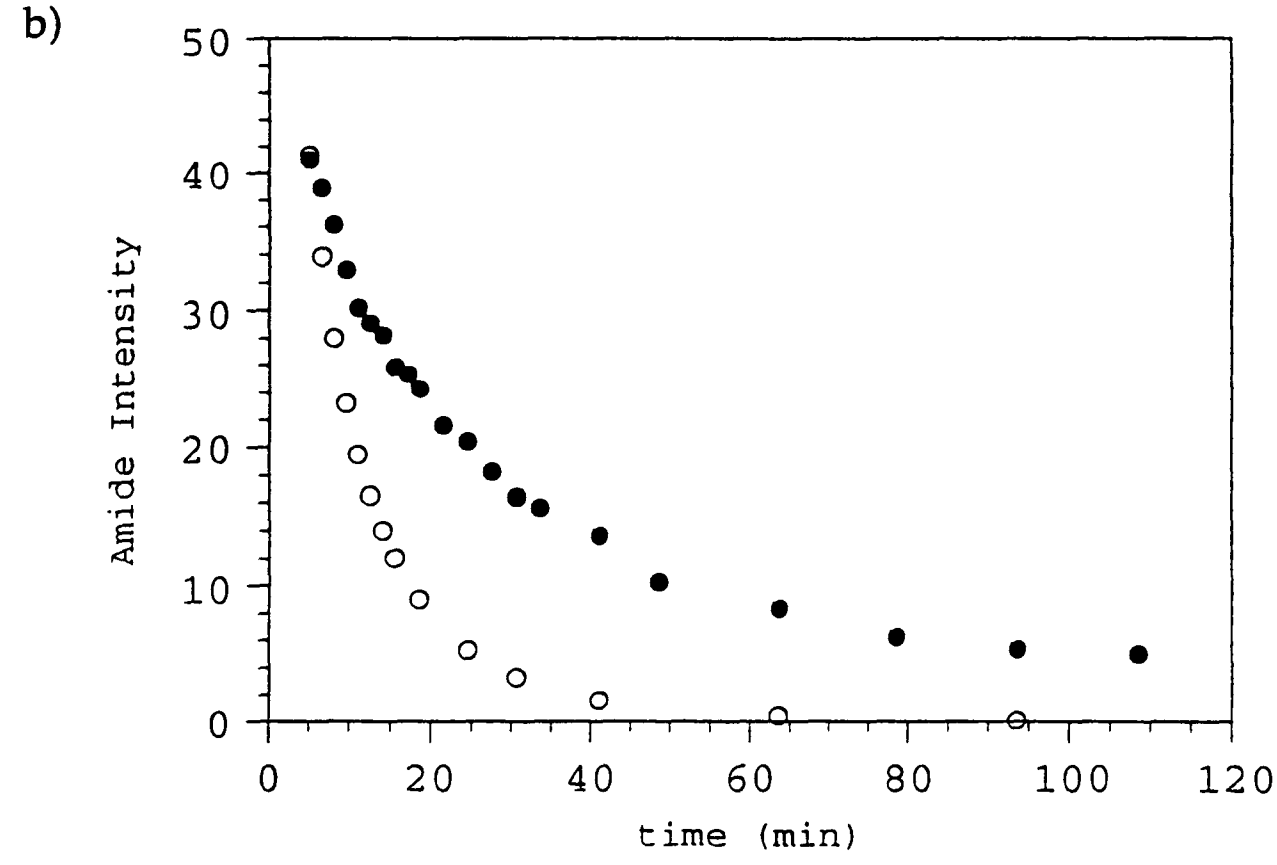


Figure 6

a) (●) Bulk hydrogen exchange from the TFE state and (O) the bulk exchange timecourse of the 66 amides measured individually. b) (●) Selective bulk exchange of rapidly exchanging amide hydrogens in the TFE state and (O) exchange timecourse computed for the 127 amide hydrogens of hen lysozyme exchanging at their intrinsic rates. The exchange of amide hydrogens into 50% TFE/D_2O (v/v), pH 2.0, 27°C was measured in each experiment. Errors in the integration of peak intensities are $<\pm 10\%$.

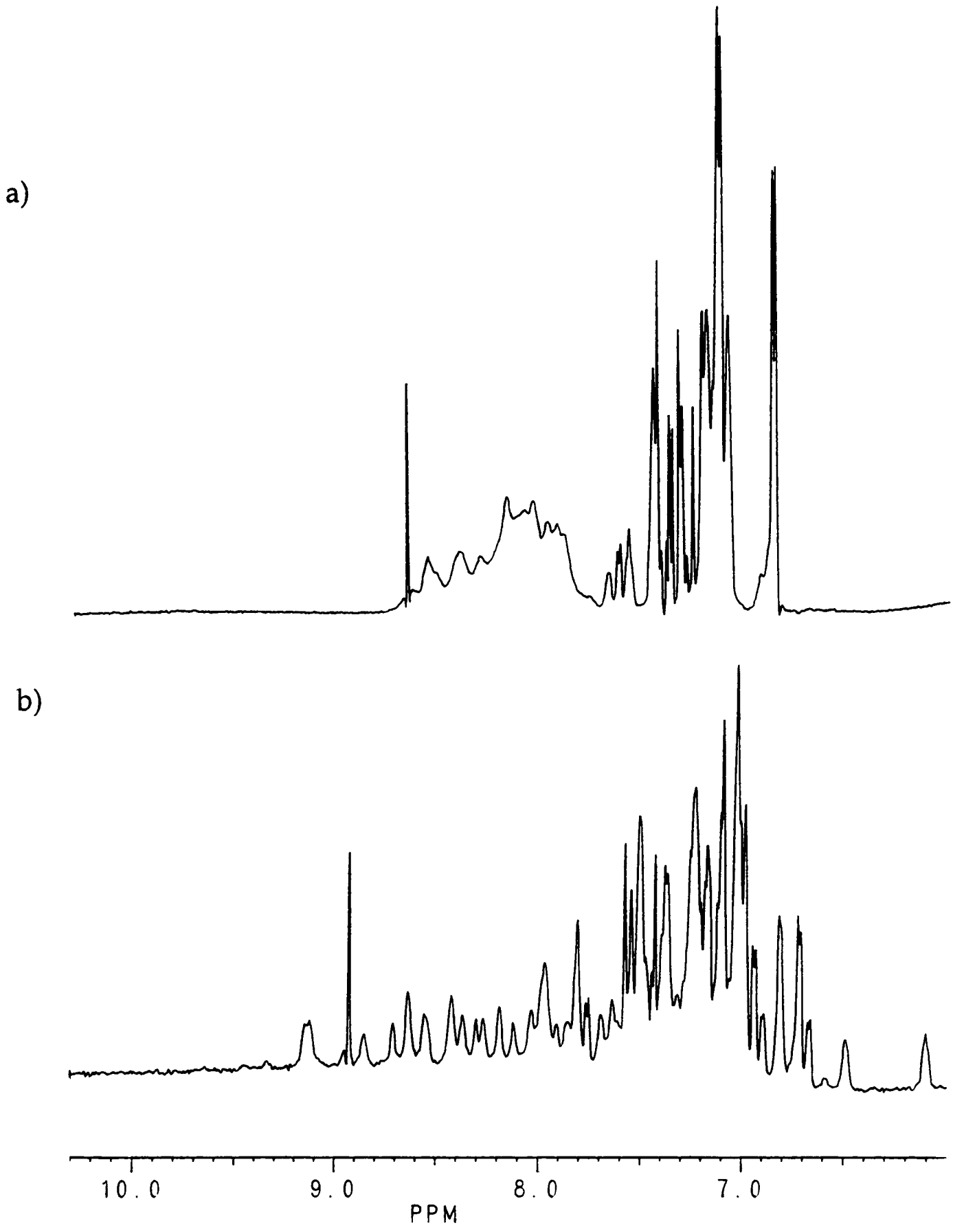


Figure 7

a) Amides remaining in the 1D ^1H NMR spectrum after 2 hours in the presence of 50% TFE-d_3/D_2O (v/v), pH 2.0. b) Amide intensity transferred to the native state spectrum after refolding of sample (a).

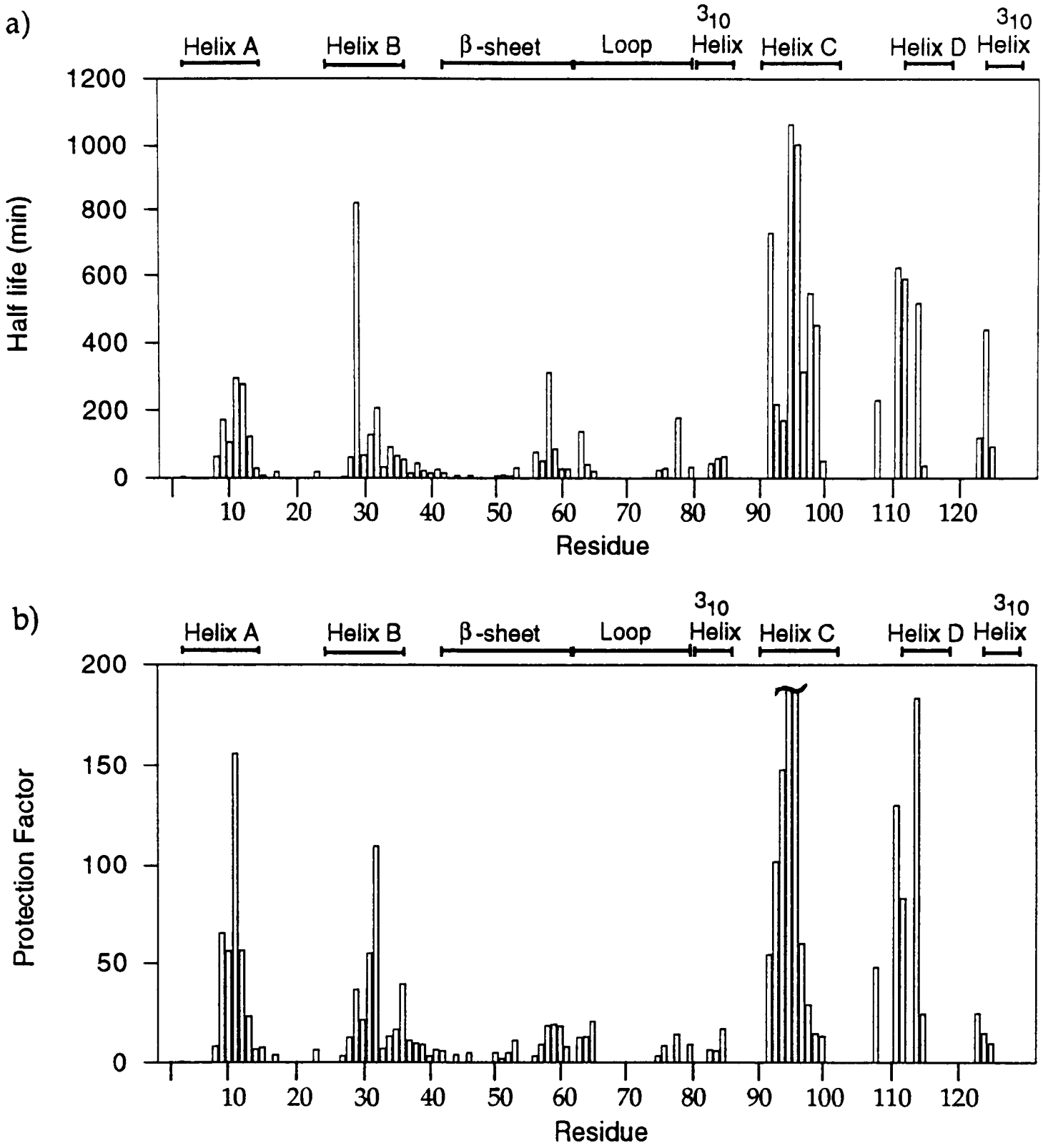


Figure 8

a) Half lives of amides protected in the TFE state of hen lysozyme versus the protein sequence. b) Protection factors (uncorrected for TFE effects) of amides protected in the TFE state. Protection factors for Ala95 and Lys96 exceed the scale with values 450 and 215. Regions of secondary structure in the native state are shown.

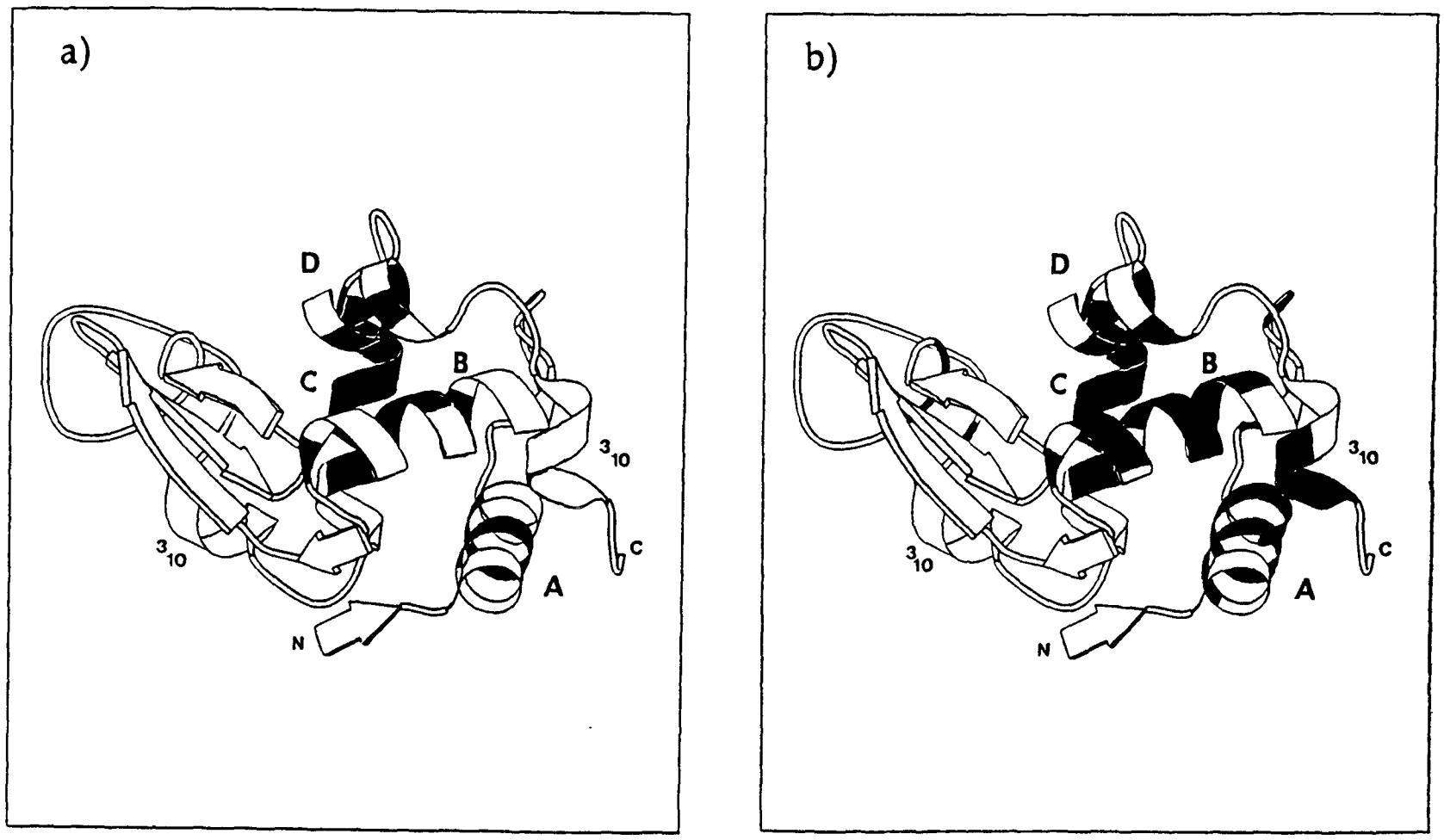


Figure 9

Schematic representation of the conformation of native lysozyme. Regions of the polypeptide chain with amide protection factors exceeding 35 in the TFE state are shaded in a). Regions which are protected in the major kinetic refolding intermediate with a time-constant of 60ms (Radford et al., 1992b) are shaded in b). This figure was drawn using the program MolScript (Kraulis, 1991).

Chapter 5

Contribution of Disulphide Bonds to Structure and Stability of a Partially Folded State of Hen Egg-White Lysozyme in Trifluoroethanol.

Summary

A partly folded state of hen lysozyme is induced by addition of 50% (v/v) 2,2,2-trifluoroethanol (TFE), at pH 2.0, 27°C and was described in the preceding chapter 4. This state is characterised by extensive secondary structure, of which only the native-like helices are stably structured, but contains only few, if any persistent tertiary interactions. However, the native disulphide bridges persist in this denatured state and in order to probe their role in the structure and stability of the TFE denatured state, these studies were extended to include a fully disulphide reduced and cysteine methylated derivative of the protein ($HEWL^{+SMe}$). By exploring a wider range of solution conditions, significant near uv circular Dichroism (CD) is detected for the unmodified, disulphide intact protein in TFE at high cosolvent concentration and especially at low temperature. This signal and a significant proportion of the chemical shift perturbations seen in ^1H NMR spectra of HEWL are absent in the TFE denatured state of $HEWL^{+SMe}$ and are therefore associated with the presence of the native disulphide bonds. However, far uv CD spectroscopy and amide hydrogen exchange suggest that the disulphide crosslinks have little, if any, effect on the overall extent and persistence of secondary structure.

Introduction

The characterisation of the structure and the dynamics of the polypeptide chain in conformations intermediate between its native and "random coil" state is central to the study of protein folding (Kim & Baldwin, 1990). Site specific examination of denatured and of partly folded states of proteins by techniques such as nuclear magnetic resonance spectroscopy, is likely to provide an understanding of the interactions which allow native-like structures to compete with alternative structures during protein folding (Dill & Shortle, 1991; Dyson & Wright, 1993). One such interaction found in many conformational states of proteins is that of disulphide bridges, which cross-link the polypeptide chain and therefore restrict its freedom (Thornton, 1981; Creighton, 1993). The role of disulphides in native states of proteins is largely thermodynamic rather than structural in origin. Extensive studies, in particular by protein engineering, suggest that the energetic consequences of disulphide bond formation are not straightforward and often dependent on their context in the native state (Betz, 1993). Disulphides have, however, considerable effects on both the structure and the dynamics of partly folded as well as substantially denatured states. This suggests, they may also have important contributions to the energetics and kinetics of protein folding (Creighton, 1992; Uversky, 1993; Clarke & Fersht, 1993).

Hen lysozyme (Figure 1) contains four disulphide cross-links in the native state (6-127, 30-115, 64-80, 76-94). They have been central to a wide range of studies, including those investigating oxidative refolding from the reduced protein (Goldberg et al., 1991; Sawano et al., 1992). *In vivo*, disulphides are not formed prior to folding but once lysozyme has been translocated into the lumen of the endoplasmic reticulum (ER). Their formation is thought to occur in concert with the folding process which is helped in protein dense media by auxiliary proteins, such as protein disulphide isomerase (PDI) (Freedman, 1992). The importance of disulphide bonds is also apparent when lysozyme is refolded in presence of the native disulphides from guanidine hydrochloride or other denaturants. A transient overshoot of ellipticity in far uv Circular Dichroism (CD) spectra (Radford et al., 1992) has been associated with disulphide strained conformations in some of the early folding intermediates (Chaffotte et al., 1992). Furthermore, studies on a three disulphide derivative of lysozyme ($CM^{6\text{-}127}$) have shown that the transient accumulation of stable intermediates during kinetic refolding may at least in part arise from the presence of disulphides (Eyles et al., 1994). The role of disulphides in partly

folded states has been studied in the homologous protein α-lactalbumin (Ewbank & Creighton, 1993; Peng & Kim, 1994). Although, hen lysozyme does not form stable molten globule states under similar equilibrium conditions (Radford et al., 1991; Cooper et al., 1992), denatured states of hen lysozyme are known to retain considerable residual structure, at least a some of which has been associated with the persistence of the native disulphides (Evans et al., 1991; Uversky, 1993).

It was shown in Chapter 4 that TFE does not induce structure indiscriminately in hen lysozyme but that α-helical structure is persistently stabilised in this state of the protein only in regions which are helical in the native state, while additional non-native structure may be induced in other parts of the sequence but is transient by nature (see Chapter 7). Several of the helices of hen lysozyme are linked by disulphide bonds (see Figure 1) and the possibility exists, therefore, that the persistence of these native interactions may be required for the generation of ordered structure in the protein. In this study, a derivative of hen lysozyme in which all the disulphides have been reduced and the cysteines blocked by methylation ($HEWL^{+SMe}$) is used to investigate this and related issues.

Materials and Methods

Hen egg white lysozyme was obtained from Sigma Chemical Company and was extensively dialysed at pH 2.0, lyophilised and stored at -20°C. The following chemicals were used: TFE-d_3 and TFE-h_3 (purchased from Fluorochem, Ltd.), 2-mercaptoethanol (BDH), ultra pure urea (BRL), p-nitrobenzesulfonic acid methylester (Aldrich).

Fully methylated lysozyme ($HEWL^{+SMe}$) was prepared by a procedure involving the denaturation of the protein in 8 M urea, the reduction of disulphides with mercaptoethanol and the reaction of free thiols with nitrobenzesulfonic acid methylester as described previously (Lee & Atassi, 1973). The reaction product was extensively washed by dialysis at pH 2.0 and its composition and purity was checked by electrospray ionisation mass spectroscopy and amino acid analysis. Further purification was not necessary as a single species was obtained, however with a mass of 14433.6 ± 2.1 Dalton corresponding to the addition of 9 methyl groups. It is possible that this additional reaction involves the single histidine of hen lysozyme, as the

reagent has been known to modify this amino acid sidechain also (Heinrikson, 1970). The protein was then lyophilised.

Circular Dichroism (CD) measurements were made on a Jasco J-720 spectropolarimeter. The concentration of lysozyme samples was typically 2-4 μM for far and 20-40 μM for near uv CD measurements in a 1 cm cell. The pH of each sample was adjusted to 2.0 after addition of TFE. The far and near uv CD spectra were accumulated as the average of 4 and 8 scans respectively at a scan speed of 20 nm/min. Molar ellipticities (Θ) ($deg.cm^2.dmol^{-1}res^{-1}$) are reported. Protein concentration was determined by amino acid analysis and by uv absorption at 280 nm. ^1H NMR spectra including those monitoring bulk hydrogen exchange were acquired as 8k complex data points over a sweep width of 7246.38 Hz at 600 MHz on a Bruker AM spectrometer belonging to the Oxford Centre for Molecular Sciences. The lyophilised enzyme was dissolved in 50% TFE-d_3/D_2O, to a final protein concentration of 2.5 mM and the pH was rapidly adjusted to 2.0 with small volumes of DCl or NaOD. One dimensional NMR spectra were acquired immediately at 27°C at intervals of 4.5 min over a period of 12 hour. Amide hydrogen intensities were integrated in the 1D spectra between 8.80 and 7.75 p.p.m. and were scaled relative to the area of the 59 aromatic protons, which resonance between 7.75 and 6.60 p.p.m. in the spectrum acquired for the same sample after all of the amide hydrogens had been exchanged with deuterons by heating the sample to 58°C for 15 min.

Results

Structure of HEWL and HEWL^{+SMe} in aqueous solution

Figures 2a and 2b show respectively the far and near uv regions of CD spectra obtained for unmodified hen egg white lysozyme (HEWL) and the fully disulphide reduced and cysteine methylated species (HEWL^{+SMe}) in aqueous solution at pH 2.0 and 27°C. HEWL is in the native state under these condition and the spectra are characterised by ellipticity arising from a mixture of α- and β-secondary structure in the far uv CD. The near uv CD is comprised of signals arising from aromatic chromophores, for example in the region 285-300 nm principally from tryptophan residues, which are constrained by the native fold in fixed orientations (Goux & Hooker, 1980). Tight packing in the protein fold is also responsible for the chemical shift dispersion seen for the resonances of aromatic residues shown in the ^1H NMR spectrum in Figure 5 (Redfield & Dobson, 1990) and the 10^5 to 10^8 fold retardation of exchange rates with solvent observed for some amides in the native state (Chapter 2).

The characteristics of hen lysozyme which has been fully reduced and cysteine methylated ($HEWL^{+SMe}$) are those of a highly denatured state. The ellipticity in the far uv CD is considerably reduced in magnitude relative to that of the disulphide intact protein and is similar to that obtained for the thermally denatured protein (data not shown), suggesting that $HEWL^{+SMe}$ contains little regular secondary structure (see also Chapters 3 and 4). Furthermore, near uv CD is absent, implying that there are no fixed tertiary interactions involving aromatic chromophores. The rate of bulk amide exchange (Figure 7) approximates to that expected for the polypeptide in an unstructured conformation within a factor of two. The NMR spectrum of the protein derivative, Figure 5, is similar to that of the reduced protein (data not shown), which suggests that also in this state significant residual clustering of aromatic sidechains is likely to exist (Evans et al., 1991; Broadhurst et al., 1991).

Titration of HEWL and $HEWL^{+SMe}$ with TFE cosolvent

Figure 3a and 3b show the structural transitions in HEWL and $HEWL^{+SMe}$ as a function of TFE concentration at pH 2.0, 27°C by the changes seen in the far and near uv CD spectra. In the case of HEWL only small perturbations are observed in the far and near uv CD spectra up to a concentration near 15% TFE (v/v) suggesting that the native fold is stable up to this concentration of cosolvent (Chapter 4). At pH 5.5 a higher concentration of TFE is required to denature HEWL corresponding to the increased stability of the native state at this pH (Lu, Buck, Radford and Dobson, unpublished data). The negative ellipticity, monitored at 222 nm in the far uv CD, increases steeply over a narrow range of TFE concentration between 15 and approximately 25% (v/v) for HEWL. At 80% TFE (v/v) the far uv CD ellipticity exceeds that of the native by a factor of almost 2.5. Previously it was suggested that this conformational transition could be a cooperative one, because changes in far uv CD and ^1H NMR were found to occur concomitantly over the same range of TFE concentration with mid-point near 20% v/v. (Chapter 4). The near uv CD data shown in Figure 3b support this analysis within error of the measurement.

The transition for $HEWL^{+SMe}$ occurs over a wider range of TFE concentration than observed for HEWL, possibly because the transition does not involve the disruption of cooperatively stabilised tertiary structure. Similar results have been obtained for polypeptides which are largely unstructured in water, such as short peptides of hen lysozyme (Bolin, Yang, Radford & Dobson unpublished data). What is most remarkable, however, is that the titration of lysozyme with and without disulphide bonds is very similar. Except at low

concentrations of TFE (< 15% v/v) the titrations of the two proteins are nearly indistinguishable. Considerable post-transitional increase in the negative ellipticity at 222 nm occurs with further addition of TFE to hen lysozyme both with and without disulphides. Therefore, the transition with TFE cannot be considered as two-state over the entire concentration range. Further evidence for the non two-state behaviour of the denaturation of HEWL is seen in the near uv CD. At 25% TFE (v/v) near uv CD spectra of hen lysozyme are nearly identical with and without disulphide bonds. However, considerable ellipticity is recovered in the near uv CD for the TFE denatured state of the disulphide intact protein when the concentration of TFE is further increased at 27°C, pH 2.0. For HEWL in 50% TFE, this ellipticity has reached 25% in magnitude of the signal obtained for the native protein. However, no significant change is observed in the near uv CD in the case of $HEWL^{+SMe}$ over the entire range of TFE concentration (Figure 3b).

The effect of disulphide bridges on CD and ^1H NMR spectra of the TFE state

Although the native disulphide cross-links have no appreciable effect on the far uv CD spectra of the TFE state, they appear to be associated with the small near uv CD signal observed for the disulphide intact protein in presence of TFE. Evidence that the residual signal in the TFE denatured state of HEWL is significant is supported by the following observations. Firstly the signals become more considerable at higher TFE concentrations and especially at low temperatures, approaching in intensity those of the native state (see Figure 4 and below). Secondly, it is not possible to recover any considerable near uv CD signal for $HEWL^{+SMe}$. The near uv CD spectra of this derivative of the protein are nearly identical in presence and absence of TFE, even at high TFE concentration or at low temperature. Disulphides in certain conformations, can significantly contribute to near and far uv spectra (Goux & Hooker, 1980; Woody, 1985; Chaffotte, et al., 1992). Taken together with the data here, this implies that the near uv signal in the TFE denatured state of HEWL is associated either directly with the persistence of the native disulphides or through tryptophan residues which are constrained by the disulphides in presence of TFE.

Figure 5 shows the aromatic region of 1D ^1H NMR spectra of HEWL and $HEWL^{+SMe}$ recorded at 27°C in aqueous solution and in the presence of 50% (v/v) TFE/D_2O, pH 2.0. Besides that of the native HEWL, the spectra are typical of proteins in denatured conformations. However, subtle differences are observed between the three spectra. For example, resonances at 6.83 p.p.m. arising from the three tyrosines of the protein, overlap almost completely in

the TFE denatured states, but are resolved into two groups for HEWL^{+SMe} in the aqueous solution. By contrast, considerable chemical shift dispersion is evident for tryptophan ring protons around 7.60 p.p.m., particularly in the case of the unmodified protein in TFE. The six tryptophan indole resonances are resolved in spectra recorded of HEWL in 50% TFE/H_2O, as are the resonances of the three phenylalanines in 2D heteronuclear spectra (Chapter 4, & data not shown). Such residual chemical shift dispersion arises in denatured states in instances when protons are brought into, at least transient proximity of other aromatic residues (Evans et al., 1991). It follows that at least some of the aromatic sidechains are partly clustered, locating them in different environments in the TFE denatured states. This residual clustering appears to be reduced in the absence of disulphides, as judged by the diminished shift perturbation of the aromatic resonances.

A second feature of the NMR spectra of the denatured states is the observation of differential linewidths for resonances. It is notable, for example that, with exception of the tyrosine resonances, many of the other resonances of the aromatic protons are considerably broadened in the spectrum of HEWL in TFE. By contrast to this, the majority of resonances in the spectrum of HEWL^{+SMe} in TFE have narrow linewidths and therefore, resonance broadening must, at least in part, be associated with the presence of disulphides (see Chapters 7 and 8). Resonance linewidths, the chemical shift perturbations and the near uv CD signal for both proteins in TFE were found to be unchanged over at least a ten fold range of protein concentration (0.5 to 5 mM by NMR, 0.03-0.3 mM by near uv CD), suggesting that these effects do not originate from intermolecular associations. They are, however, sensitive to TFE concentration and temperature. Thus, the NMR data reflect considerable conformational averaging in the TFE state of HEWL, approaching the intermediate exchange regime on the NMR time-scale. The differences between the NMR spectra of HEWL and HEWL^{+SMe} in TFE imply that the presence of disulphides has significant effects on the conformation and dynamics of amino acid sidechains. Despite this the overall structural content of HEWL and HEWL^{+SMe} in TFE appears to be nearly identical as monitored by far uv CD and it is interesting to assess whether disulphides and the residual tertiary structure affects the stability of this structure in the TFE denatured state of hen lysozyme.

Effect of disulphides on the stability of structure in the TFE denatured states. Thermally or chemically induced denaturation transitions are commonly used to assess the thermodynamic stability of proteins (e.g. Taddei et al., 1994). This analysis is not possible when proteins exist in an ensemble of states which

interconvert by structural changes which are non-cooperative. The changes seen in the near and far uv CD for HEWL in presence of 50% TFE (Figure 4a,b and 6a,b) are almost linear with temperature and urea concentration, suggesting that melting of residual tertiary structure and secondary structure are each non-cooperative; It is intriguing to note that the sensitivity of structure to denaturation, taken as the gradient $\Delta CD_{222nm}/\Delta T$ and $\Delta CD_{222nm}/\Delta urea$ of the denaturation profiles, are within experimental error identical for the TFE denatured states with and without the disulphides.

Amide protons which are hydrogen bonded or buried by protein structure are generally slowed from exchange with solvent deuterons. Thus, protection of amides against exchange with solvent is a sensitive test for the presence of persistent secondary structure in proteins (Chapter 3). $HEWL^{+SMe}$ can not be refolded to the native state and therefore it is not possible to use this method to identify protected sites site-specifically by the methods described in Chapters 2, 3 and 4. Figure 7 shows the bulk exchange profiles of HEWL and $HEWL^{+SMe}$ in presence of 50% TFE at pH 2.0, 27°C. The bulk hydrogen exchange time course shows the exchange processes of all the labile protons as a sum and is clearly multi-exponential, as depicted by its non-linearity when plotted as ln(amides remaining) versus time. It is clear that a number of amide hydrogens exchange at rates substantially slower than are anticipated under these conditions on the basis of the rates predicted (Bai et al., 1993) for an unstructured polypeptide with the sequence of hen lysozyme. These rates were summed to reconstruct a profile of predicted bulk exchange in absence of protection and this is shown for comparison in Figure 7. The rate of exchange after 100 min can be fitted to a linear relationship, the rate of which is found to be around 24 and 19 times slower for HEWL and $HEWL^{+SMe}$, respectively, than the predicted intrinsic bulk exchange rates. Studies on model peptides showed that the effect of TFE on intrinsic exchange rates is small at pH 2.0 (Chapter 4; Bolin, 1994) and is unlikely to account for this retardation of exchange rates. The differences in the rates of exchange of the slowly exchanging amides in the TFE denatured states of HEWL and $HEWL^{+SMe}$ is small (around 25%). Considering the errors which are anticipated to arise, for example from slight discrepancies of sample pH and ionic strength, in this type of experiment, not only the number of around 30 amides involved in the persistent structure, but also their extent of protection appear to be similar in the TFE states of hen lysozyme in presence and absence of disulphides. The origin of their protection, therefore, is likely to be similar and is not affected by disulphides.

Discussion

Secondary Structure

The transition monitored by far uv CD, which is induced by addition of TFE in a derivative of hen lysozyme lacking disulphide bonds (HEWL^{+SMe}) is highly similar to that observed for the unmodified, disulphide intact protein (HEWL). This suggests that very similar structures are stabilised in the reduced protein, even though the initial state of the protein is highly denatured in absence of disulphides. A remarkable observation of the present study is that the TFE denatured states of lysozyme in presence and absence of the four native disulphides are identical as judged by far uv CD. This surprising conclusion has consequences for the interpretation of the far uv data itself: The far uv CD signal at 222 nm in 50% TFE is increased by more than 2.3 fold over that of the native state of HEWL, and 4.1 times over that seen for HEWL^{+SMe} in absence of TFE. Previous interpretations of this result, elaborated in Chapter 4, suggested that the α-helical content of the protein could be substantially increased, either by elongation of the native helices or by induction of additional novel helical structure. In addition, it is also possible that helices are essentially native-like but are of greater regularity in TFE due to the absence of at least the majority of tertiary contacts. A number of alternative interpretations are also plausible, however. These include the possibility of contributions to the far uv signal from aromatic residues or strained disulphide bonds, which could contribute either positive ellipticity to the native state or negative ellipticity to the TFE state (Woody, 1985; Chaffotte, 1992). Although the possibility remains that the far uv CD signal which arises purely from secondary structure in the TFE states of HEWL and HEWL^{+SMe} could be substantially different, and that the spectra are made similar by chance due to such compensatory effects, this seems unlikely considering that their equivalence is observed over a wide range of experimental conditions.

The simplest interpretation of the far uv CD data is that the extent of α-helical structure is represented rather accurately, with the implication that the overall α-helix content in the TFE state of hen lysozyme is increased over and above that found in the native state (see Chapter 7). Secondly, the results described here show that disulphides play no role in this process. The latter is supported by the recent observation that a far uv CD spectrum nearly identical to that of intact lysozyme in TFE may be constructed as the sum of spectra obtained from four peptides, spanning the sequence of lysozyme in segments of 22-46 residues in length, when placed under identical solution conditions (Yang, unpublished data). In this respect, the TFE denatured state

may resemble rather closely a collection of peptides of hen lysozyme in presence of cosolvent. Evidence that this interpretation may be correct at the level of secondary structure, comes from the observation that the melting behaviour of HEWL and HEWL^{+SMe} with temperature and urea is non-cooperative and very similar (Figure 6). Furthermore, the hydrogen exchange behaviour of the TFE state is only marginally affected by removal of the four disulphides (Figure 7). Similarly, studies of a 17 residue peptide corresponding to the C-helix of hen lysozyme in TFE suggest that protection factors are only marginally reduced in magnitude compared with those of the corresponding residues in the TFE state of intact lysozyme (Bolin, 1994). Therefore, the data reported in this chapter show that the extent and persistence of secondary structure in the TFE state are closely similar, irrespective whether the native-like disulphides persist in these partly folded states.

Tertiary Interactions

Despite the similarity of secondary structure in TFE denatured states of hen lysozyme, there is also evidence that certain features of the structure and dynamics of the protein are different in presence and absence of disulphides. The near uv CD spectrum of HEWL, but not of HEWL^{+SMe}, is characterised by considerable signal intensity, particularly at low temperature or high TFE concentration. These signals could arise directly from the disulphides or from constrained aromatics as a consequence of the presence of the native disulphide bonds. The site specific origin of the CD signals cannot be defined because several aromatic and cysteine residues are present in the protein. Similarly, chemical shift perturbation in NMR spectra indicate that some aromatic residues are clustered in the TFE state in presence of disulphides. It has not yet been possible to identify by heteronuclear ^{15}N-edited NMR the residues involved in this tertiary structure (Chapter 7). It is clear, however, from the differences in resonance linewidths, seen in the spectrum of HEWL and HEWL^{+SMe} in TFE, that disulphides have considerable effect on the overall conformation and the dynamics of these denatured states. The variation of the near uv CD spectra, ^{1}H chemical shift dispersion as well as linewidths with TFE concentration and especially temperature (data not shown) suggests that as helices become dynamically more rigid, disulphides and turns are likely to act as hinges for the structure, bringing sidechains into close proximity. This is not possible when disulphides are absent. Indeed in this case the NMR data, shown in Figure 5, and low angle X-ray scattering data (Goto, personal communication) suggest that the TFE state could be considerably more expanded in the absence of disulphides. Preliminary studies with a three disulphide derivative of hen lysozyme (CM$^{6-127}$ lysozyme

characterised in Radford et al., 1991) suggest that possibly all of the four disulphides are required to give rise to a significant near uv CD signal in the TFE denatured state (data not shown).

The remarkable result of this study is that the tertiary interactions, associated with disulphides, apparently do not affect the secondary structural content of the protein or its stability. This is surprising because disulphides, and the contraints they introduce in the polypeptide chain, might be envisaged to act as helix stop signals. Alternatively, disulphides may be required to stabilise helical structure in cases were helix formation depends on the additional stability provided by tertiary, albeit often non-specific contacts, from other regions of the polypeptide chain. This appears not to be a requirement in the TFE state, possibly because helices are formed in regions with sufficient intrinsic propensity (see Chapter 7). In addition, it is possible that the tertiary interactions provided through disulphides may be replaced efficiently by interactions with TFE cosolvent. The nature of these interactions is not understood at present (Chapter 4, 7 and Epilogue). Moreover, further experiments are needed to establish whether disulphide formation in the TFE state is favourable and whether particular disulphide pairings are preferred in the partially structured states, as seen in the case of the molten globule state α-lactalbumin (Ewbank & Creighton, 1993; Peng & Kim, 1994). However, preliminary work on he formation of disulphides in presence of TFE suggested that, as in the case of the reduced protein in aqueous solution (Goldberg et al., 1992), intermolecular association is dominant at the protein concentrations used for the experiments described in this chapter (Lu, Buck - unpublished data).

Comparisons with kinetic intermediates and other partially folded states.
It was suggested in the previous chapter, that the TFE denatured state may resemble folding intermediates populated transiently on the folding pathway of lysozyme in that it has a persistently structured α-subdomain in absence of a stably structured β-domain (Radford et al., 1992 & Chapter 4). A transient overshoot has been observed in the far uv CD spectrum during refolding of lysozyme from guanidine denaturant and has been assigned to strained conformations of disulphides or a contribution of aromatics to the spectrum of the folding intermediate (Chaffotte et al., 1992). The comparison of far uv CD spectra of HEWL and HEWL^{+SMe} in TFE demonstrates that disulphides are not a possible explanation of the increased ellipticity of the far uv CD in the TFE state itself. Furthermore, significant near uv CD in the TFE state of HEWL is not observed in the kinetic refolding experiments until both the α- and β-

subdomains of the protein have folded (Radford, 1992b) and therefore these observations appear to set the TFE denatured state apart from the kinetic folding intermediates which accumulate transiently on the pathway.

However, the CD and NMR spectra of the TFE denatured state of hen lysozyme suggest that it shares many characteristics with other partly folded states, including an increasingly diverse group of states referred to as "molten globules" (Ptitsyn, 1992; Dobson, 1994). Principal characteristics of these states include the observation of secondary structure in absence of extensive and specific tertiary contacts, and protection against hydrogen exchange. The data also imply that at least some of the secondary structure is non-native, which is typical of many cosolvent denatured states (Fan et al., 1993; Cox et al., 1994; Lui et al., 1994; Smith et al., 1994), and that considerable tertiary interactions exist, at least in presence of the native disulphide crosslinks. By contrast to other molten globule states (Jeng & Englander, 1991; Baldwin, 1993), non-specific tertiary interactions and compactness of the conformation may not be required for formation of persistent structures in the TFE denatured state of $HEWL^{+SMe}$, presumably because these can compensated by interactions with cosolvent. Similar to the majority of molten globule states secondary structure in the TFE denatured state is found to melt in a non-cooperative manner with temperature. By contrast to the TFE state, however, the denaturation of many molten globule states, with urea or guanidine displays some of the characteristics of cooperative transitions (Ptitsyn, 1992; Uversky et al., 1992, 1993). Further study of the nature of the cosolvent and its interactions with the polypeptide chain of intact proteins are likely to provide critical insights into this issue.

The studies presented in this chapter have allowed a further refinement of the model for the TFE denatured state and highlighted additional similarities as well as differences to folding intermediates and other partially folded states. With the advent of ^{15}N labelled hen lysozyme in the laboratory, the TFE denatured state of hen lysozyme proved to be a particularly tractable systems for the study of structure and dynamics in a partially ordered state of a protein by heteronuclear NMR. These studies are presented in Chapters 7 and 8. Application of these heteronuclear NMR techniques to the native state of hen lysozyme gave important insights into the relationship between structure and dynamics in a folded protein. This work is described in the following chapter.

References

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

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

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

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

Broadhurst, R.W., Dobson, C.M., Hore, P.S., Radford, S.E. & Rees, M.L. (1991) Biochem. 30, 405-412.

Chafotte, A.F., Guillou, Y. & Goldenberg, M.E. (1992) Biochem. 31, 9694-9702.

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

Cooper, A., Eyles, S.J., Radford, S.E. & Dobson, C.M. (1992) J.Mol.Biol. 225, 939-43.

Cox, J.P.L., Evans, P.A., Packman, L.C., Williams, D.H. & Woolfson, D.N. (1994) J.Mol.Biol. 234, 483-92.

Creighton, T.E. (1992) in Protein Folding, Creighton, T.E. eds. pp. 301-349. Freeman & Company, New York.

Dill, K.A. & Shortle, D. (1991) Annu.Rev.Biochem. 60, 795-825.

Dobson, C.M. (1994) Curr.Biol. 4, 636-640.

Dyson, H.J. & Wright, P.E. (1993) Curr.Opin.Struct.Biol. 3, 60-65.

Evans, P.A., Topping, K.D., Woolfson, D.N. & Dobson, C.M. (1991) Proteins: Struct.Funct. & Genetics 9, 248-266.

Ewbank, J. J., & Creighton, T.E. (1993) Biochem. 32, 3694-707.

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

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

Freedman, R.B. (1992) in Protein Folding, Creighton, T.E. eds., pp. 475-541, Freeman & Company, New York.

Goldberg, M.E., Rudolph, R. & Jaenicke, R. (1991) Biochem. 30, 2790-2797.

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

Heinrickson, R.L. (1970) Biochem.Biophys.Res.Comm. 41, 967-972.

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

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

Lee, C-L. & Atassi, M.Z. (1973) Biochem. 12, 2690-2695.

Lui, Z-P., Rizo, J. & Gierasch, L.M. (1994) Biochem. 33, 134-142.

Peng, Z. & Kim, P.S. (1994) Biochem. 33, 2136-2141.

Ptitsyn, O.B. (1992) in Protein Folding, Creighton, T.E. eds. pp. 243-300. Freeman & Company, New York.

Radford, S.E., Dobson, C.M. & Evans, P.A. (1992) Nature 358, 302-305.

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

Redfield, C. & Dobson, C.M. (1990) Biochem. 29, 7201-7214.

Sawano, H., Koumoto, Y., Ohta, K., Sasaki, Y., Segawa, S-I & Tachibana, H. (1992) FEBS.Lett. 303, 11-14.

Segawa, S-I., Fukuno, T., Fujiwara, K., & Noda, Y. (1991) Biopolymers 31, 497-509.

Taddei, N., Buck, M., Broadhurst, R.W., Stefani, M., Ramponi, G. & Dobson, C.M. (1994) Eur.J.Biochem. 225, 811-817.

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

Uversky, V.N. (1993) Biochem. 32, 13288-13298.

Uversky, V.N., Semisotnov, G.V., Pain, R.H. & Ptitsyn, O.B., (1992) FEBS Lett. 314, 89-92.

Woody, R.W. (1985) "Circular Dicroism of peptides" The peptides, vol. 7, 15-114.

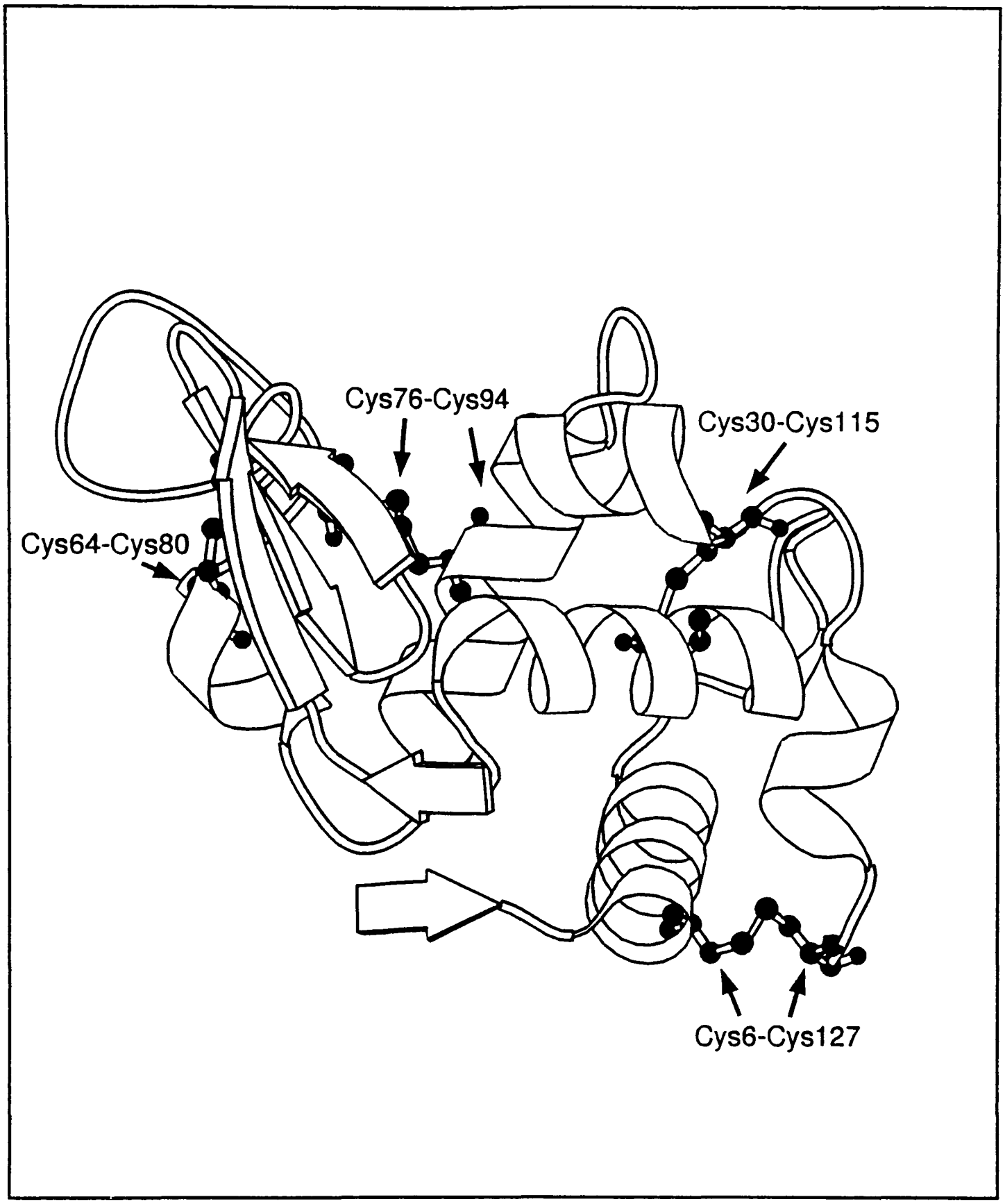
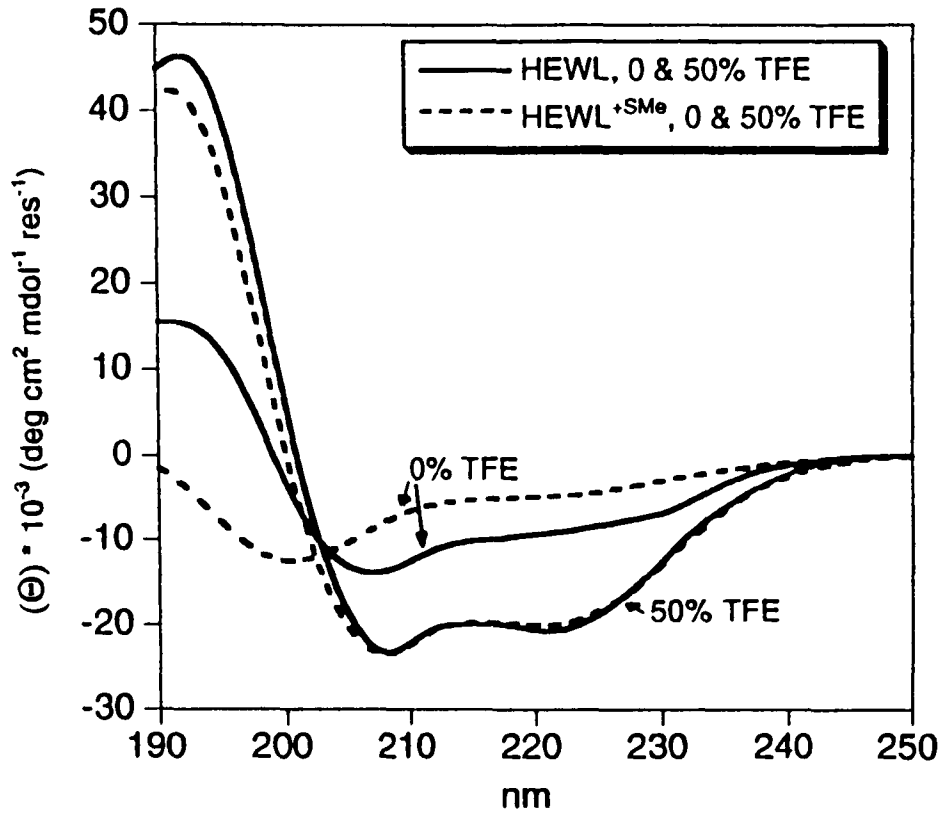


Figure 1
Schematic representation of the native fold of hen lysozyme using Molscript (Kraulis, 1991). The four disulphide bonds are shown.

a)



b)

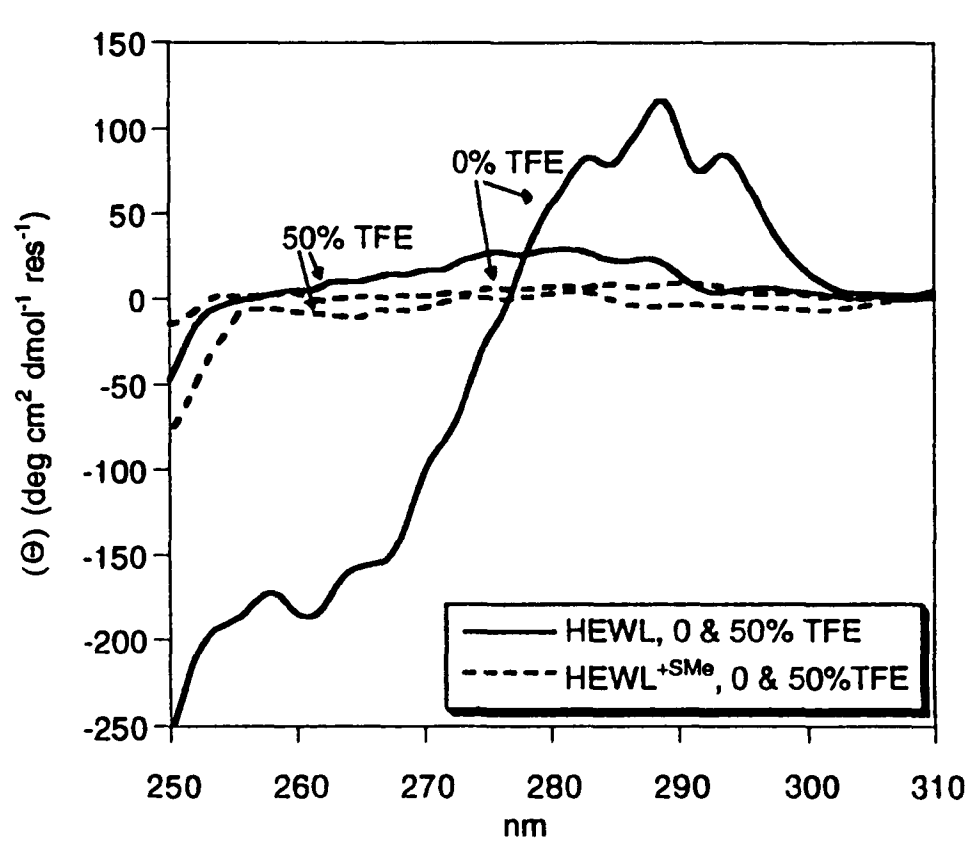
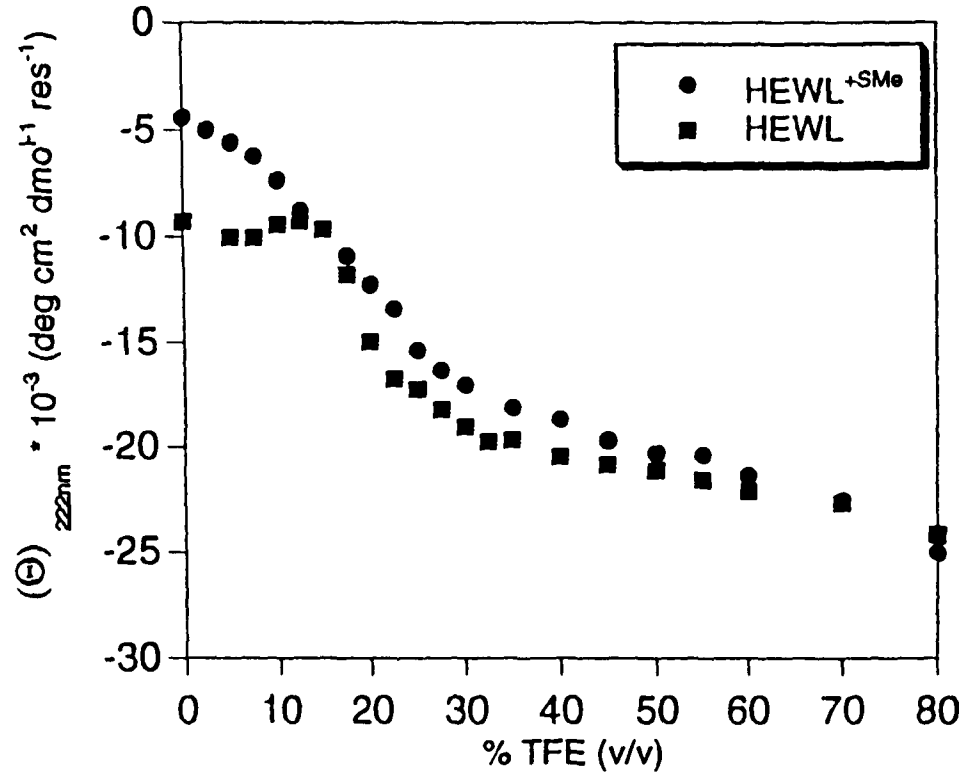


Figure 2

a) Far and b) near uv CD spectra of unmodified hen lysozyme (HEWL) and fully disulphide reduced and cysteine methylated hen lysozyme ($HEWL^{+SMe}$) in 0 and 50% TFE (v/v), pH 2.0, 27°C.

a)



b)

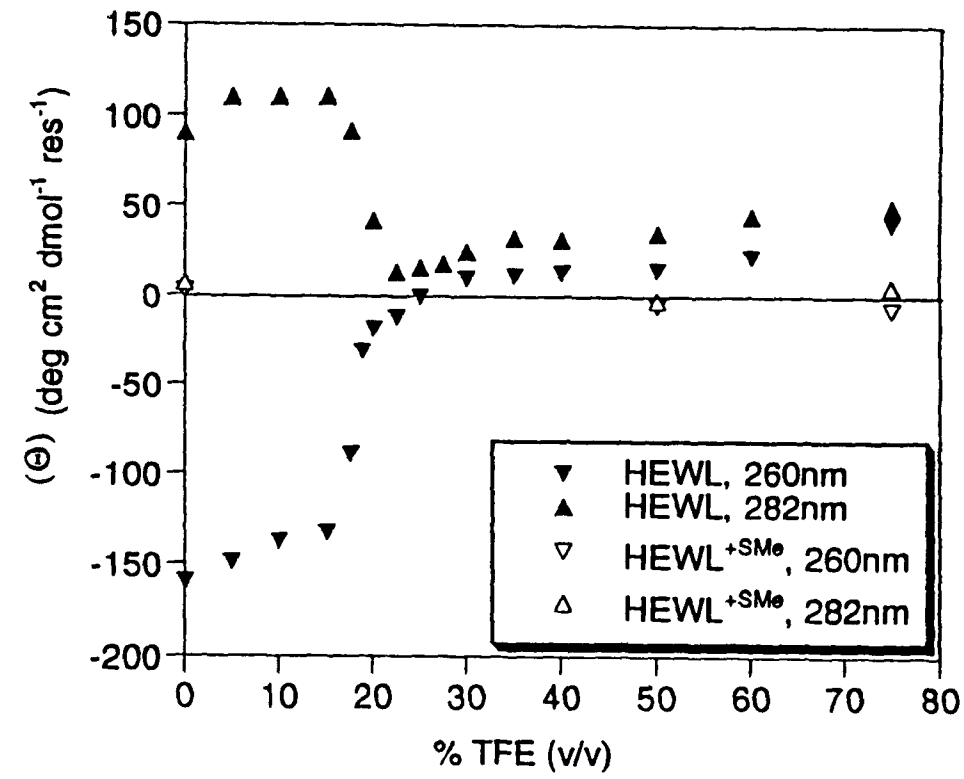
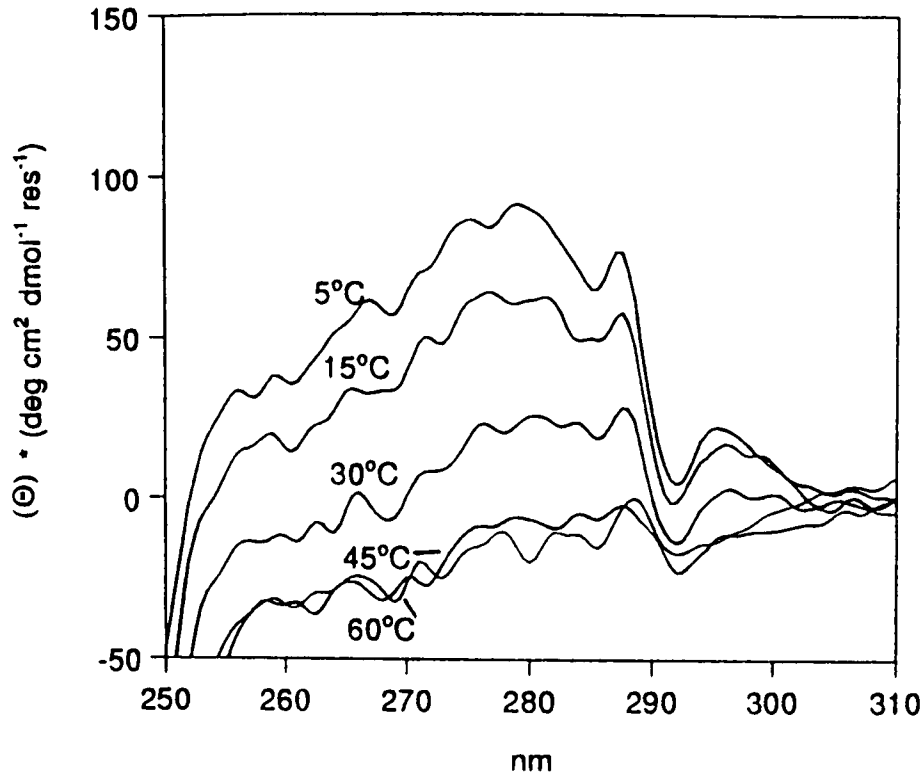


Figure 3
Structure monitored for HEWL and HEWL^{+SMe} at 27°C, pH 2.0 as a function of TFE concentration a) in the far uv CD at 222nm b) in the near uv CD at 260nm and 282nm.

a)



b)

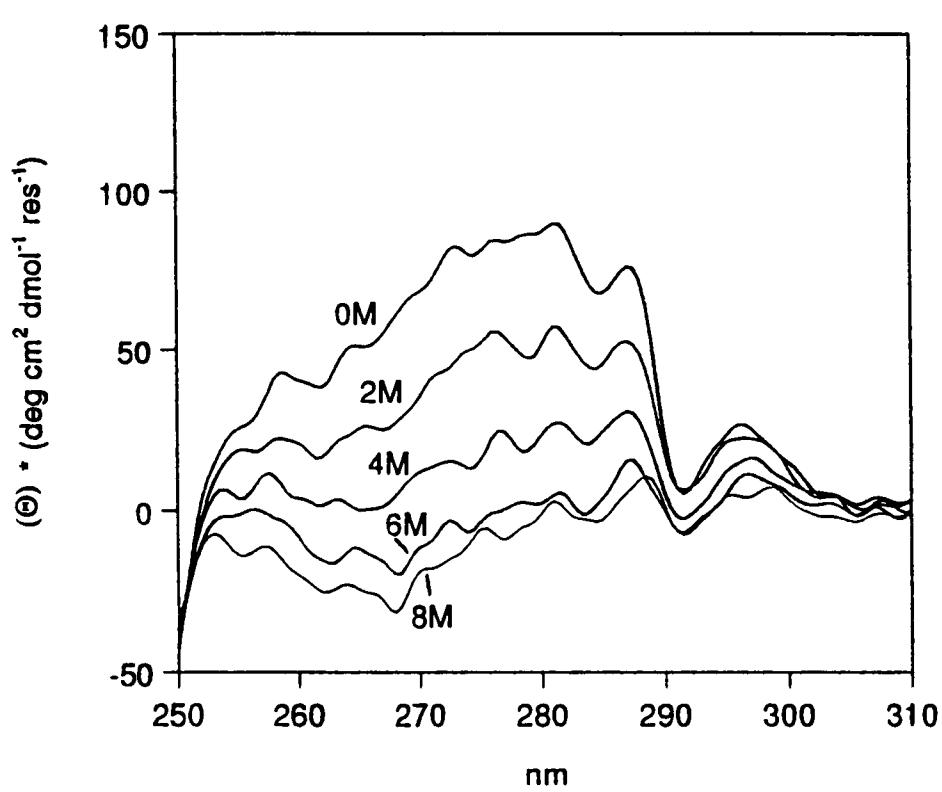


Figure 4

Near uv CD of HEWL as a function of a) increasing temperature and b) urea concentration at 5°C, both in presence of 50% TFE (v/v) at pH 2.0.

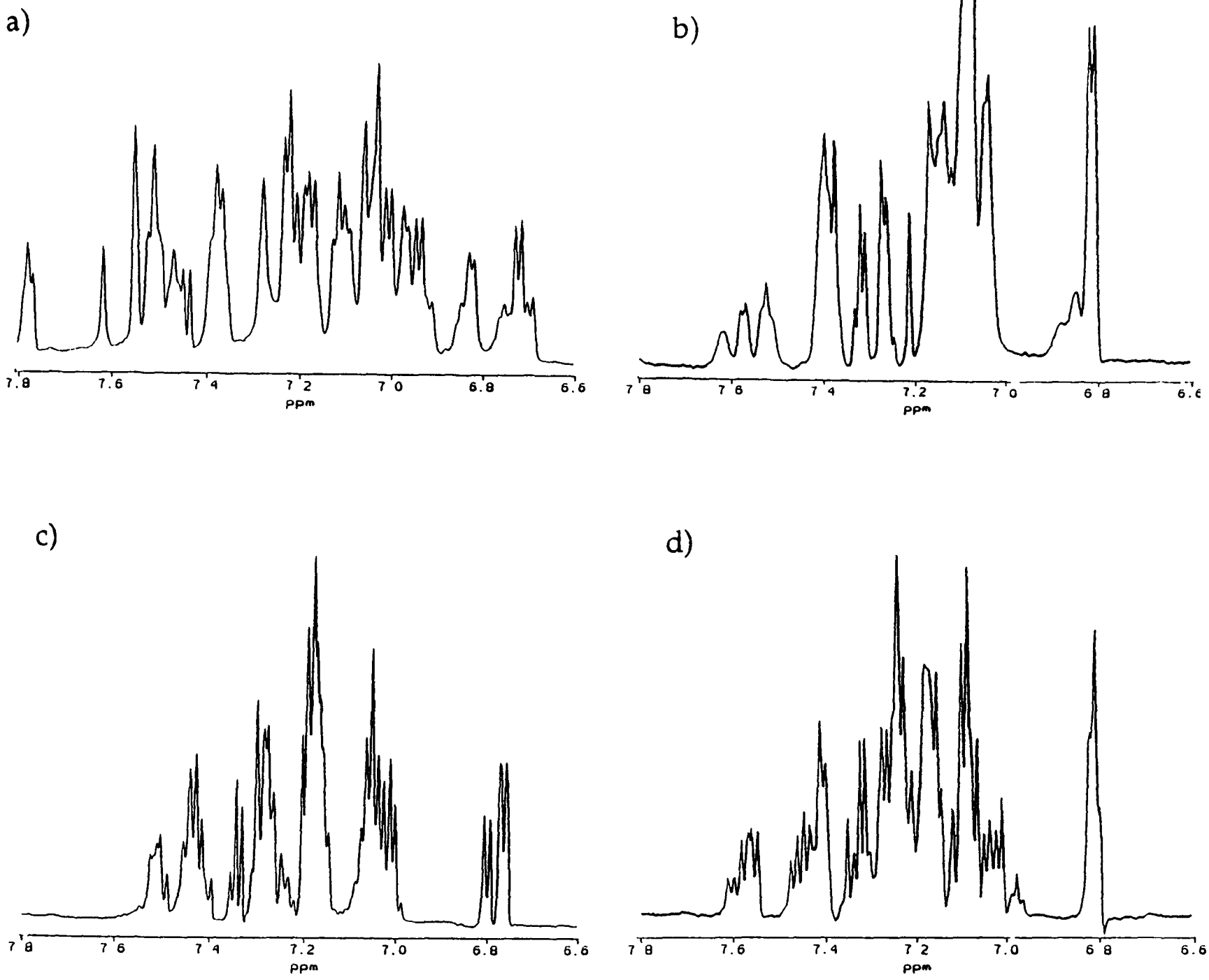


Figure 5
Aromatic region of the 1D ^{1}H NMR spectra of HEWL a) in absence, b) in presence of 50% TFE and HEWL^{+SMe} c) in absence, d) in presence of 50% TFE (v/v). The spectra were resolution enhanced using the same Gaussian window function.

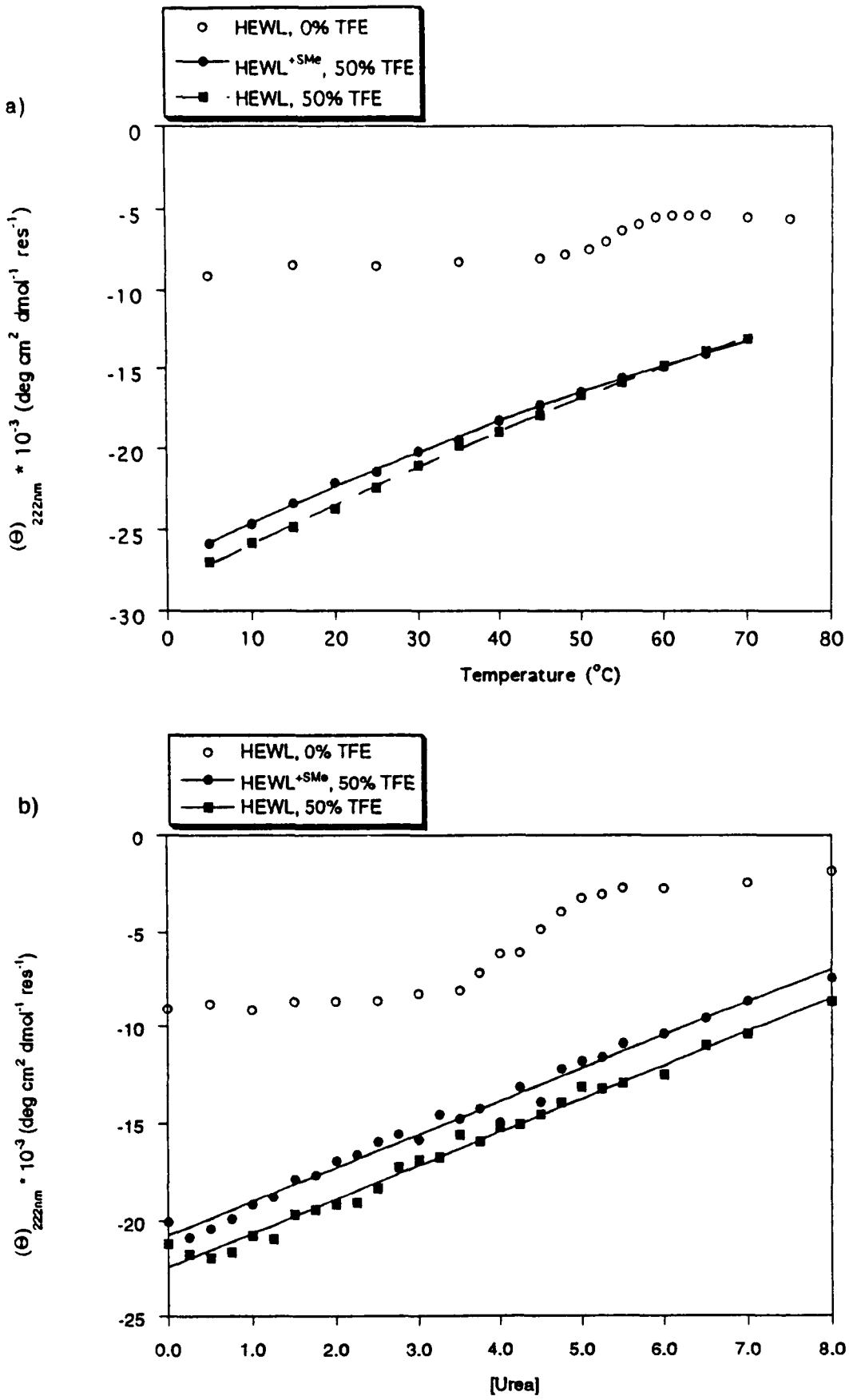


Figure 6

Denaturation of structure as monitored at 222nm in far uv CD of HEWL and $HEWL^{+SMe}$ as a function of a) increasing temperature and b) urea concentration at 27°C, both in presence of 50% TFE (v/v) at pH 2.0. The denaturation of HEWL in absence of TFE with temperature and urea denaturant is shown for comparison in a) and b) respectively.

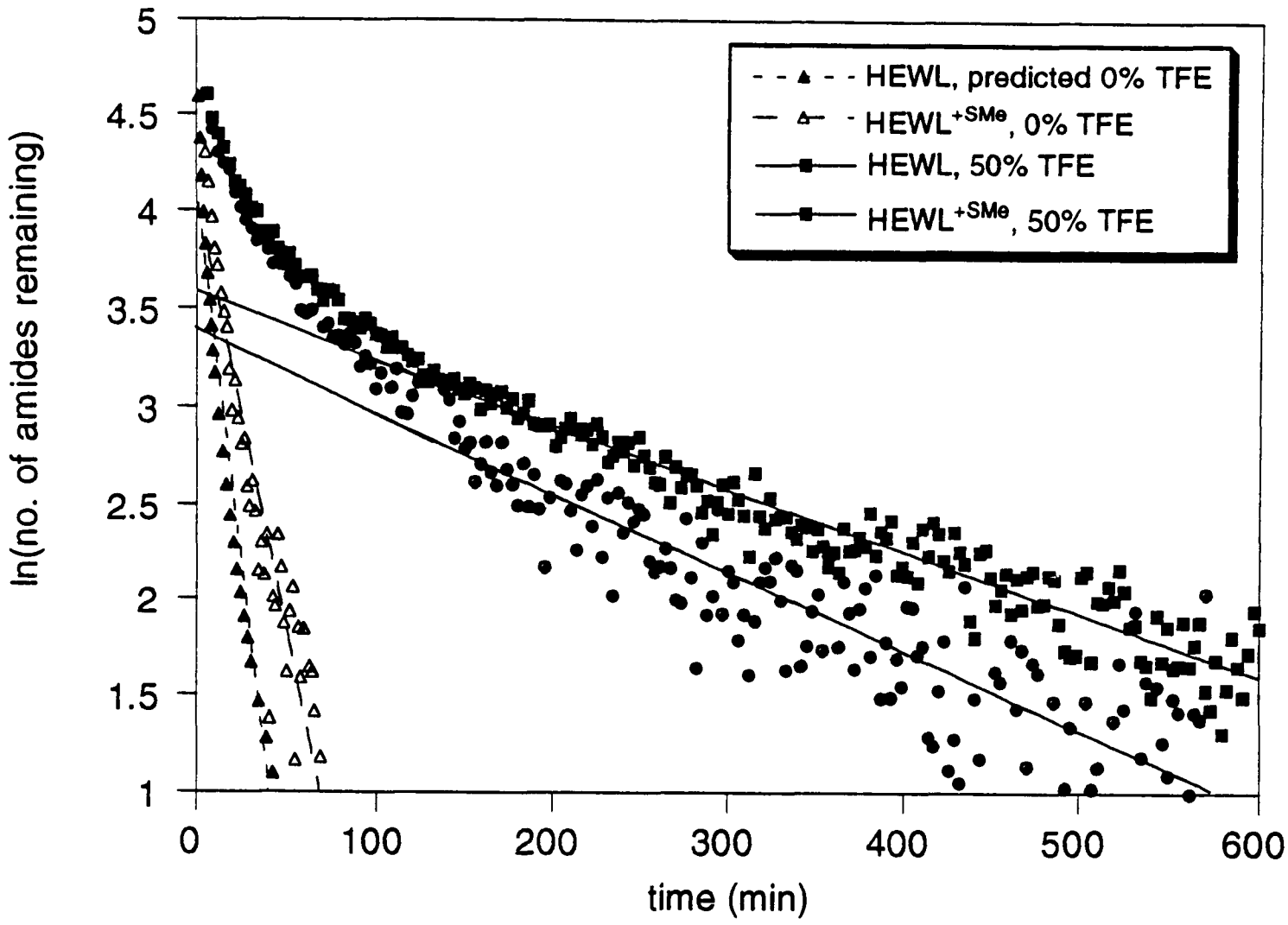


Figure 7: Bulk hydrogen exchange from HEWL and HEWL^{+SMe} in presence of 50% TFE at pH 2.0, 27°C compared to the bulk exchange profile from HEWL^{+SMe} under these conditions, but in absence of TFE and compared to the profile predicted on the basis of the intrinsic exchange rates (Bai et al. 1993). The exchange profiles are plotted as ln(no. of amides remaining) to show the multiexponential nature of the exchange from partly folded states in TFE. The slow phase of the exchange profiles (data points > 100min) are fitted to a linear equation. The rate constants are $3.3*10^{-3}$ min^{-1} and $4.1*10^{-3}$ min^{-1} for HEWL and HEWL^{+SMe} in 50% TFE respectively; $4.5*10^{-2}$ min^{-1} for HEWL^{+SMe} in 0% TFE and $7.7*10^{-2}$ min^{-1}, for the sum of intrinsic rates predicted for unfolded lysozyme, all at pH2.0, 27°C. The data are consistent with those in Chapter 4. Errors in amide intensity appear particularly large at longer time-points when plotted logarithmically because the number of amides remaining is small.

Chapter 6

Structural Determinants of Protein Dynamics: Analysis of ^{15}N NMR Relaxation Measurements for Mainchain and Sidechain Nuclei of Hen Egg-white Lysozyme

Summary

^{15}N labelled hen lysozyme has been studied by 2D and 3D NMR in order to characterise its dynamic behaviour. Resonances of all mainchain amide nitrogen atoms were assigned, as were resonances of nitrogen atoms in 28 sidechains. Relaxation measurments for the mainchain, arginine and tryptophan sidechain ^{15}N nuclei used standard methods, and for the ^{15}N nuclei of asparagine and glutamine sidechains used pulse sequences designed to remove unwanted relaxation pathways in the NH_2 groups. The calculated order parameters (S^2) show that the majority of mainchain amides undergo only small amplitude librational motions on a fast timescale ($S^2 \geq 0.8$). Increased mainchain motion ($0.5 < S^2 < 0.8$) is observed for a total of 19 residues located at the C-terminus, in loop and turn regions, and in the first strand of the main β-sheet. Order parameters derived for the sidechains range from 0.05 to 0.9; five of the six tryptophan residues have high order parameters ($S^2 \geq 0.8$) consistent with their location in the closely packed core of the protein, whereas the order parameters between 0.05-0.3 for arginine residues confirm increased sidechain mobility at the protein surface. Order parameters for the sidechains of asparagine and glutamine residues range from 0.2 to 0.8; high values are found for sidechains which have low solvent accessible surfaces and have well defined χ_1 values as measured by $^3J_{\alpha\beta}$ coupling constants. Many of the mainchain and sidechain groups with low order parameters have higher than average temperature factors in X-ray crystal structures and increased positional uncertainty in NMR solution structures. They also tend to lack persistent hydrogen bond interactions and protection against amide hydrogen exchange. The most significant correlations are found between residues with low order parameters and high surface accessibility in both crystal and solution structures. The results suggest that a lack of van der Waals contacts is a major determinant for sidechain and mainchain mobility in proteins.

Introduction

It is now accepted that protein molecules are not rigid entities but experience a variety of motions covering a wide range of timescales and amplitudes, from fast local oscillations about bonds to slower cooperative motions of segments of the overall structure (Karplus & McCammon, 1983; Gerstein et al., 1994). The manner in which these motions are related to molecular function is a subject of intense interest and some controversy. Molecular motion has been related to binding and catalytic events and to the mode of interaction of proteins with other macromolecules such as receptors, antibodies and nucleic acids. As well as the relationship of molecular motion to function, it is clear that the dynamical properties of proteins are intimately related to their stability and the manner in which they fold (Richards, 1992; Dobson, 1993). Unfolded and partly folded proteins have very extensive mobility compared to the native state and only when the specific interactions between sidechains are established, and close packing of different sections of the polypeptide has been achieved, does the characteristic order of a globular protein structure appear. Information about the dynamic properties of the native state as well as denatured states should therefore provide clues as to the nature of the interactions within proteins and their influence on structure and stability. Despite the importance of these in fields such as protein modelling and design, the structural determinants for protein dynamics are not well established.

Of particular value in the investigation of molecular dynamics is NMR spectroscopy (Dobson & Karplus, 1986; London, 1989). ^{15}N relaxation measurements have recently become an attractive means of studying mainchain dynamics of proteins in solution (Kay et al., 1989; Clore et al., 1990; Peng & Wagner, 1992). The analysis of the relaxation data in conjunction with the model free approach of Lipari & Szabo (1982) yields order parameters (S^2) commonly interpreted as a measure of the geometrical restriction of the ^{15}N-^{1}H bond vectors on a time scale faster than the stochastic rotational tumbling of the molecule in solution. Molecular dynamics simulations have reproduced experimentally determined order parameters with some success (Olejniczak et al., 1984; Chandrasekhar et al., 1992). The main strategy for the interpretation of order parameters has become focused on their magnitude, their variation along the polypeptide chain, and their relationship with other structural parameters. For example, good correlations between order parameters and B-factors of X-ray structures or rmsd values from ensembles of NMR structures have been reported for at least some of the proteins studied (Kördel et al., 1992; Redfield et al., 1992; Grasberger et al., 1993; Powers et al. 1993). Similarly, a

tendency of order parameters to be related to secondary structure or hydrogen bonding was suggested in several but not all studies (e.g. Schneider et al., 1992; Kay et al., 1989; Stone et al., 1993).

In order to identify features which may account for dynamic behaviour, we have derived order parameters from ^{15}N relaxation data for a protein for which a wide range of structural data is available. Hen egg-white lysozyme was the first enzyme whose three dimensional structure was determined by X-ray crystallography (Blake et al., 1965). Since then a series of crystal structures showing essentially the same conformation have been refined to high resolution (Handoll, 1985; Ramanadham et al., 1990; Kundrot & Richards, 1991; Wilson et al., 1992; Young et al., 1994). Recently, restraints from 2D ^{1}H NMR measurements have been used to calculate an ensemble of NMR structures in good agreement with the X-ray structures (Smith et al., 1993). The protein has a globular fold and its structure, shown schematically in Figure 1, consists of two subdomains, predominantly comprised of either α- or β-secondary structure, with the active site cleft at their interface. Although hen lysozyme is known to unfold cooperatively under equilibrium conditions, the two subdomains appear to become structured largely independently during kinetic refolding (Radford et al., 1992a).

The dynamic behaviour of hen lysozyme in crystals has been investigated by analysis of B-factors. In order to attempt to differentiate between static and dynamic disorder, structures were solved at different temperatures, pressures and levels of hydration (Kundrot & Richards, 1987; Kodandapani et al., 1990; Young et al., 1994), in different crystal forms (Sternberg et al., 1979; Ramanadham et al., 1989; Young et al., 1994) and of different species (Artymiuk et al., 1979). Mobility in the native state of the protein has also been inferred from changes in the conformation of the enzyme on antibody or ligand binding (Sheriff et al., 1987; Strynadka & James, 1991). However, comparison of hydrogen exchange behaviour in the crystalline state and in solution (Pedersen et al., 1991) has suggested that at least some fluctuations may be significantly dampened by crystal packing contacts.

NMR spectroscopy has also been used to investigate the dynamic behaviour of hen lysozyme. Structural variation amongst the members of the ensemble of NMR structures arises as a consequence of insufficient or conflicting experimental restraints, some of which reflect the dynamic behaviour of the protein in solution but some reflect limitations in the experimental ^{1}H NMR methods for a protein of this size (Smith et al., 1993). Direct evidence for

structural dynamics was obtained by detection of aromatic ring flips (Campbell et al., 1975), ^{1}H cross-relaxation dynamics (Olejniczak et al., 1984) and analysis of hydrogen exchange (Radford et al., 1992b; Thomson & Poulsen, 1993) and by measurement of coupling constants (Smith et al., 1991; Bartik et al., 1993). These NMR studies concluded that the mainchain of hen lysozyme is well ordered, but that sidechains, particularly at the surface of the protein are undergoing fluctuations which are significantly greater than is suggested by their definition in crystal structures (Smith et al., 1991).

Parameters calculated from molecular dynamics simulations of hen lysozyme have been compared with experimental data such as diffuse x-ray scattering (Clarage et al., 1992; Faure et al., 1994) and NMR proton relaxation measurements (Olejniczak et al., 1984). Such simulations indicate that fluctuations of the loop regions are significantly greater than those of the helices and β-sheet (Post et al., 1989). In addition, computational techniques have been used to investigate the possibility of hinge bending (Bruccoleri et al., 1986; Gibrat & Go, 1990). Its existence was suggested by the small reorientation of the subdomains in crystals on substrate binding (Imoto et al., 1972), but little additional evidence has been found in support of such motion in hen lysozyme.

In order to investigate these phenomena in more detail we report in this paper relaxation measurements and order parameters for the mainchain ^{15}N nuclei of lysozyme and for a comprehensive set of sidechain ^{15}N nuclei. Many of these sidechains are located at the protein surface. The relaxation data are compared with the rmsd values derived from the ensemble of NMR structures and B-factors from several crystal structures. Hydrogen bonding and surface accessibility are investigated as possible primary determinants of the mobility of NH and NH_2 groups.

Materials and Methods

Collection of NMR data

Hen egg white lysozyme was expressed in *Aspergillus niger* using $^{15}NH_4Cl$ as the sole nitrogen source and purified from filtered culture medium (Roberts et al., 1992) with the following modifications. *A. niger* transformant B1, an improved lysozyme production strain, was used. This strain has been transformed with a glucoamylase-lysozyme fusion vector (Jeenes et al., 1993) and produces about 60mg ^{15}N-labelled lysozyme/L. Prior to FPLC purification on Mono S, lysozyme was isolated from filtered culture medium using ion exchange chromatography on ProductiveTMS cartridge columns (bps Separations Ltd., Durham, UK). The NMR samples were prepared to contain ca. 4mM protein in 95% H_2O/5% D_2O (v/v) and the pH was adjusted to 3.8. NMR experiments were performed at 35°C on three homebuilt NMR spectrometers belonging to the Oxford Centre for Molecular Sciences with ^1H operating frequencies of 500.1, 600.2 and 750.2 MHz. The ^{15}N assignments were made using data recorded from a single 3D NOESY-HMQC experiment (Messerle et al., 1989) with a mixing time of 200 ms, in conjunction with the known ^1H assignments (Redfield and Dobson, 1988). The data set comprised (t_1, t_2, t_3) 128, 32, 512 complex points recorded with a digital resolution of 42.7, 37.8, 12.4 Hz/point respectively at 500 MHz and was zero filled to 256, 64, 1024 real points.

Pulse sequences for the measurement of the longitudinal (T_1) and transverse (T_2) relaxation time and the heteronuclear nuclear Overhauser effect (NOE) of ^{15}N nuclei in NH groups have been described previously (Kay et al., 1989; Boyd et al., 1990; Palmer et al., 1992). Suppression of the solvent resonance was achieved by low power presaturation or by exploiting RF inhomogeneity (Messerle et al., 1989). The T_1 and T_2 relaxation measurements used a series of 9 experiments with relaxation delays ranging from 20 to 1400 ms and 8.5 to 340 ms respectively. A recycle delay of 4.0 s was used for the T_1 relaxation and NOE experiments and 2.4 s for the T_2 experiments. The data sets were acquired using 128 complex data points in T_1 with resolution of 9.5 Hz/point.

Experiments involving $^{15}NH_2$ resonances used the pulse sequences described in Figure 2. These data were recorded at 600 MHz with 64 complex data points in t_1 giving a resolution of 11.4 Hz/point. The T_1 and T_2 data sets were recorded with a digital resolution in t_2 of 6.2 Hz/point and the NOE experiments were recorded using a digital resolution of 7.4 Hz/point. The 2D spectra were zero filled once in F_2 and twice in F_1 and were processed with Felix (Hare Research

and Biosym, Inc.). The residual water signal was removed during data processing by application of a low frequency deconvolution in t_2 (Marion et al., 1989).

Analysis of NMR relaxation data

The heteronuclear NOE effect was calculated as the ratio of peak heights in spectra recorded with and without ^1H saturation. Rates R_1 ($=1/T_1$), and R_2 ($=1/T_2$), were fitted as single exponential decays to peak height data; uncertainties were estimated as described previously (Palmer et al., 1991; Stone et al., 1992). The errors in the primary intensity data were taken as 1-3 times the spectral noise. The ^{15}N relaxation parameters obtained were analysed using the formalism of Lipari & Szabo (1982) with the extension of Clore et al., (1990). In the case of analysis of the relaxation parameters for ^{15}N in NH_2 groups the dipolar contribution to spectral density functions was increased by a factor of two to account for the presence of two ^1H nuclei. The appropriate model for the spectral density function used to calculate T_1, T_2 and NOE was selected according to the criteria of Clore et al., (1990) on the basis of the observed T_1/T_2 ratio and the magnitude of the NOE effect. Parameters calculated using the appropriate spectral density function, S^2, τ_e, Δex for simple models and S_s^2, S_f^2 and τ_s for the extended model, were optimised by minimisation of the rmsd between the calculated and the experimental relaxation data using published procedures (Clore et al., 1990; Palmer et al., 1991; Stone et al., 1992; Constantine et al., 1993). The order parameters and their uncertainties were estimated by Monte Carlo sampling using MODELFREE2.1 (Arthur G. Palmer, Columbia University) and in-house software.

Analysis of crystal and NMR structures

Six crystal structures of hen lysozyme were analysed, five of which are deposited in the Brookhaven Databank. 2lym (Kundrot & Richards, 1987), 2lzt (Ramanadham et al., 1990) are tetragonal and triclinic structures both at 2.0Å resolution; 4lyt and 6lyt (Young et al., 1994) are monoclinic and tetragonal structures at 1.9 Å resolution, and 1hel (Wilson et al., 1992) is a tetragonal lysozyme structure solved to 1.7 Å resolution. Hentyp2 is a tetragonal structure refined to 2 Å by Handoll (1985) on the basis of the original data of Blake et al., (1967) and has been used frequently for comparisons with NMR data. B-factors of individual nitrogen atoms were used for comparison of mainchain and tryptophan sidechain dynamics in crystals with the ^{15}N relaxation data. Positioning of arginine, asparagine and glutamine sidechain atoms can be more ambiguous in X-ray structures, and the average of B-factors for the groups of $N^{\varepsilon}H+(N^{\eta}H_2)_2$ and of $O+N^{\delta/\varepsilon}H_2$ atoms were used for arginine $N^{\varepsilon}H$

and for asparagine and glutamine $N^{\delta/\varepsilon}H_2$ sidechain groups, respectively. Global and local (superposition window size of 2-5 residues) rmsd values were calculated for the mainchain superposition of the 6 crystal structures and for the ensemble of NMR structures (Smith et al., 1993).

Accessible surface area of the heavy atoms in the structures was calculated using the van der Waals radii of Chothia (1976) and a 1.4Å radius probe in an implementation of the algorithm of Lee and Richards (1971) as part of the program NACCESS (Dr. Simon Hubbard, EMBL, Germany and Prof. Janet Thornton, University College London, UK). All hydrogen bond interactions involving a particular atom were determined in the crystal structures and their energies were calculated explicitly using X-PLOR (Brünger, 1992). When several hydrogen bonds are found to a particular atom their energies were summed. A stable hydrogen bond is taken to exist if this energy is more negative than -0.5 kcalmol^{-1} (Kabsch & Sander, 1983). Hydrogen bond interactions are considered as a barrier to mobility and therefore plotted as an "activation energy" term, exp(H-bond energy), throughout, giving better correlations with S^2 than the linear relationship. The peptide bond unit (C_{i-1}, O_{i-1} and N_i) and the group of sidechain atoms $O+N^{\delta/\varepsilon}H_2$ for asparagine and glutamine and $N^{\varepsilon}H+(N^{\eta}H_2)_2$ for arginine, comprise structural units whose mobility is likely to be determined in a concerted manner by the surrounding environment. For example, hydrogen bonding with one $N^{\eta}H_2$ group of an arginine residue is likely to affect mobility of the entire sidechain end unit, but the effects on accessibility and hydrogen bond energy are predominantly on the one of the three donating nitrogen atoms involved. Therefore, the absolute accessibilities (in Å^2) and hydrogen bond energies are reported as averaged over these atom groups. This grouping yielded better correlations with S^2 compared to the parameters of the single atoms. The B-factors, accessibilities and hydrogen bond energies were calculated for individual structures and then averaged to yield their mean values over the 6 crystal structures and over the ensemble of NMR structures.

Results

^{15}N Resonance assignments

^{1}H NMR assignments previously reported for hen lysozyme (Redfield & Dobson, 1988) were used together with a 2D ^{1}H COSY experiment as the starting point for ^{15}N resonance assignment. All the remaining information was provided by a single 3D NOESY-HMQC spectrum (Messerle et al., 1989) which allowed connectivities to be traced along the entire mainchain using next neighbour NOE connectivities. Assignments for the ^{15}N-^{1}H correlations for the 126 mainchain amides and for the 6 tryptophan indole ^{15}N-^{1}H correlations in the HSQC spectrum of hen lysozyme are shown in Figure 3. Additional resonances in the HSQC spectrum belong to the sidechains newly assigned in this study; these include pairs of resonances in the region around 112 p.p.m. which arise from sidechain $^{15}NH_2$ groups of the 14 asparagine and 3 glutamine residues in hen lysozyme. Resonances in the region boxed around 108 p.p.m. are folded into the spectrum and arise from arginine $^{15}N^{\varepsilon}$-^{1}H sidechain correlations. These residues could be assigned in the 3D spectrum by their NOEs to mainchain and sidechain protons of their own and neighbouring residues, and in the case of 5 arginines also by use of through bond information. None of the resonances arising from arginine $(^{15}N^{\eta}$-$^{1}H_2)_2^+$ sidechain correlations is resolved and they are likely to correspond to the broad resonances near 118p.p.m. which are folded in the HSQC spectrum. [Table 1 presents a full listing of the ^{15}N and additional ^{1}H assignments made in this study.] The ^{1}H chemical shifts of resonances identified in the spectra are in agreement with those reported previously (Redfield & Dobson, 1988) to within ± 0.06 p.p.m.. Assignments for mainchain ^{15}NH and asparagine and glutamine sidechain $^{15}NH_2$ resonances have also been reported for human lysozyme by Ohkubo et al. (1991). As has been found for ^{1}H chemical shifts (Redfield & Dobson, 1991), there is a good correlation between ^{15}N mainchain and sidechain chemical shifts, in particular for the residues which are conserved in these two homologous proteins.

Relaxation data and order parameters for main chain ^{15}N nuclei

Relaxation parameters were measured for 121 of the 129 mainchain ^{15}N nuclei of hen lysozyme (Figure 4). A value for the global rotational correlation time, τ_R, was obtained from the average T_1/T_2 ratio (Clore et al., 1990) of 100 mainchain ^{15}N nuclei which did not show evidence for significant fast motion or exchange broadening. The mean T_1/T_2 ratio was 3.32 ± 0.13 and yielded an optimised value for τ_R of 5.7 ± 0.2 ns. This value is consistent with previous

experimental data for hen lysozyme after correction for temperature and viscosity effects (Olejniczak et al., 1984) and compares well with the τ_R value of 6.0 ns which can be calculated for hen lysozyme under these experimental conditions (Cantor & Schimmel, 1980).

The mainchain order parameters S^2, the internal correlation times τ_e or τ_s, and the magnitude of the exchange effects Δex are plotted as a function of the protein sequence in Figures 4 and 5 [experimental data and derived parameters are listed in Appendix S-II]. The uncertainties estimated for these parameters are found to be small for S^2 (< 4%), larger for Δex (typically 10-30%) and much larger for certain τ_e values (several exceeding 100%). These estimated uncertainties correspond in magnitude to those reported for other proteins (Palmer et al., 1991; Constantine et al., 1993). It is notable, comparing Figures 4 and 5, that the profile of S^2 in the range 0.75-0.95 is predominantly determined by the variations in the T_1 and T_2 data, the NOE value being uniformly high. By contrast, values of S^2 below 0.75 reflect slower motions of increased amplitude which result in a reduced magnitude for the NOE. Residues with such behaviour are not randomly distributed in the protein (Figure 4). It is noticeable that residues undergoing motions described by similar values of the internal correlation time, τ_e, are clustered in the A-helix, A-loop-B and B-helix region as well as the C-helix and loop-to-D-helix region, all located in the α-subdomain of the protein (Figure 1). Residues in the latter region are undergoing motions of increased amplitude and one residue (Asn103) undergoes motions which require an extended form of the model for the spectral density function by Clore et al., (1990) involving internal motions on two timescales. Furthermore motions on a μs to ms timescale which result in an exchange contribution to relaxation are clustered in the region of the first 3_{10} helix and the D-helix. Such clustering is indicative of the regional nature of the environments of the polypeptide chain and may, at least in part, reflect segmental flexibility. Residues Gly22 and Ser85 have order parameters significantly different from their neighbours. At present the structural reasons for their relaxation behaviour are not understood. However, it is interesting to consider that both are located near the termini of helices.

The order parameters, S^2, derived from the analysis of the mainchain ^{15}N relaxation data are greater than 0.80 for the majority of residues (Figure 1, 6). The protein backbone is, therefore, well ordered except for a few mainchain ^{15}N atoms which undergo significant motions associated with S^2 below 0.70 for a few residues in hen lysozyme. These are located in the long loop (residues 65-75) and in a short loop connecting the C- and D-helices (residues 101-104), the

C-terminal residue as well as residue 85. The lowest order parameters are around 0.5. Therefore lysozyme does not belong to the class of proteins which includes, for example, staphylococcal nuclease (Kay et al., 1989) that lack any appreciable internal mobility; with the exception of residues in the termini no value of S^2 for the latter protein is below 0.8. Lysozyme can be best compared to proteins, such as ubiquitin, thioredoxin and interleukin-8 (Schneider et al., 1992; Stone et al., 1993; Grasberger et al., 1993 respectively), in which several residues, usually located in loop regions, are undergoing motions of moderate amplitude (lowest value of S^2 usually > 0.5). By contrast, in several proteins for example interleukin-1β, interleukin-4, calbindin, calmodulin and eglin c (Clore et al., 1990; Redfield et al., 1992; Kördel et al., 1992; Barbato et al., 1992; Peng & Wagner, 1992, respectively) continuous stretches of residues are found with values of S^2 close to or less than 0.5, substantially reduced from those of the remainder of the protein.

Proteins with regions characterised by low order parameter values all possess long and disordered loops.Correlations of S^2 with other structural parameters have been made most convincingly for this class of protein. By contrast, most of the loops and turns in hen lysozyme are short. Turns joining strands of the β-sheet and loops joining the two 3_{10}-helices to other structural elements are 4 residues long, the loops which join α-helices A-B and C-D are 7 and 8 residues long respectively. There is, however, one long loop comprising residues 61 to 78 in hen lysozyme. Of these loops only the C-D loop and the centre of the long loop appear to undergo significant fluctuations on a timescale faster than that of the overall rotational tumbling of the protein. Hinge bending fluctuations modelled by dynamics simulations (Bruccoleri et al., 1986; Gibrat & Go, 1990) have periods of approximately 20 ps and are within the frequency range contributing to the NMR relaxation parameters studied here. The amplitudes of the fluctuations are small, however, and predicted to be less than 3°. Such an angle would cause a maximal atomic displacement of 1 Å. Theoretical analysis suggests that order parameters described here would not be sufficiently sensitive to monitor such small hinge bending dynamics, which are obscured by more local fluctuations (Xiang Li & Raphael Brüschweiler, personal communication).

Relaxation data and order parameters for side-chains

Relaxation data for ^{15}N nuclei in tryptophan and arginine sidechain NH groups were extracted from the same spectra and using the same procedure as for those of the mainchain NH groups [experimental data are listed in Appendix S-II]. The data for the tryptophan $^{15}N^{\varepsilon}$ atoms are similar to those

obtained for the majority of mainchain atoms, with the exception of Trp62 which has a lower NOE value of 0.25 (S^2 of 0.56). NOE ratios for 5 arginine $^{15}N^{\varepsilon}$ atoms are, by contrast to those of tryptophan sidechains, significantly smaller, all values being less than 0.04 and in some cases being negative, such as Arg125 with the lowest NOE value in this study of -2.8. In addition, T_1 and T_2 relaxation time constants are longer than those for typical mainchain nuclei suggesting that the arginine sidechains are undergoing considerable fluctuations. Order parameters for these sidechains are shown in Table 2(i,ii).

The relaxation of ^{15}N nuclei in NH_2 groups of asparagine and glutamine sidechains is more complicated than that of ^{15}N in NH groups, and pulse sequences designed to monitor selectively the ^{15}N longitudinal and transverse relaxation of these sidechain nuclei were used in their measurement (see Materials and Methods, Figure 2 and Table 2). For this purpose a simplified model was adopted to describe the relaxation properties of the ^{15}N nucleus in the asparagine and glutamine side chains. Justifications for the use of pulse sequences in Figure 2 have recently been presented (Boyd, submitted for publication). The ^{15}N nucleus is considered to be part of an isolated weakly coupled 3-spin system where the two heteronuclear one-bond coupling constants and bond lengths are assumed equal. However, the two protons have distinct chemical shifts, giving rise to an I_1I_2S spin system where $J(I_1I_2) << \Delta\nu(I_1I_2)$. The relaxation mechanisms which must be considered in addition to those possible in N-H spin systems include dipolar auto and cross-correlation, chemical shift anisotropy of the ^{15}N nucleus (including CSA-dipole cross-correlation) and a chemical exchange term defining the intramolecular bond rotation. The intramolecular chemical exchange is not anticipated to contribute to relaxation of either the single spin transverse or longitudinal processes. The importance of cross-correlation terms, both ^{15}N CSA-dipolar and dipole-dipole, must, however, be considered. The magnitude of these cross-correlations will dictate the extent of the deviation from an exponential longitudinal or transverse relaxation process for the ^{15}N nucleus. The experimental protocols developed for the suppression of the ^{15}N CSA-dipolar cross-correlation process in two-spin heteronuclear systems were found to be effective here also (Kay et al., 1989; Palmer et al., 1992). The deviations from exponential behaviour caused by the dipolar cross-relaxation terms (Overhauser effects), are small and may be neglected. Pulse sequences shown in Figure 2a,b create spin operators as a result of the dipole-dipole cross-correlation process during the relaxation delay which would, if significant, lead to observable magnetisation and to amplitude distortion. However, the magnitude of the dipole-dipole cross-correlation process between the pair of

internuclear vector $^{15}NH_1$ and $^{15}NH_2$ is scaled by the angular term $(3\cos^2\theta_{NH1NH2} -1)/2$ which for $\theta_{NH1NH2} = 120^o$ is 0.125. Simulations using a rotational correlation time of 5.7 ns appropriate for hen lysozyme show that this cross-correlation process does not strongly affect the longitudinal or transverse recovery. Therefore, the very small deviations from exponential behaviour introduced by this dipole-dipole cross-correlation process were ignored, in both cases, and the relaxation treated as an exponential decay. The order parameters calculated from the data (listed in Appendix S-II) are shown in Table 2(iii); in contrast to the order parameters for the mainchain or the tryptophan and arginine sidechains, the order parameters for asparagine and glutamine sidechains show considerable variation in their values, ranging from 0.16 to 0.82.

The dynamic behaviour of the sidechains studied here can be grouped into three classes by their S^2 values. The first class of sidechains (S^2 0.8-0.9) comprises 5 of the 6 tryptophan and 1 of the 14 asparagine sidechains. These S^2 values are similar to those obtained for the majority of mainchain residues. Similar results have been reported for the two tryptophan sidechains in thioredoxin which were studied by ^{15}N relaxation (Stone et al., 1993). A second class of sidechains (S^2 0.4-0.7) comprises Trp62 and the majority of asparagine and glutamine sidechains, which are located at or near the protein surface. A third class of sidechains undergoes significantly greater fluctuations, having S^2 values less than 0.4. It comprises those of the remaining 2 asparagine and 2 glutamine residues and all arginines. The location of these sidechains is shown in the hentyp2 crystal structure in Figure 1; the classification of sidechains by their dynamics is indicated by shading. This classification of sidechain dynamics is not in general simply by amino acid type. A report of NOE measurements in eglin c suggested that two of the four arginine residues in the protein were significantly restrained in their mobility (Peng & Wagner, 1992). However, the data of the present study indicate that all arginine sidechains of hen lysozyme for which measurements were made experience considerable fluctuations.

Discussion

Correlation of main chain S^2 values with features of X-ray and NMR structures. Good correlations of mainchain S^2 values with B-factors have been reported in several studies (Clore et al., 1990; Kördel et al., 1992; Grasberger et al., 1993; Powers et al., 1993). Figure 5a,b shows the comparison of the experimental order parameters with B-factors of the crystal structure of hen lysozyme frequently used in comparison with NMR data (referred to as hentyp2). It is apparent that all regions which have high B-factors in this tetragonal crystal structure also have moderately low values of S^2. However, the converse does not appear to hold, as residues 83-88 have low S^2 values but do not appear to have B-factors higher than the average. However, B-factors vary considerably between the 6 crystal structures used in this analysis, especially when comparing structures solved in different crystallographic space groups (Table 3a), and appear to depend, at least in part, on the precise experimental conditions. However, several regions of the mainchain, for example the long loop residues 61-78, have increased B-factors in the majority of crystal structures. This suggests that B-factors do indeed reflect internal dynamics, except when they are dampened by intramolecular contacts (Artymuik et al., 1979; Handoll, 1985; Smith et al., 1986; Young et al., 1994).

In order to explore this further, a mean B-factor was calculated by simply averaging the reported B-factors from the 6 crystal structures, shown in Figure 6b. As shown in Table 3b, such average B-factors show a better correspondence with S^2 values than those taken from individual structures. Furthermore, the variations both in S^2 values and B-factors are largely regional, suggesting that averaging over the structural segments of hen lysozyme may lead to further improvement of the correlations. The correlation of S^2 with these regional B-factors, averaged over 17 structural units of the protein (Handoll, 1985 and Table 4), is shown in Figure 6. The rmsd, resulting from a global superposition of the ensemble of NMR structures (Smith et al., 1993) also correlates well with individual S^2 values. A similar result has been found for other proteins (e.g. Redfield et al., 1992; Peng et al., 1992; Grasberger et al., 1993). In the case of lysozyme the correlation is comparable to that found for B-factors. No significant improvement in the correlation with S^2 is seen when local rmsd values, derived from the superposition of structures over a window of 2-5 residues, were used. Correlations between smoothed and regionally averaged NMR data are also compared in table 3.

Given that these different indicators of molecular dynamics correlate well, it is possible to examine the factors that determine the extent of dynamics in the different regions of the protein. A molecular dynamics study of hen lysozyme has suggested a possible relationship between increased fluctuations and a greater than average surface accessibility of regions of the protein (Post et al., 1989). In contrast to B-factors, surface accessibilities of peptide bond units are nearly identical when calculated for single protein molecules taken from different crystal structures having the same space-group, and very similar when calculated from structures in other space groups and the ensemble of NMR structures. Therefore, the effect of crystal contacts is, at least in part, removed when individual molecules are considered in this manner. Comparisons of individual S^2 values with peptide bond accessibilities (Figure 6c, Table 3) indicate there is a significant correlation between them. Furthermore, the correlation of S^2 with accessibilities are considerably enhanced when the data are averaged over structural segments of the protein (Figure 6a,b, Table 3,4). Interestingly, comparisons of S^2 values with average accessibilities derived for the ensemble of NMR structures are even better than those with crystal structure accessibilities, especially when smoothed over several residues or averaged over the structural segments of the protein. These results imply that some of the determinants of the dynamic behaviour may often be similar for neighbouring residues.

The extent of mainchain fluctuations at particular sites is also likely to be affected by the participation of the latter in hydrogen bonding and in secondary structure. This has been found in some studies (Schneider et al., 1992; Kördel et al., 1992) but not in others (Kay et al., 1989; Stone et al., 1993). Figure 5d suggests that there is significant correspondence between regions of the mainchain with strong hydrogen bonding and high order parameters. From Figure 7a, it is apparent that peptide groups which are hydrogen bonded in crystal structures have on average higher order parameters than those which are not. However, these populations overlap considerably, as there are sites which, according to their S^2 values, are highly restrained despite the absence of hydrogen bonding (energy > -0.5 $kcalmol^{-1}$). Thus, while for example 14 of the 31 peptide units which are not persistently hydrogen bonded (as either donor or acceptor) in the crystal structures have values of $S^2 \leq 0.83$, 12 have values of $S^2 > 0.86$. This suggests that structural features other than hydrogen bonding significantly determine mobility of mainchain sites.

It is apparent that hydrogen bonding strength is, at least in part, related to surface accessibility. Thus the most strongly hydrogen bonded groups are

completely inaccessible in hen lysozyme. However, this relationship is complicated because inaccessibility does not always result in strong hydrogen bonding (McDonald & Thornton, 1994) and in these cases the lack of mainchain mobility must be rationalised solely in terms of the van der Waals packing interactions within the interior of the protein. The effects of hydrogen bonding and accessibility on the dynamics are therefore largely complementary for NH groups considered here. Comparing the distribution of S^2 values for amides with and without protection against hydrogen exchange (Class II-IV in Pedersen et al., 1991), Figure 7b, it is clear that the latter class have on average lower values of S^2. As both these factors contribute to hydrogen exchange (Pedersen et al., 1991) it might be expected that there would be a good quantitative correlation between S^2 values and amide exchange protection factors. This is not, however, found experimentally presumably because fluctuations which give rise to hydrogen exchange from protected amides represent rare excursions and are not reflected on the timescale of the dynamic processes examined in this study by relaxation measurements.

In addition to the general correlations discussed above, some more specific features of the protein structure do, however, appear to be of considerable importance in determining mainchain dynamics. As an example, we consider the observation that only the central region, residues 68-72, of the long loop spanning residues 61-78 in hen lysozyme experiences substantial fluctuations. Several features of the protein primary sequence in this region are pertinent in this respect. The region contains two proline residues (70 and 79) and several bulky hydrophobic residues (Trp62, Trp63, Leu75, Ile78), and the sidechains of several residues may form salt-bridges (Asp66, Arg61, Arg68 and Arg73) or hydrogen bonds (Asn65, Asn73 and Asn77). These factors are all likely to contribute to the attenuation of the mobility of this loop. The presence of cysteine residues participating in disulphide cross-links (Cys64, Cys76 and Cys80) is, however, particularly likely to be of importance for the reduction of mobility of this and other regions of the protein (shown in Figure 5). By contrast to these features which provide conformational stability to the loop terminal residues, glycine residues 67 and 71 are likely to be responsible for the relatively high flexibility in the central region of the loop. Interestingly, a second mobile loop in the protein, residues 101-104 is also bounded by glycine residues 102 and 104. Other mainchain regions with glycine repeats are, however, not as flexible. An example is the region of the mainchain joining the A- and B-helices (residues 16-23) which does not experience substantial motions on the fast time scales considered here, although Gly22 and Gly26 are in close proximity, and despite the fact that the mainchain in this region has

higher than average surface accessibility and lower than average hydrogen bonding interactions (Figure 7b,c). Three of the loop residues, however, have sidechains which participate in the hydrophobic core of the protein. Thus, as in the case of disulphides, above, it is possible that specific sidechain interactions may dampen and restrict mainchain dynamics and may therefore well be superimposed on the correlation of mobility with accessibility. Similar structural themes have been described as the determinants of loop and turn conformation in antibodies (Tramontano et al., 1989) and are consistent with overall design strategies to build in rigidity to *de novo* protein structures (Chemeris et al., 1994)

Correlation of sidechain S^2 values with features of the crystal and NMR structures. S^2 values of sidechains have been compared to structural parameters similar to those considered in the analysis of the mainchain values; see Table 2(i-iii). Again, B-factors of sidechains correlate well with their order parameters derived in this study. The majority of sidechains with low B-factors have, as expected, high values for S^2 and many of the sidechains with low S^2 values have high B-factors in all crystal structures. Indeed, electron density has often been reported as poor or non-existent in crystallographic maps for several of these sidechains; this is the case for example for Arg73, Arg125 and Gln121 in tetragonal crystal structures (Grace, 1980; Handoll, 1985). There are, however, sidechains with low S^2 values but with low B-factors in all structures showing exceptions to this trend. As with mainchain atoms, B-factors for some of the sidechains can vary considerably between different crystal structures, presumably as a result of crystal packing contacts specific to a certain structure or space group. The sidechains of Gln41, Asn106 and Asn113, which deviate from the approximately linear relationship shown in Figure 8a between S^2 and B-factors for the asparagine and glutamine sidechains in the hentyp2 structure, are part of the group of sidechain atoms known to make intermolecular contacts in tetragonal crystal structures (Handoll, 1985). Overall, however, the results indicate that both B-factors and S^2 values reflect similar motions.

The correlation of S^2 with accessibility found for the mainchain is even more evident for the sidechains, where the variation in both of these parameters is greater. Thus an excellent correlation is obtained when the S^2 values of asparagine and glutamine sidechains are compared with their accessibilities in crystal structures, plotted for hentyp2 in Figure 8b. The location of the sidechains of Asn27 and Asn77 in the hentyp2 crystal structure are shown in Figure 9; these represent examples of low and high sidechain accessibility

respectively. Sidechain group accessibilities averaged over individual members of the ensemble of the NMR solution structures also correlate well with S^2 values, although not as well as the crystallographic accessibilities. This may be due to the fact that few NOE restraints were included for these sidechain atoms in the structure calculations and the restraints on their position are therefore indirect. Low values of S^2 for the arginine sidechains and for Trp62 are, on average, also accompanied by both high B-factors and high accessibilities in crystal structures, Table 2(i & ii). Furthermore, S^2 values which could be obtained for 4 of the remaining 6 unassigned arginine sidechains are also less than 0.3; all arginines have high accessibilities in the crystal structures.

Many of the sidechains considered here are found to be hydrogen bonded in one or more of the 6 crystal structures. Four tryptophan indole hydrogens are on average involved only in weak hydrogen bonding, but these residues are confined to the hydrophobic core of the protein which is likely to be responsible for their high S^2 values. The indole hydrogens of Trp62 and Trp63 are not hydrogen bonded in any of the crystal structures. Two of the five assigned arginine residues are on average hydrogen bonded in the crystal structures (bond energy $\leq$ -0.5 $kJmol^{-1}$). The hydrogen bond partners are however different in different individual structures and the low order parameters suggest that hydrogen bonding is unlikely to be persistent in solution. The asparagine and glutamine sidechains can be divided roughly into two groups by the hydrogen bonding energy associated with them in crystal structures. When hydrogen bond energies are averaged over the 6 crystal structures, eight of these sidechains are hydrogen bonded; they have a mean order parameter of 0.69 $\pm$ 0.10, well above that for sidechains which are on average not hydrogen bonded in crystal structures (mean S^2 0.40 $\pm$ 0.15). S^2 is plotted as a function of hydrogen bond energy in Figure 8c and shows a similar overall trend to that observed for the mainchain atoms. Sidechains with high order parameters are therefore more likely to be involved in hydrogen bonding, as judged from the frequency and strength of their hydrogen bonding in different crystal structures, and are invariably buried from the protein surface.

Evidence for sidechain mobility for hen lysozyme in solution has previously come from $^3J_{\alpha\beta}$ coupling constants, measured directly (Smith et al., 1991) or from the extent of averaging of the χ_1 angle identified from peak-shapes (Bartik & Redfield, 1993). Thus, the extent of conformational averaging about the Cα-Cβ bond is known for the great majority of the sidechains considered here and is summarised together with the order parameters in table 2(i-iii). The 6

asparagine sidechains with a highly averaged χ_1 value in solution have a mean S^2 value of 0.40 ± 0.12. By contrast the 8 asparagine sidechains found to exist in a single conformation about χ_1 have a mean S^2 value of 0.67 ± 0.09. Therefore there is a correspondence between lack of rotational averaging about the Cα-Cβ bond and the degree of restraint of the ^{15}N-^{1}H bond vector as measured by ^{15}N relaxation. This is despite the fact that there is a further Cβ-Cγ bond (torsion angle χ_2) about which the asparagine sidechain end groups may rotate. For the longer sidechains of glutamines, this correspondence may break down as seen in Gln41 which has a S^2 value of 0.19. For arginines, which have even longer sidechains, there appears to be no correlation with the extent of averaging about the Cα-Cβ bond. Interestingly, however, for some of the groups of sidechains there is a significant correspondence between mainchain and sidechain mobility (shown in Figure 1a,b). This may be anticipated if the features which are the determinants of conformational freedom are general and local to particular regions of the protein.

Determinants of conformational freedom. Close packing of the majority of the mainchain and of a significant fraction of sidechains in the interior of proteins is a feature specific to their native folds (Chothia, 1975, 1976; Miller et al., 1987; Richards, 1992; Lee & Richards, 1971, 1994). Equally, it has been observed that the great majority of mainchain and sidechain groups with hydrogen bonding potential are involved in such interactions within the interior as well as near and at the surface of folded proteins (Baker & Hubbard, 1984; Jeffrey & Saenger, 1991; McDonald & Thornton, 1994). Thus burial of the main and sidechains and the optimisation of their hydrogen bonding are recognised as the main determinants of protein structure and folding (Rose & Wolfenden, 1993). It follows that these features of protein structure must also significantly affect, if not determine, the dynamic behaviour of the polypeptide mainchain and the sidechains of its component residues. Evidence for such relationships has been provided from previous analysis of the crystal and solution structures of hen lysozyme and other proteins. Examples are the correlation of the apparent thermal motions (a derivative of the B-factor) with the square of distance from the molecular centroid (Sternberg et al., 1979; Artymiuk et al., 1979) and the correlations found by MacArthur & Thornton (1993) between surface accessibility and a lack of sidechain rotamer definition in NMR structures.

In this study we find statistically significant correlations of S^2 values with B-factors of crystal structures and rmsd values of the ensemble of NMR structures. This suggests that these indicators of positional disorder have at least some common determinants. The data derived here from ^{15}N relaxation

measurements on the dynamics of the mainchain and the collection of 28 sidechains of hen lysozyme in solution have allowed us to explore further some of the possible origins of these determinants. A significant correspondence is found between limited mobility, as defined by high S^2 values, and the strength of hydrogen bonding, although the origin of this relationship is complicated by the known dependence of hydrogen bond strength on low solvent accessibility. However, the most striking correlations both for the mainchain (when regionally averaged) and particularly for sidechains is between S^2 values and surface accessibilities. The quality of these correlations implies that van der Waals contacts or proximity to neighbouring sidechain and mainchain groups, here crudely estimated by static solvent accessibility, impose the most effective constraints on mobility.

It will be interesting as further data become available to assess how similar the details of correlations between accessibility and structure are for different proteins, and to what extent differences between them can be attributed to specific features of the structures. Thus, for example the long loop of lysozyme (residues 61-78), whilst mobile compared to other regions of this protein, experiences more limited fluctuations than loops of some other proteins such as the four helix bundle interleukin-4 (Redfield et al, 1992). This may be attributed to the specific nature of the residues in this region of the sequence and to the presence of disulphide bonds. Furthermore, although the location of binding sites of proteins for other molecules will of necessity be on or near to surfaces of proteins, and hence the residues located in them will frequently be highly mobile in the unbound state, this may not always be the case. In the case of lysozyme, for example, a number of side-chain residues located in the active site are actually more restricted in the extent of their motions than other residues of the same amino acid type located elsewhere in the molecule (Figure 1). This can be attributed to the fact that this region of the protein surface forms a deep cleft which gives rise to closer packing and reduced solvent accessibility for at least some of the residues within it. The extent to which such variation in behaviour is of biological significance, or simply reflects the overall architecture of the protein, remains to be established.

References

Artymiuk, P.J., Blake, C.C.F., Grace, D.E.P., Oatley, S.J., Phillips, D.C. & Sternberg, M.J.E. (1979) Nature 280, 563-568.

Baker, E.N. & Hubbard, R.E. (1984) Prog.Biophys.Mol.Biol. 44, 97-179.

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

Bartik, K. & Redfield, R. (1993) J.Biol.NMR 3, 415-428.

Barbato, G., Ikura, M., Kay, L.E., Pastor, R.W., Bax, A. (1992) Biochem. 31, 5269-5278.

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.

Boyd, J., Hommel, U. & Campbell, I.D. (1990) Chem.Phys.Lett. 175, 477-482.

Bruccoleri, R.E., Karplus, M. and McCammon, J.A. (1986) Biopolymers, 25, 1767-1802.

Brünger, A.T. (1992) "X-PLOR version 3.1: A system for X-ray crystallography and NMR", Yale University Press.

Campbell, I.D., Dobson, C.M. & Williams, R.J.P. (1975) Proc.R.Soc.Lond.B. 189, 503-509.

Cantor, C.R. & Schimmel, P.R. (1980) "Biophysical Chemistry", Vol II; Freeman & Co, San Francisco.

Chandrasekhar, I., Clore, M.G., Gronenborn, A.M. & Brooks, R. (1992) J.Mol.Biol. 226, 239-250.

Chemeris, V.V., Dolgikh, D.A., Fedrov, A.N., Finkelstein, A.V., Kirpichnikov, M.P., Uversky, V.N. & Ptitsyn, O.B. (1994) Protein Eng. 7, 1041-1052.

Chothia, C. (1975) Nature 254, 304-308.

Chothia, C. (1976) J.Mol.Biol. 105, 1-14.

Clarage, J.B., Clarage, M.S., Phillips, W.C., Sweet, R.M. & Caspar, D.L.D. (1992) Proteins: Struct.Funct. & Genetics 12, 145-157.

Clore, G.M., Driscoll, P.C., Wingfield, P.T. & Gronenborn, A.M. (1990) Biochem. 29, 7387-7401.

Constantine, K.L., Friedrichs, M.S., Goldfarb, V., Jeffrey, P.D., Sherriff, S. & Mueller, L. (1993) Proteins: Struct.Funct. & Genetics 15, 290-311.

Cross, T.A., & Opella, S.J. (1983) J.Am.Chem.Soc. 105, 306-308.

Dobson, C.M. & Karplus, M. (1986) Meth.Enzym. 131, 362-389.

Dobson, C.M. (1993) Curr.Biol. 3, 530-532.

Faure, P., Micu, A., Perahia,D., Doucet, J., Smith, J.C. & Benoit, J.P. (1994) Nature: Struct.Biol. 1, 124-128.

Gerstein, M., Lesk, A.M & Chothia, C. (1994) Biochem. 33, 6739-6749.

Gibrat, J-F & Go, N. (1990) Proteins: Struct.Funct. & Genetics 8, 258-279.

Grace, D.E.P. (1980) D.Phil. Thesis, University of Oxford.

Grasberger, B.L., Gronenborn, A.M. & Clore, G.M. (1993) J.Mol.Biol. 230, 364-372.

Handoll, H.H.G. (1985) D.Phil. Thesis, University of Oxford.

Herzfeld, J., Roberts, J.E. & Griffin, R.G. (1987) J.Chem.Phys. 86, 597-602.

Hoel, P.G (1984) "Introduction to Math. Statistics", 5th Ed., J. Wiley & Sons, N.Y.

Imoto, T., Johnson, L.N., North, A.C.T., Phillips, D.C., & Rupley, J.A. (1972) "The Enzymes" (Boyer, P.D., Ed.) Vol VII, Academic Press, N.Y.

Jeffrey, G.A. & Saenger, W. (1991) "Hydrogen Bonding in Biological Structures", Springer, N.Y.

Jeenes, D.J., Marczinke, B., MacKenzie, D.A. & Archer, D.B. (1993) FEMS Microbiol.Lett. 107, 267-272.

Kabsch, W. & Sander, C. (1983) Biopolymers 22, 2577-2637.

Karplus, M. & McCammon, J.A. (1983) Ann.Rev.Biochem. 53, 263-300.

Kay, L.E., Torchia, D.A. & Bax, A. (1989) Biochem. 28, 8972-8979.

Kay, L.E., Nicholson, L.K., Delagio, F., Bax, A. & Torchia, D.A. (1992) J.Magn.Reson. 97, 359-375.

Kodandapani, R., Suresh, C. & Vijayan, M. (1990) J.Biol.Chem. 27, 16126-16131.

Kördel, J., Skelton, N.J., Akke, M., Palmer, A.G. & Chazin, W.J. (1992) Biochem. 31, 4856-4866.

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

Krishna, N.R., Sarathy, K.P., Huang, D-H., Stephens, R.L., Glickson, J.D., Smith, C.K. & Walter, R. (1982) J.Am.Chem.Soc. 104, 5051-5053.

Kundrot, C.E. & Richards, F.M. (1987) J.Mol.Biol. 193, 157-170.

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

Lipari, G. & Szabo, A. (1982) J.Am.Chem.Soc. 104, 4559-4570.

London, R.E. (1989) Meth.Enzym. 176, 358-375.

MacArthur, M.W. & Thornton, J.M. (1993) Proteins: Struct.Funct. & Genetics 17, 232-251.

Marion, D., Ikura, M. & Bax, A. (1989) J.Mag.Reson. 84, 425-430.

McDonald, I.K. & Thornton, J.M. (1994) J.Mol.Biol. 238, 777-793.

Messerle, B.A., Wider, G., Otting, G., Weber, C. & Wüthrich, K. (1989) J.Mag.Reson. 85, 608-613.

Miller, S., Janin, J., Lesk, A. & Chothia, C. (1987) J.Mol.Biol. 196, 641-656.

Ohkubo, T., Taniyama, Y. & Kikuchi, M. (1991) J.Biochem. 110, 1022-1029.

Olejniczak, E.T., Poulsen, F.M., & Dobson, C.M. (1984) J.Am.Chem.Soc. 106, 1923-1930.

Palmer, A.G., Rance, M. & Wright, P.E. (1991) J.AmChem.Soc. 113, 4370-4380.

Palmer, A.G., Skelton, N.J., Chazin, W.J., Wright, P.E. & Rance, M. (1992) Mol.Phys. 75, 699-711.

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, J.W. & Wagner, G. (1992) Biochem. 31, 8571-8586.

Post, C.B., Dobson, C.M. & Karplus, M. (1989) Proteins: Struct.Funct. & Genetics 5, 337-354.

Powers, R., Clore, G.M., Garret, D.S. & Gronenborn, A.M. (1993) J.Mag.Reson. B 101, 325-327.

Ramanadham, M., Sieker, L.C. & Jensen, L.H. (1990) Acta Cryst. B46, 63-69.

Radford, S.E., Dobson, C.M. & Evans, P.A. (1992a) Nature 358, 302-305.

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

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

Redfield, C. & Dobson, C.M. (1990) Biochem. 29, 7201-7214.

Redfield, C., Boyd, J., Smith, L.J., Smith, R.A.G & Dobson, C.M. (1992) Biochem. 31, 10431-10437.

Richards, F.M. (1974) J.Mol.Biol. 82, 1-11.

Richards, F.M. (1992) in "Protein Folding", T.E. Creighton (Ed.), Freeman, N.Y.

Roberts, I.N., MacKenzie, D.A., Jeenes, D.B., Archer, D.B., Radford, S.E., Robinson, C.V., Aplin, R.T. & Dobson, C.M. (1992) Biotech.Lett. 14, 897-902.

Rose, G.D. & Wolfenden, R. (1993) Ann.Rev.Biophys.Biomol.Struct., 22, 381-415.

Schneider, D.M., Dellwo, M.J. & Wand, J.A. (1992) Biochem. 31, 3645-3652.

Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C. & Davis, D.R. (1987) Proc.Natl.Acad.Sci (USA) 84, 8075-8079.

Smith, L.J., Sutcliffe, M.J., Redfield, C. & Dobson, C.M. (1991) Biochem. 30, 986-996.

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

Smith, J.L., Hendrickson, W.A., Honzatko, R.B. & Sheriff, S. (1986) Biochem. 25, 5018-5027.

Sternberg, M.J.E., Grace, D.E.P. & Phillips, D.C. (1979) J.Mol.Biol. 130, 231-253.

Strynadka, N.C.J. & James, M.N.G. (1991) J.Mol.Biol. 220, 401-424.

Stone, M.J., Fairbrother, W.J., Palmer, A.G., Reizer, J., Saier, M.H. & Wright, P.E. (1992) Biochem. 31, 4394-4406.

Thomson, N.K. & Poulsen, F.M. (1993) J.Mol.Biol. 234, 234-411.

Tramontano, A., Chothia, C., & Lesk, A.M. (1989) Proteins: Struct.Funct. & Genetics 6, 382-394.

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

Wilson, K.P., Malcolm, B.A. & Matthews, B.W. (1993) J.Biol.Chem. 267, 10842-10849.

Young, A.C.M., Tilton, R.F. & Dewan, J.C (1994) J.Mol.Biol. 235, 302-317.

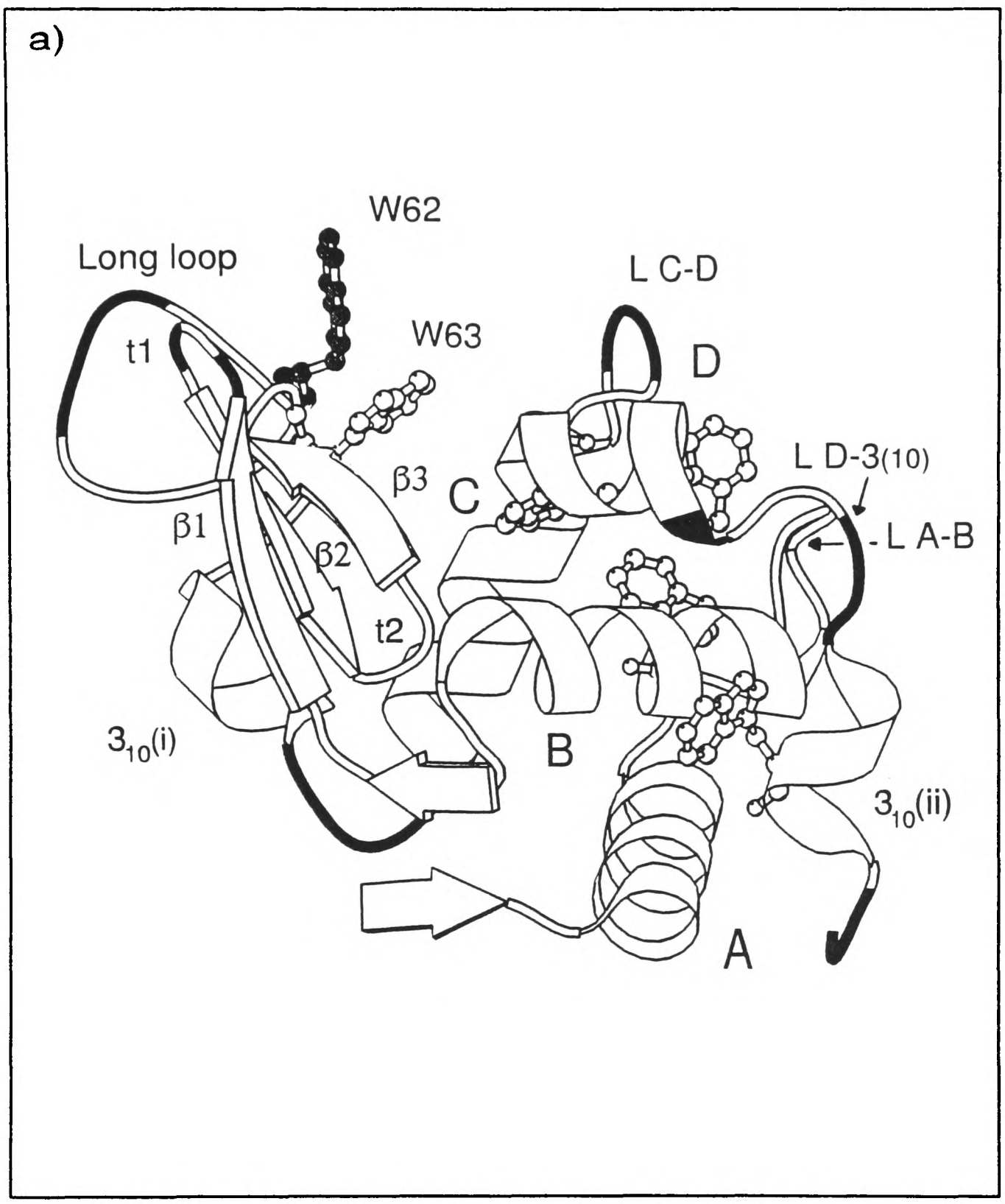


Figure 1a

Schematic outline of the structure of hen lysozyme drawn using Molscript (Kraulis, 1991). In the helical sub-domain the loops are labelled L.

The four α-helices, and the N- and C- termini are labelled. The regions of the mainchain and the tryptophan sidechain (Trp62) with order parameters S^2 less than or equal to 0.80 are shaded.

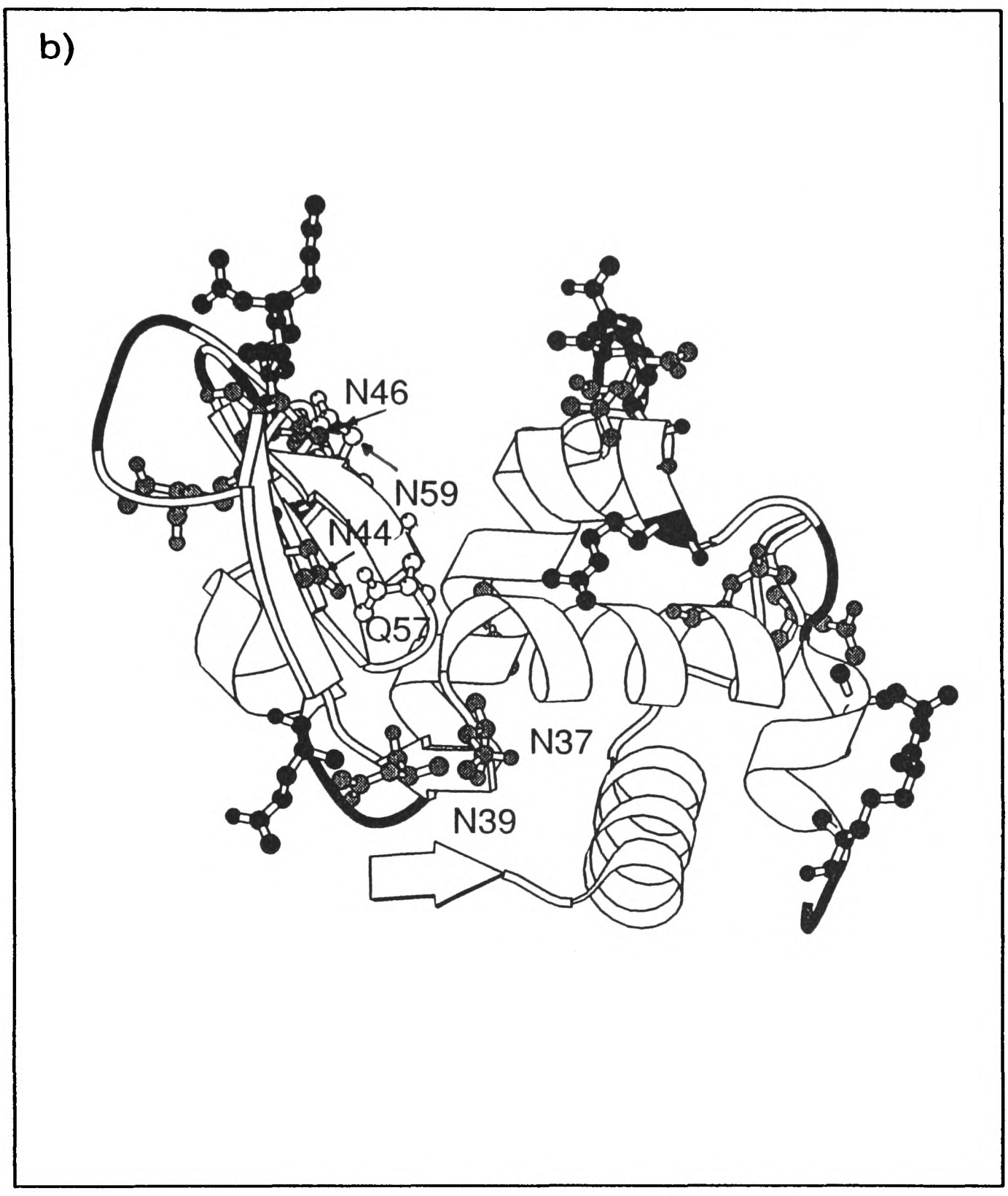


Figure 1b

Schematic outline of the structure of hen lysozyme drawn using Molscript (Kraulis, 1991). The sidechains of arginine, asparagine and glutamine are shaded according to the magnitude of the S^2 for their terminal NH or NH_2 groups. Black, $S^2 < 0.4$; grey, $0.4 < S^2 < 0.7$; white, $S^2 > 0.7$. S^2 values for the 7 asparagine/ glutamine sidechains located in the active site cleft range from 0.5-0.8 (mean S^2 value 0.65 ± 0.13); S^2 values for the 10 asparagine/ glutamine sidechains located elsewhere range from 0.2-0.7 (mean S^2 value 0.46 ± 0.19). The active site is located in the cleft between the α-helical and β-sheet domains

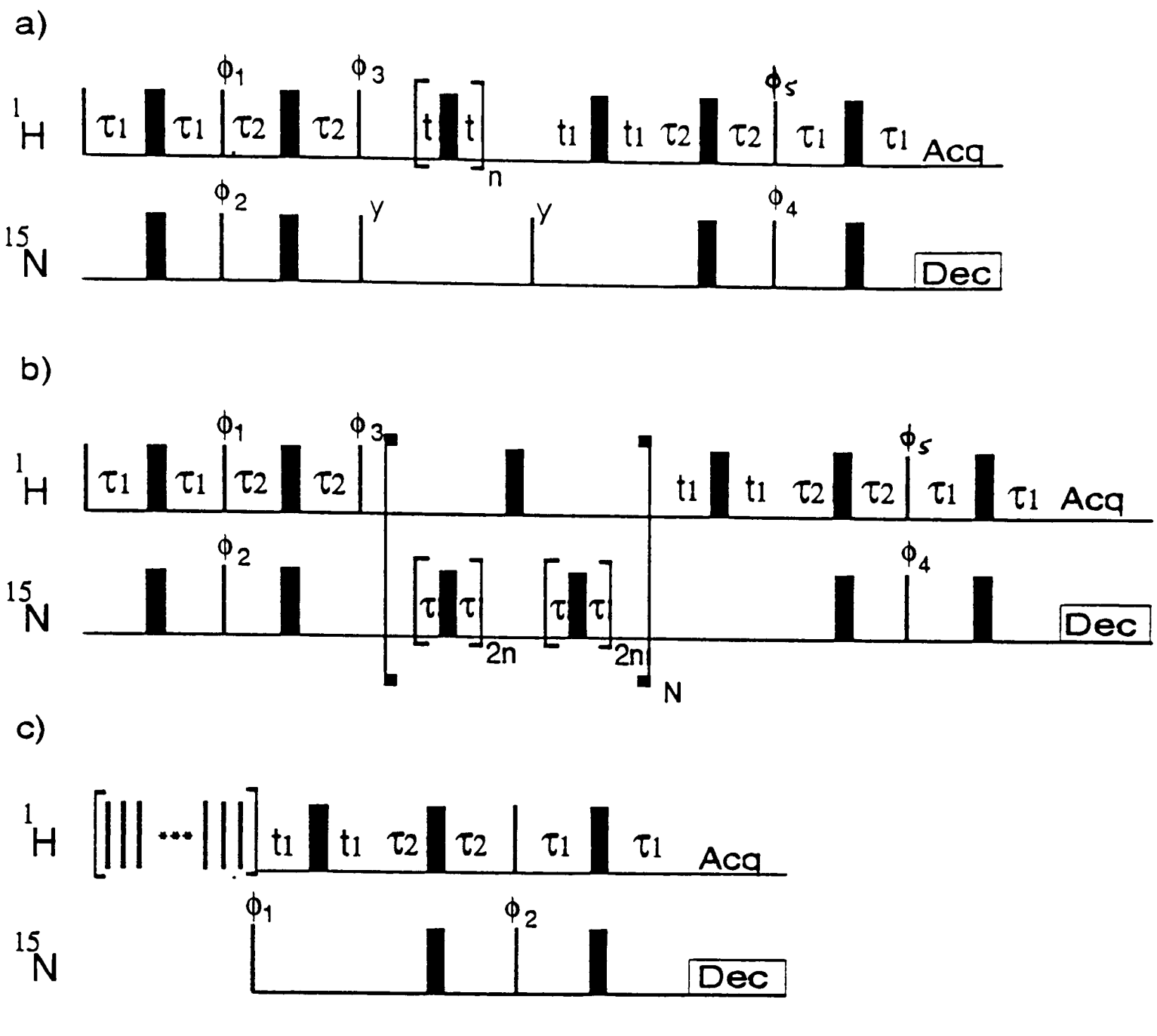


Figure 2

Pulse schemes for the measurement of a) T_1 b) T_2 and c) ^{15}N-^1H NOE relaxation parameters of ^{15}N nuclei in asparagine and glutamine side chains. All π pulses have phase +y. The phase cycling used for sequences a) and b) was φ1=8(y),8(-y); φ2=4(x),4(-x); φ3=2(x),2(-x), φ4=16(y),16(-y) and the acquisition was 2(x), 4(x), 2(x), 2(x), 4(x), 2(x). The receiver phase was further inverted every 16 scans. Sign discrimination in F_1 was achieved by incrementing the phase of φ4. The ^1H π pulses applied during the relaxation period of sequences a) and b) serve to remove the effects of ^{15}N-^1H CSA-dipolar cross-correlations. (Palmer et al., 1992; Kay et al., 1992). Solvent suppression, where used, was achieved with a combination of RF pulse presaturation and trim pulses (Messerle et al., 1989). For the ^{15}N-^1H NOE experiment no RF presaturation was used. φ2 was incremented for F_1 sign discrimination and the phase cycling closely followed that described in Kay et al., (1989). The delays τ_1 and τ_2 were set to 2.45ms and 1.38ms respectively for all experiments. τ in sequence b) was typically set to 300μs.

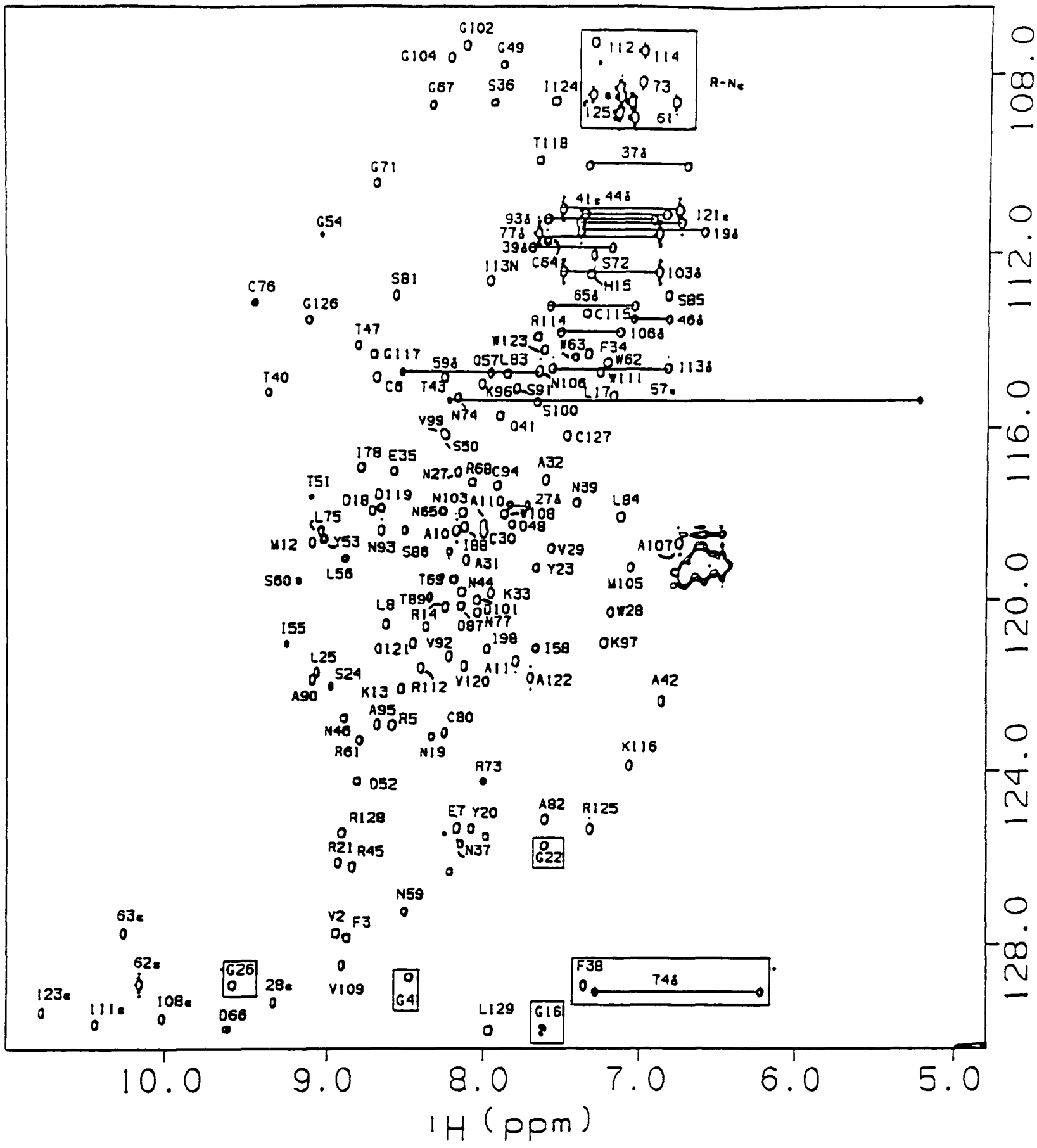


Figure 3

500MHz ^{15}N-^{1}H HSQC spectrum of uniformly ^{15}N labelled hen lysozyme in H_2O at 35°C and pH 3.8. All the expected 126 mainchain ^{1}H-^{15}N correlations are labelled, together with the 6 tryptophan $^{15}N^{\varepsilon}$-^{1}H and pairs of the resonances of the 17 asparagine/ glutamine $^{15}N^{\delta/\varepsilon}$-$^{1}H_2$ sidechain groups. The 5 arginine $^{15}N^{\varepsilon}$-^{1}H resonances which could be assigned are labelled. Cross peaks shown in boxes are contoured at negative levels as they are folded in from outside the region of the spectrum displayed. Unlabelled peaks are low level impurities. Cross peaks C30/A110, S50/V99 and G16/impurity overlap significantly and were not used for subsequent analysis of ^{15}N relaxation.

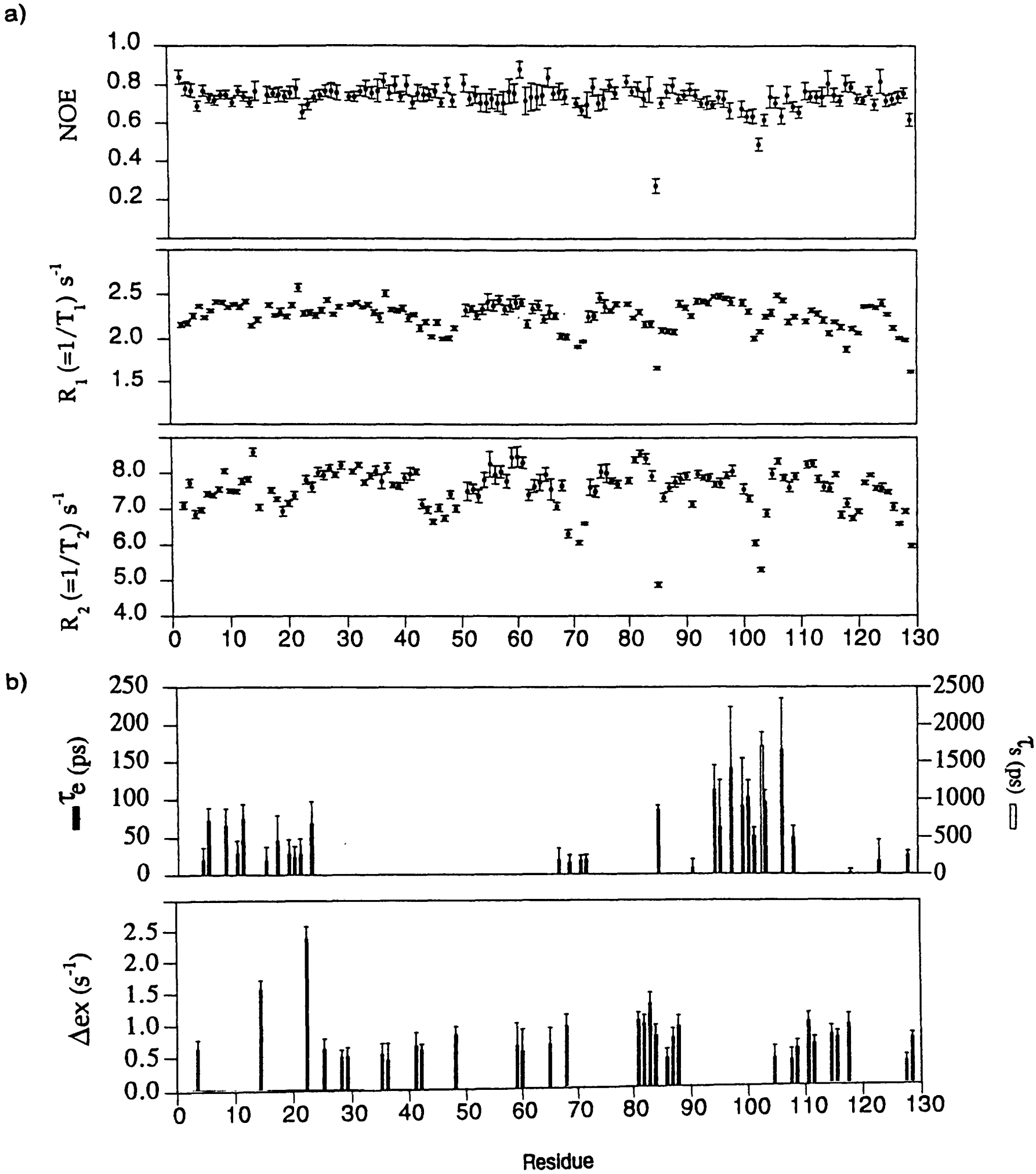


Figure 4 (legend see next page)

Figure 4

a) NOE; R_1 (=$1/T_1$) and R_2 (=$1/T_2$) values for mainchain amide nitrogens and b) τ_e (black), τ_s (white) and Δex values are each plotted versus lysozyme sequence. Values of -160 ppm for the CSA constant (Stone et al., 1992; Schneider et al., 1992) and 5.7 ± 0.2ns for the rotational correlation time (τ_R) were used in the calculations. All order parameters in this study were calculated with the assumption of fixed and equal NH bonds lengths of 1.02Å throughout the protein. A lengthening of the bond by 1% would cause an increase of approximately 6% in the value of S^2; such vibrational lengthening may itself be part of internal motion (Kay et al., 1989; Schneider et al., 1992). The R_2 value for residue 22 in plot a) is $10.8s^{-1}$.

Figure 5

a) Main chain S^2 values and their uncertainties, calculated as described in the text, as a function of sequence. The units of protein secondary structure are shown and the positions of the cysteine residues, involved in disulphide bonds, the glycine and the proline residues are indicated as (■), (G) and (P) respectively. b) B-factors of mainchain nitrogen atoms in the hentyp2 crystal structure (dashed) and as the average of the 6 hen lysozyme crystal structures, which are listed in the Methods section. c) Mean peptide unit accessibilities of the 6 crystal structures (continuous line) and the members of the ensemble of NMR structures (Smith et al., 1993) (dashed). d) Mean of hydrogen bond energies for the peptide units in the 6 crystal structures, plotted in the exponential form. The data in c) and d) were smoothed (unweighted) by averaging data points over a sliding window of 5 residues .

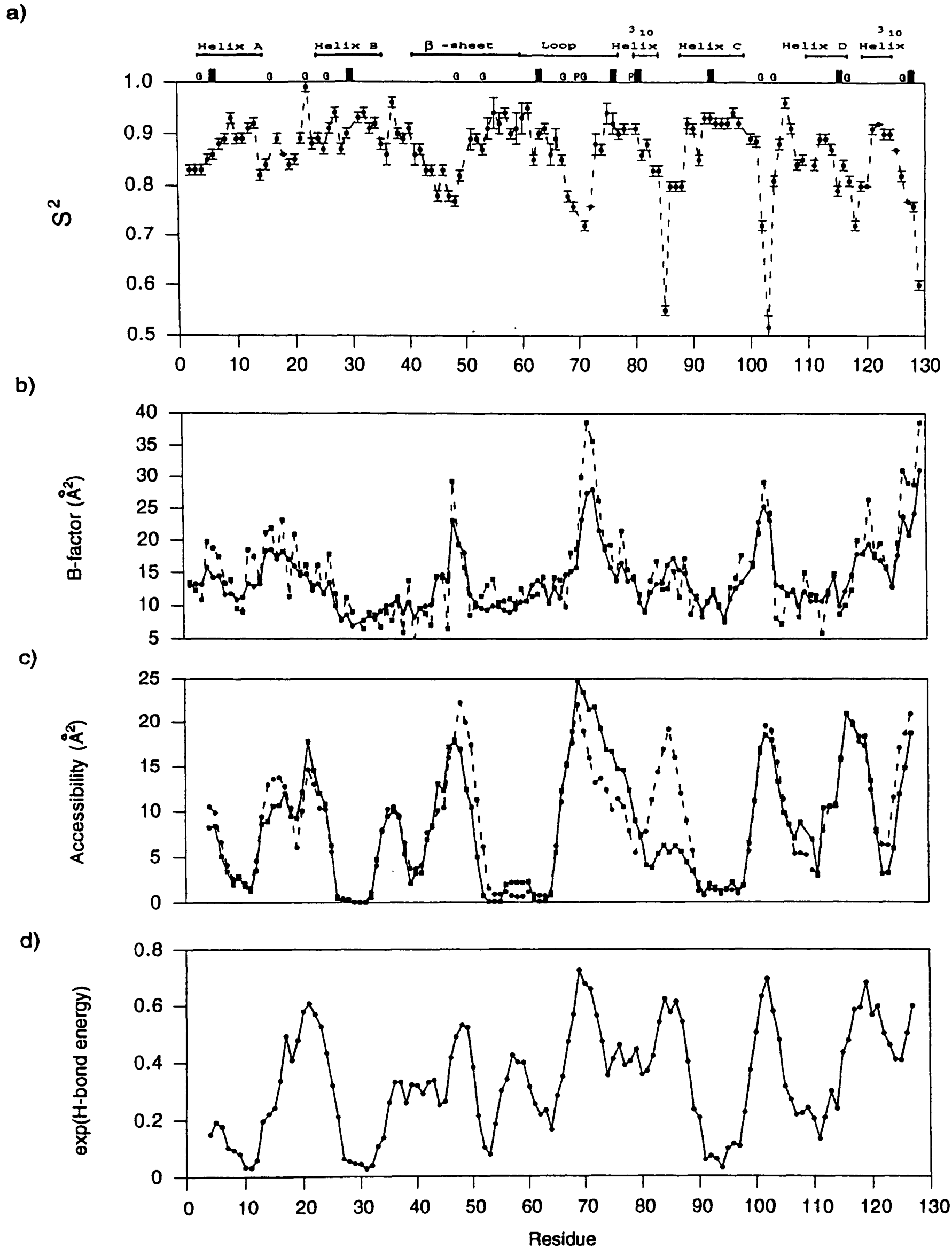


Figure 5 (legend on previous page)

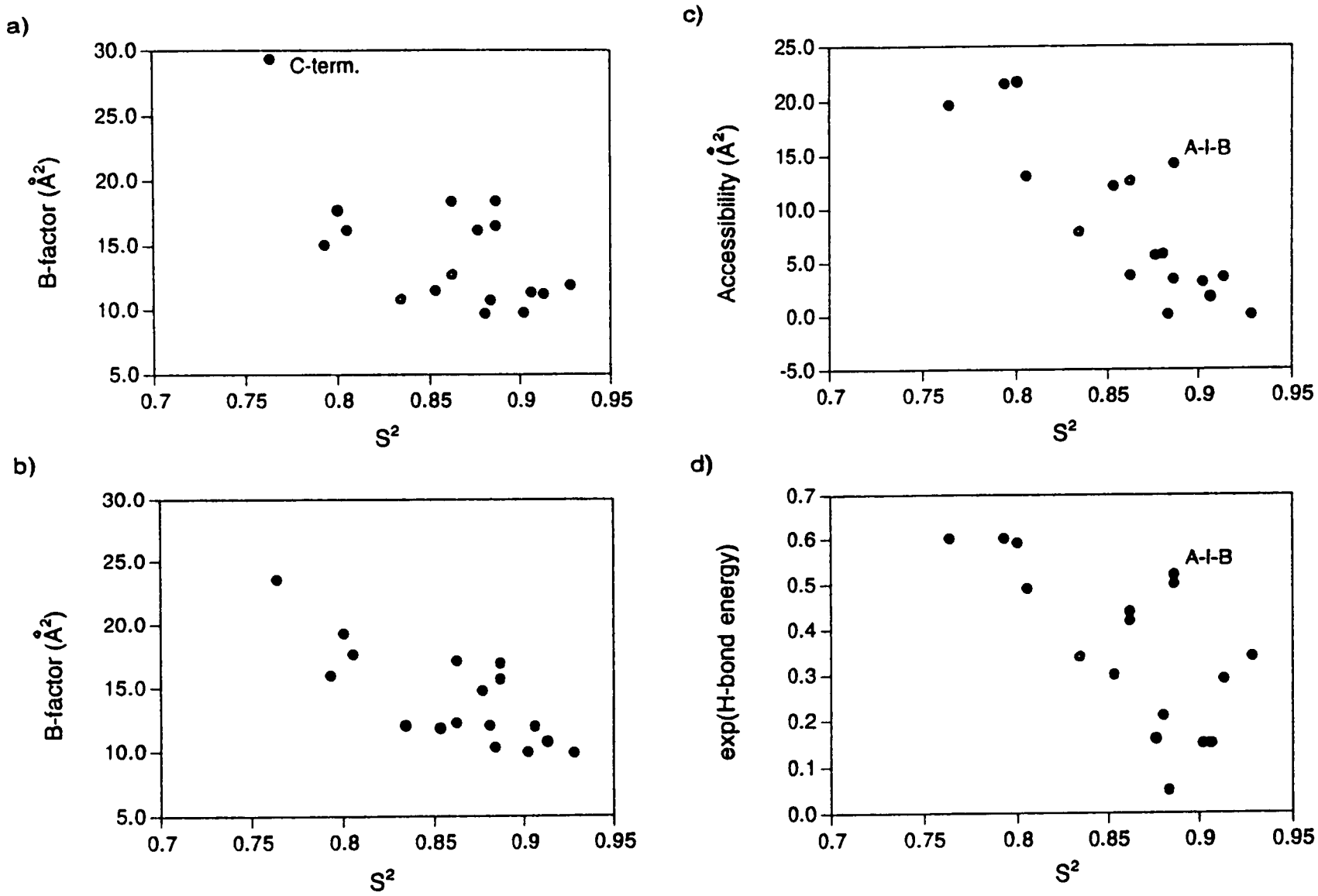


Figure 6
Correlations of S^2 values averaged over structure segments with a) B-factors of mainchain nitrogen atoms in the hentyp2 crystal structure b) mean of B-factors of the 6 crystal structures c) mean peptide unit accessibilities of the 6 crystal structures d) mean hydrogen bond energies of the peptide units in the 6 crystal structures. The correlation coefficients are given in Tables 3 and 4.

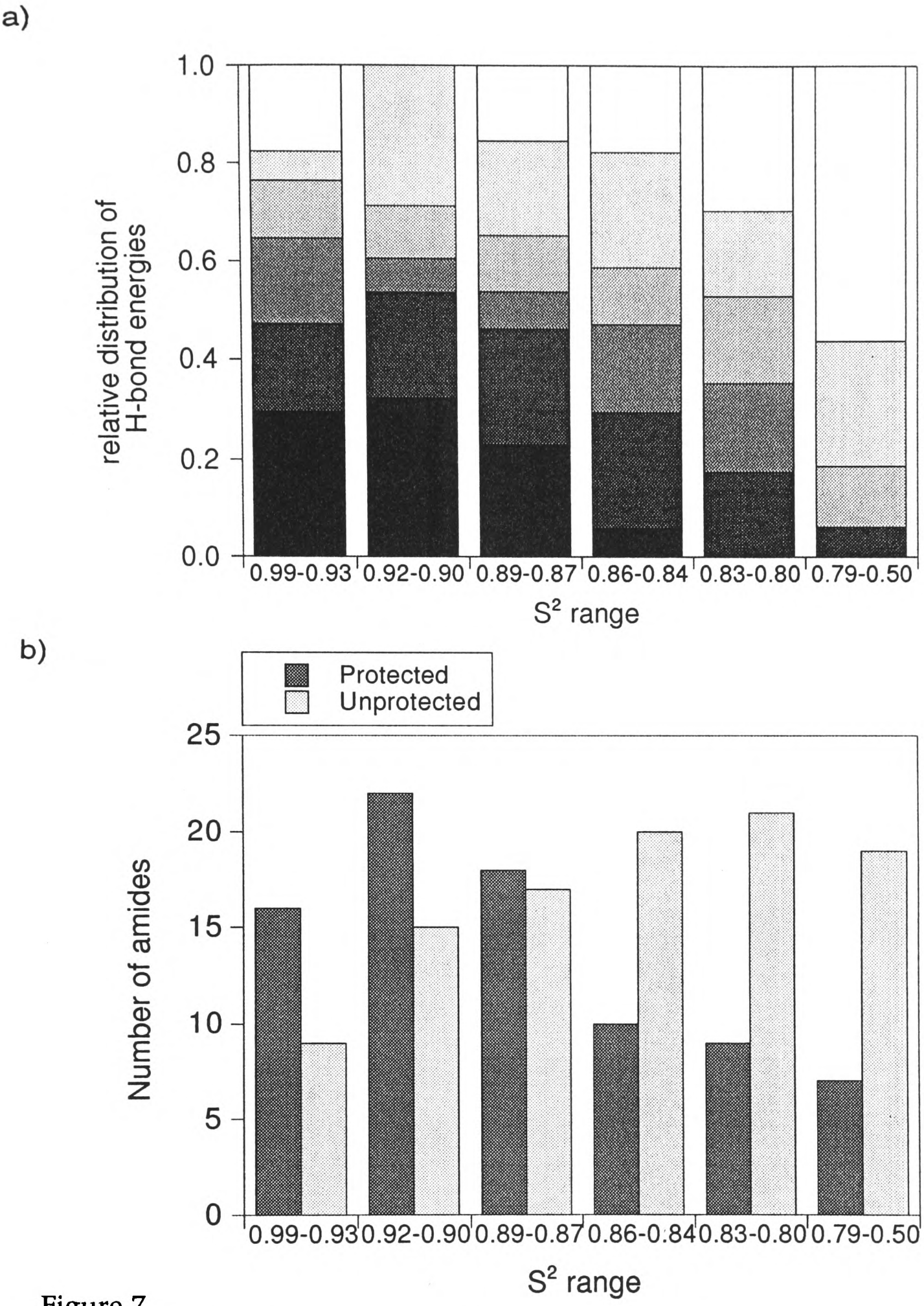


Figure 7

a) Relative distribution of the peptide units in categories of hydrogen bond energy within ranges of S^2 values. Hydrogen bond energy categories are 0.0 to -0.5, -0.5 to -1.0, -1.0 to -1.5, -1.5 to -2.0, -2.0 to -3.0 and less than -3.0kcalmol^{-1} in order of increased shading. b) The number of amides having slowed hydrogen exchange (class II-IV defined as protected in Pedersen et al., 1991) (black) and fast exchange (grey) within ranges of S^2 values.

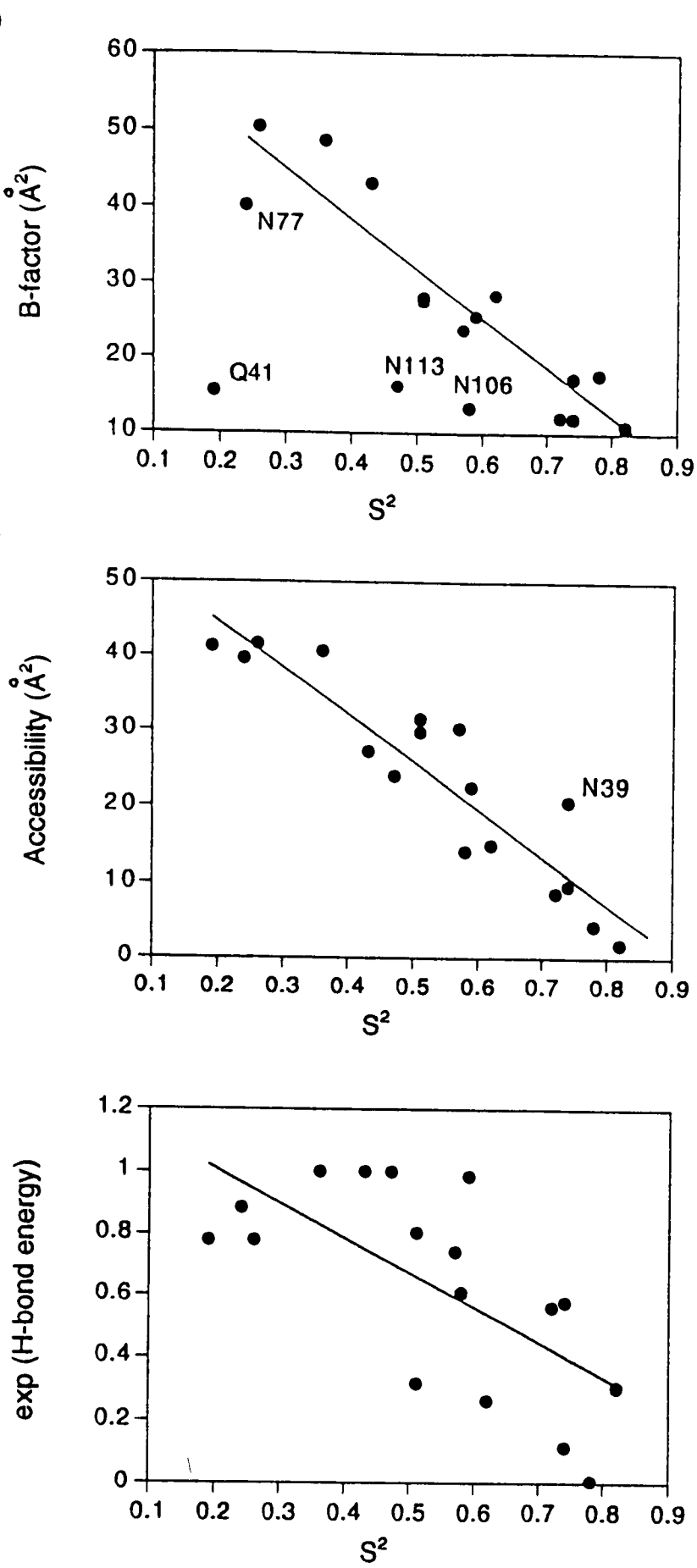


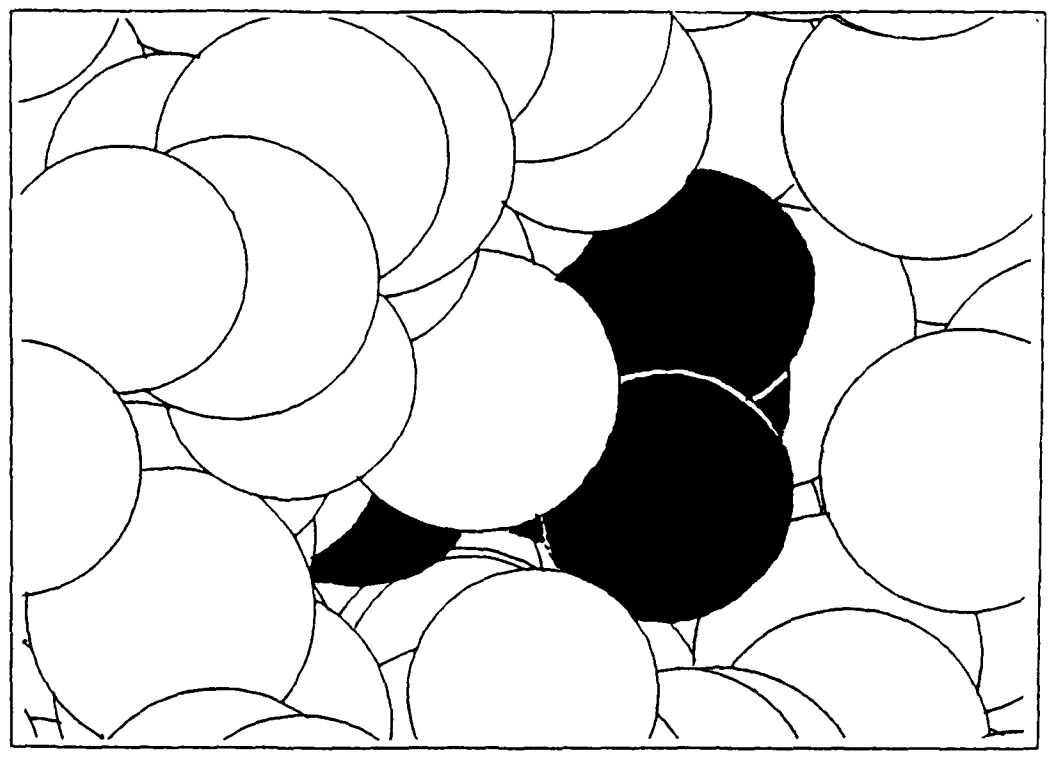
Figure 8

S^2 values for asparagine and glutamine sidechain groups as a function of a) B-factors, b) accessibility, both of the hentyp2 crystal structure, c) hydrogen bond energy plotted in exponential form, averaged over the 6 crystal structures. The correlation coefficients for these relationships are 0.65 for a), 0.92 for b) and 0.67 for c). Direct comparison of results for tryptophan and arginine with those for asparagine and glutamine sidechains is difficult given their fewer numbers, the differences in the absolute accessibility scale used and the likely dependence of the minimum value of S^2 attainable on the number of intervening bonds to the mainchain. The data for tryptophan and arginine sidechains are summarised in Table 1; the trends are in good agreement with those seen for asparagine and glutamine sidechains.

Asn 27

a)

S^2 0.72
Accessibility 8.8Å^2 (20%)



Asn 77

b)

S^2 0.24
Accessibility 39.6Å^2 (88%)

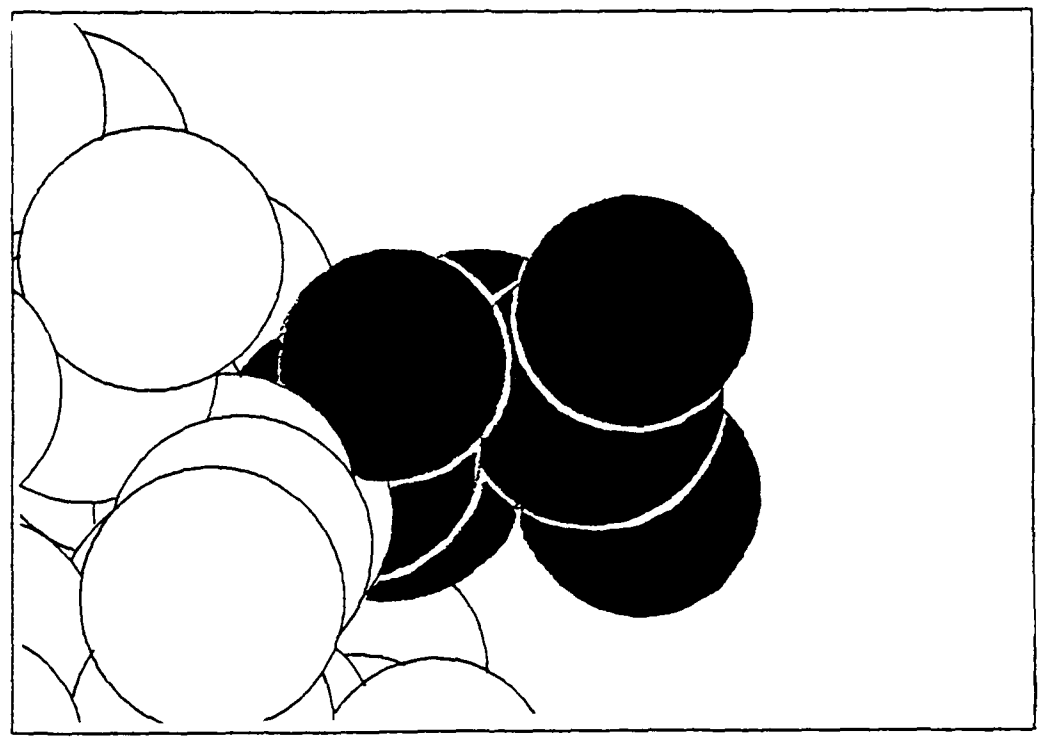


Figure 9
Examples of the environments of asparagine sidechains a) Asn27 with a well defined conformation b) Asn77 with a poorly defined conformation in all the 6 crystal structures. Accessibilities of the hentyp2 structure and S^2 values are indicated. Van der Waals radii were used on the hentyp2 crystal structure to generate this picture using InsightII (Biosym. Inc.)

Table 1

^{15}N mainchain and ^{1}H, ^{15}N sidechain assignments for hen lysozyme.

Residue		Chemical Shifts (ppm) a) Backbone ^{15}N	Sidechain ^{15}N		Sidechain ^{1}H
Lys	1				
Val	2	127.72			
Phe	3	127.80			
Gly	4	104.86			
Arg	5	122.92			
Cys	6	114.84			
Glu	7	125.30			
Leu	8	120.56			
Ala	9	121.56			
Ala	10	118.40			
Ala	11	121.42			
Met	12	118.66			
Lys	13	122.06			
Arg	14	120.14			
His	15	112.48			
Gly	16	106.14			
Leu	17	115.24			
Asp	18	117.92			
Asn	19	123.18	Nε	111.55	6.61/ 7.39
Tyr	20	125.30			
Arg	21	126.10			
Gly	22	101.88			
Tyr	23	119.20			
Ser	24	121.96			
Leu	25	121.68			
Gly	26	105.02			
Asn	27	117.00	Nδ	117.73	7.72/ 7.83
Trp	28	120.26	Nε	129.28	
Val	29	118.78			
Cys	30	118.42			
Ala	31	119.06			
Ala	32	117.18			
Lys	33	119.82			
Phe	34	114.28			
Glu	35	116.98			
Ser	36	108.64			
Asn	37	125.66	Nδ	110.32	6.72/ 7.33
Phe	38	105.04			
Asn	39	117.72	Nδ	111.89	7.19/ 7.70
Thr	40	115.18			
Gln	41	115.70	Nε	111.05	6.77/ 7.50
Ala	42	122.34			
Thr	43	114.84			
Asn	44	119.80	Nδ	111.13	6.85/ 7.37
Arg	45	126.20			
Asn	46	121.68	Nδ	113.52	6.83/ 7.05
Thr	47	114.10			
Asp	48	118.22			
Gly	49	107.80			
Ser	50	116.14			
Thr	51	117.60			
Asp	52	124.22			
Tyr	53	118.58			
Gly	54	111.60			
Ile	55	121.02			
Leu	56	119.04			
Gln	57	114.70	Nε	115.36	5.23/ 8.24
Ile	58	121.10			
Asn	59	127.22	Nδ	114.74	7.69/ 8.54
Ser	60	119.56			
Arg	61	123.26	Nε	84.75	6.79
Trp	62	114.46	Nε	128.90	
Trp	63	114.34	Nε	127.74	
Cys	64	111.72			
Asn	65	117.94	Nδ	113.21	7.04/7.58

Asp	66	129.88			
Gly	67	108.70			
Arg	68	117.24			
Thr	69	119.50			
Pro	70				
Gly	71	110.44			
Ser	72	112.04			
Arg	73	124.20	Nε	84.30	7.00
Asn	74	115.28	Nδ	105.24	6.23/ 7.29
Leu	75	118.40			
Cys	76	113.14			
Asn	77	120.28	Nδ	111.59	6.89/ 7.66
Ile	78	116.90			
Pro	79				
Cys	80	123.08			
Ser	81	112.96			
Ala	82	125.10			
Leu	83	114.74			
Leu	84	118.04			
Ser	85	112.94			
Ser	86	118.36			
Asp	87	120.12			
Ile	88	118.28			
Thr	89	119.92			
Ala	90	121.86			
Ser	91	115.08			
Val	92	120.60			
Asn	93	118.38	Nδ	111.26	6.85/ 7.60
Cys	94	117.34			
Ala	95	122.90			
Lys	96	114.98			
Lys	97	120.98			
Ile	98	121.12			
Val	99	116.14			
Ser	100	115.38			
Asp	101	119.98			
Gly	102	107.36			
Asn	103	117.94	Nδ	112.48	6.89/ 7.50
Gly	104	107.62			
Met	105	119.22			
Asn	106	114.68	Nδ	113.81	7.14/ 7.52
Ala	107	118.66			
Trp	108	117.98	Nε	129.67	
Val	109	128.44			
Ala	110	118.50			
Trp	111	114.70	Nε	129.68	
Arg	112	121.30	Nε	83.42	7.31
Asn	113	112.64	Nδ	114.62	6.83/ 7.57
Arg	114	113.90	Nε	83.61	6.99
Cys	115	113.34			
Lys	116	123.84			
Gly	117	114.32			
Thr	118	109.92			
Asp	119	117.88			
Val	120	120.32			
Gln	121	121.00	Nε	111.36	6.75/ 7.40
Ala	122	121.82			
Trp	123	114.20	Nε	129.53	
Ile	124	108.60			
Arg	125	125.34	Nε	84.59	7.32
Gly	126	113.54			
Cys	127	116.16			
Arg	128	125.40			
Leu	129	129.94			

Table 1 Footnotes

(a) ^{1}H chemical shift were referenced relative to dioxan at 35°C and ^{15}N chemical shifts referenced externally to $^{15}NH_3$. Chemical shifts are accurate to 0.01 and 0.03 p.p.m. respectively.

Table 2
Summary of order parameters and structural data for sidechains[a].

Residue (a)	S^2	+/-	Sf^2, Ss^2	τe [τs] (ps)	+/-	Δex (s^{-1})	+/-	B-factor ($Å^2$) (f)	Accessibility ($Å^2$) (f)	χ_1 (g)	exp(H-bond E) (h)
Trp (b)											
28	0.90	0.01		49.73	10.54			11.8	0.0	no	0.67
62	0.41	0.01	0.73, 0.56	[1317.1]	31.90			42.4	19 6	-	1.00
63	0.88	0.01		46.03	14.82	0.84	0.17	19.4	11.4	-	1.00
108	0.87	0.01		65.59	10.36			15.8	2.8	no	0.63
111	0.88	0.01				0.83	0.16	10.9	0.0	no	0.68
123	0.85	0.01						13.4	10.6	no	0.67
Arg (c)											
61	0.28	0.02		74.96	8.69			40.2	17.7	no	0.53
73	0.12	0.01		101.16	12.63			78.9	36.2	yes	0.93
112	0.31	0.04		72.51	13.24	1.59	1.40	46.4	23.4	yes	0.33
114 (d)	0.27	0.02		34.92	4.16	0.52	0.53	13.2	28.9	no	0.97
125	0.05	0.01		120.53	6.15	2.37	0.97	57.9	19.4	yes	0.89
Asn/Gln (e)											
N19	0.43	0.00		24.74	0.92			43.2	27.2	yes	1.00
N27	0.72	0.08				2.82	1.35	12.2	8.8	no	0.56
N37	0.51	0.04		20.02	1.94			28.3	29.8	yes	0.80
N39 (d)	0.74	0.02				1.29	0.37	12.0	20.7	no	0.11
Q41 (d)	0.19	0.01	0.51, 0.38	[228.2]	[12.7]			15.6	41.2	no	0.78
N44 (d)	0.51	0.01		25.70	1.86			27.8	31.5	yes	0.31
N46	0.62	0.02				1.88	0.48	28.5	15.1	no	0.26
Q57	0.82	0.04						10.9	1.9	no	0.30
N59	0.78	0.13				4.68	2.25	17.8	4.4	no	0.01
N65	0.57	0.01		11.69	2.87			23.8	30.3	no	0.74
N74	0.74	0.03				1.75	0.57	17.4	9.7	no	0.58
N77	0.24	0.01	0.57, 0.42	[389.2]	[36.2]			40.2	39.6	yes	0.88
N93	0.59	0.01		26.79	3.13	0.65	0.21	25.7	22.6	no	0.98
N103	0.26	0.02	0.56, 0.47	[699.5]	[57.8]			50.4	41.6	yes	0.78
N106 (d)	0.58	0.02		21.80	4.10	0.82	0.29	13.4	14.2	no	0.61
N113 (d)	0.47	0.01	0.71, 0.66	[397.5]	[31.8]			16.2	24.0	yes	1.00
Q121	0.36	0.00		55.73	1.05			48.5	40.6	yes	1.00

Table 2 Footnotes

(a) Data at pH3.8, 35°C. Hydrogen exchange rates for the sidechains are estimated to be between $0.25s^{-1}$ and $0.7s^{-1}$ (Krishna et al., 1982; Wedin et al., 1982; Bai et al., 1993) and are too slow to cause significant errors in the relaxation measurements. (b) calculated using a CSA constant of -89 ppm (Opella & Cross, 1983). (c) the CSA constant is not known. A value of -160ppm was used in the calculation. S^2 values vary by ±20% with values of -90 and -210ppm respectively. (d) Sidechain atom groups involved in crystallographic contacts. (e) A CSA constant of -163 ppm was used; this was calibrated for NH_2 in asparagine (Herzfield et al., 1987) and was presumed to be identical for glutamine. The differences in derived parameters, fitted to data for the two components of the 15N-1H correlation separately, are small and are shown as average. (f) hentyp2 crystal structure. (g) A "yes" indicates that the sidechain is found in a single $\chi 1$ rotamer state and a "no" indicates that averaging about $\chi 1$ has been detected by NMR (Smith et al., 1991; Bartik et al., 1993). (h) Hydrogen bond energy as exponential, mean of the 6 crystal structures.

Table 3.1
Correlation coefficients for comparisons made between structural parameters describing crystallographic and NMR structures of hen lysozyme.

(a)	B-factor	rmsd NMR	Accessibility crystal	Accessibility NMR	exp(H-bond E) crystal
B-factor	0.84	0.46	0.59	0.58	0.57
rmsd NMR	0.50	N/A	0.38	0.54	0.45
Accessibility crystal	0.76	0.50	0.96	0.84	0.57
Accessibility NMR	0.74	0.69	0.71	0.89	0.58
exp(H-bond E) crystal	0.71	0.61	0.74	0.76	0.94

Table 3.1 Footnotes
(a) Diagonal: average correlation coefficient for the comparisons of parameters of individual structures with the parameters averaged over the structures. Above diagonal: comparisons of the 6 crystal structures or NMR ensemble averaged structural parameters. Below diagonal: comparison of parameters smoothed over a window of 5 residues.

Table 3.2
Correlation coefficients for comparisons made between structural parameters describing crystallographic and NMR structures and order parameter S^2. (a) (b)

Comparison of S^2 with	individual residues (121)	smoothed window of 5 residues	average of structural units (17)
B-factor crystal (c)	0.53	0.53	0.63
B-fact mean crystals	0.65	0.69	0.77
rmsd NMR	0.53	0.68	0.62
exp(H-bond E)	0.50	0.67	0.69
Accessibility			
single crystal (c)	0.52	0.66	0.85
mean crystal	0.53	0.66	0.85
NMR ensemble	0.62	0.81	0.92

Table 3.2 Footnotes
(a) Statistical significance (at 5% critical value) requires correlation coefficients of greater than 0.14 for 121 data points (residues) and 0.41 for 17 data points (structural units) (Hoel, 1984). (b) The relationships need not be linear, or in the case of hydrogen bond energy, exponential and there can be no proof of a causal link. (c) representative data from hentyp2 crystal structure.

Table 4
S^2 values and structural parameters averaged over 17 different structural units of hen lysozyme.

Structural Units (a)	Residues	S^2	B-factor (b) ($Å^2$)	rmsd NMR (c) ($Å^2$)	Accessibility crystal (b) ($Å^2$)	Accessibility NMR (c) ($Å^2$)	exp(H-bond E) (d)
Helix A	4-15	0.88	16.2	1.3	5.7	6.8	0.16
Helix B	24-36	0.90	9.8	1.1	3.2	3.7	0.15
Helix C	89-99	0.91	11.4	1.0	1.8	1.2	0.15
Helix D	108-115	0.85	11.5	1.7	12.1	9.6	0.30
3_{10} (i)	80-84	0.86	12.7	4.4	3.8	11.3	0.42
3_{10} (ii)	120-124	0.89	18.4	2.2	3.4	6.3	0.50
strand1	41-45	0.83	10.9	1.9	7.9	8.3	0.34
turn1	46-49	0.80	17.7	5.4	21.8	21.8	0.59
strand2	50-53	0.88	10.7	2.4	0.1	1.4	0.05
turn2	54-57	0.93	11.9	1.1	0.1	0.9	0.34
strand3	58-60	0.91	11.2	1.6	3.6	0.9	0.29
loopA-B	16-23	0.89	16.5	1.3	14.2	11.1	0.52
loopC-D	100-107	0.81	16.2	1.5	13.1	13.7	0.49
Long loop (e)	61-78	0.86	18.4	2.3	12.6	10.8	0.44
loopD-3_{10} (ii)	116-119	0.79	15.1	4.1	21.6	21.0	0.60
C-terminus	125-129	0.76	29.3	1.5	19.7	20.9	0.60
small sheet	1-3/37-40	0.88	9.7	2.9	5.8	7.7	0.21

Table 4 Footnotes
(a) defined in Handoll (1985). (b) hentyp2 crystal structure, (c) ensemble of structures, (d) mean of the 6 crystal structures. (e) The average hydrogen bond energy and peptide unit accessibility over the central region of the long loop (residues 68-72) are -0.4kcal mol^{-1} and 23.4$Å^2$ respectively. The parameters for the terminal residues of the loop (residues 61-67 + 73-78) are -1.0kcal mol^{-1} and 9.8$Å^2$. The average residue hydrogen bond energy and the accessibility of peptide units for the entire protein are -1.1kcal mol^{-1} and accessibility 8.4$Å^2$ respectively.

Chapter 7

A Heteronuclear NMR Study of a Partially Ordered State of a Protein: Structural Preferences of Hen Egg-White Lysozyme in Trifluoroethanol

Summary

2D and 3D heteronuclear NMR have been used to characterise the structure of hen egg white lysozyme in a partially ordered state using uniformly ^{15}N labelled protein. This state is formed by the denaturation of the protein in 70% trifluoroethanol (TFE)/water (v/v) and is characterised by substantial α-helical secondary structure in the absence of extensive tertiary interactions. ^{15}N-filtered 3D NOESY and TOCSY experiments have allowed the sequential assignment of resonances for all but 2 of the 126 mainchain amide nitrogen atoms and of the majority of main and sidechain proton resonances. The conformation of the polypeptide chain was characterised by analysis of the pattern of NOEs , H_α chemical shift perturbations, $^3J(H^N,H_\alpha)$-coupling constants and hydrogen exchange. α-helical structures are found to be located in parts of the polypeptide which are helical in the native state, and are accompanied by significant protection of amide hydrogens against exchange with solvent. By contrast to these persistent and native-like structures, non-native helical structures of increased flexibility are observed in this state of the protein as extensions to the native-like helices and in the region which forms in the native state the C-terminal part of the β-sheet. No specific structural preferences are detected in regions corresponding to the long loop and N-terminal part of the β-sheet of native lysozyme, confirming that TFE is unable to stabilise ordered structures in these regions. A combination of local features of the polypeptide chain, including the predicted propensities of residues for helix formation and for their participation in N- and C-terminal helix capping interactions can be used to rationalise the conformational behaviour of the polypeptide chain of hen lysozyme in this partially ordered state.

Introduction

It is generally accepted that the refolding of small monomeric proteins occurs along pathways which are defined by partially folded states (Kim & Baldwin, 1990; Creighton, 1990; Matthews, 1993). The nature and the dynamics of structures in these states, intermediate between the natively folded and fully unfolded conformations of the polypeptide chain, are of critical importance for the elucidation of the mechanisms by which proteins fold (Barrick & Baldwin, 1993; Fersht, 1993; Dobson, 1991, 1994). Detailed studies of such partially ordered states are likely to provide an understanding of the competing interactions which determine the stability of structures as well as the manner of their interconversion during the folding process.

In practice, the characterisation of intermediates sampled during kinetic refolding is often complicated by their transient nature, the availability of structural probes, the possible conformational heterogeneity of intermediates and parallel nature of some folding pathways. Despite these difficulties, the folding pathway of hen egg white lysozyme has been analysed in some detail using a wide variety of different spectroscopic techniques (Radford et al., 1992b; Dobson et al., 1994). Equilibrium studies are made difficult by the high degree of cooperativity of the folding process of many proteins under denaturing conditions. Studies of protein fragments (Dyson & Wright, 1991; Dill & Shortle, 1991) and of peptide models (Staley & Kim, 1990; Peng & Kim, 1994), provide a complementary strategy, allowing the propensity of a given amino acid sequence to form local structures to be delineated (Dyson et al., 1992; Kemminck & Creighton, 1993; Chakrabartty et al., 1994).

Another strategy used to overcome the difficulties arising from the cooperativity of folding of intact proteins under equilibrium conditions employs denaturants, such as urea and GuHCl, or extremes of pH or temperature. While in some studies low concentrations of GuHCl or low pH have been used to promote local unfolding of the most labile structures (Mayo & Baldwin, 1993; Redfield et al., 1994), other studies have used denaturants at high concentrations in order to probe the most persistent structures of proteins (Neri et al., 1992; Logan et al. 1994). It is clear, that depending on the protein and the solution conditions, structures in these partially ordered states, including the species often referred to as "molten globule states", may range from very local conformational preferences, in otherwise unfolded states to structures in states which are substantially ordered and involve larger regions of the polypeptide chain (Dill & Shortle, 1991; Ptitsyn, 1992). In the latter case

the ordered regions of a partially folded protein may encompass significant fractions of structural domains which comprise the conformation of the native protein (Dobson, 1994).

NMR has become the method of choice for the study of non-native proteins at high resolution and several techniques have been applied for this purpose. Protection of amide hydrogens against exchange with solvent, for example, is afforded by structures which are persistently ordered (Woodward et al., 1982; Englander & Kallenbach, 1984) and can be detected also by indirect methods, such as pH or denaturant concentration pulse quench techniques (Baum et al., 1989; Roder, 1989; Chyan et al., 1993; Woodward, 1994; and see Chapters 2 to 4), which make use of favourable characteristics of the native state. The NMR spectrum of native proteins has also been important for direct studies of denatured states under equilibrium conditions by magnetisation transfer experiments (Jeener et al., 1979). Correlations of resonances between native and non-native states are possible if the two states can be brought into exchange with kinetics of typically greater than 1 s^{-1} and this has allowed the identification of resolved resonances in denatured states of a number of proteins (Dobson et al., 1984; Dobson & Evans, 1984; Baum et al., 1989; Evans et al., 1991; Harding et al., 1991). Nevertheless, both these and other methods provide only limited information about denatured states, principally because overlap of resonances in homonuclear NMR spectra of denatured states severely complicates their complete assignment and the extraction of NMR parameters for all but small peptide systems of typically fewer than 40 residues in length.

Recently methods have been developed, exploiting the greater sensitivity and resolution offered by indirect detection involving ^{15}N and ^{13}C nuclei, which overcome these problems, either in conjunction with heteronuclear magnetisation transfer techniques (Wider et al., 1991; Zhang et al., 1994) or by NOE or through bond correlation techniques (Kay et al., 1990; Stockman et al., 1993; Logan et al., 1993). Site specific characterisation of the conformational and dynamic behaviour in non-native states of proteins may now, in favourable circumstances, be possible to the extent common for native proteins and small peptides (Wüthrich, 1994). Although, several such heteronuclear studies have been reported on denatured and partially folded proteins it is apparent that these involve proteins either in states which are substantially structured, possessing at least part of an ordered core (van Mierlo et al., 1993; Feng et al., 1994; Redfield et al., 1994) or states which are largely disordered, containing few local clusters of structure (Neri et al., 1992a,b; Logan et al., 1994; ; Alexandrescu

et al., 1994c; Arcus et al., 1994). Proteins in states with characteristics between these two extremes, such as classical "molten globule" type states, have been difficult to analyse by homonuclear methods (Baum et al., 1989; Alexandrescu et al., 1993), because structures in them experience conformational averaging on a timescale of milliseconds, leading to extensive broadening of resonances. Thus it is of considerable interest to examine the recently developed heteronuclear NMR methods in their application to a states of proteins with a wide range of structural and dynamic behaviour.

Recently, we have described a model for a partially ordered state of hen lysozyme by denaturing the protein in trifluoroethanol (TFE) cosolvent. This partially ordered state differs substantially from both the native and highly denatured states of the protein (Chapter 2 to 6). Far uv CD spectroscopy suggests the α-helical structural content exceeds that of the native state. However, direct as well as indirect hydrogen exchange measurements demonstrate that amides located in the most stable structures in the TFE state are also protected in the native state and are situated in regions corresponding to regions which are α-helical in the native state. Extensive and persistent tertiary contacts are absent in this denatured state (Chapter 4 and 5). Nevertheless, low angle X-ray scattering data implies that the protein is at least partially compact in the presence of the native disulphides which persist in this denatured state (Goto - personal communication). In a number of these respects the TFE denatured state, therefore, resembles a "molten globule" type state, although in others it does not.

The TFE denatured state of hen lysozyme has also been studied as peptides of various lengths (Segawa et al., 1991; Bolin, 1994 & in preparation; Yang et al., 1994 & in preparation). These and the previous studies suggested that some of the additional helix content, which is induced by the cosolvent, may be located in part of the region of the polypeptide chain corresponding to the native β-sheet of hen lysozyme. An important question in the study of denatured states at equilibrium studies is, therefore, whether direct or indirect interactions of proteins with denaturants promote native-like structures, or whether such interactions can alter and induce novel structures.

2,2,2-Trifluoroethanol (TFE) has been employed in many studies of the structural propensities of peptides and protein fragments in order to stabilise their conformations (e.g. Segawa et al., 1991; Dyson et al., 1992). Thus, there is considerable incentive to understand the effect of TFE on protein structure. Although at present the thermodynamic and structural aspects of TFE-protein

interactions are incompletely understood, a large number of studies suggest that TFE stabilises α-helical structures preferentially in regions of the polypeptide chain which are helical in the native state (e.g. Nelson & Kallenbach, 1989; Lehrman et al., 1992). However, TFE also induces α-helical structures in regions of polypeptides which are part of β-structures in the native folds (Fan et al., 1993; Lui et al., 1994). In several instances, however, these polypeptides have been shown to have considerable intrinsic propensity to adopt α-helical structures (Sönnichsen et al., 1992; Goto, Y. - personal communication). In some instances TFE has also been reported to stabilise β-turn and hairpin structures (Sönnichsen et al., 1992; Shin et al., 1993; Blanco et al., 1994).

The native fold of hen lysozyme has been characterised in great detail by both X-ray and NMR spectroscopy (Handoll, 1985; Wilson et al., 1992; Smith et al., 1993) and contains both α-helical and β-sheet secondary structural elements (Figure 1a). Comparisons of the conformational preferences detected in the partially ordered state with the structures in the native state are likely to be of critical importance for the understanding of the interactions which give rise to their formation. Similarly, the conformations can be put into the context of the structural features found in peptide model systems. For example, the role of interactions which stabilise α-helices have received particular attention in this respect (Scholtz & Baldwin, 1992) and it should be possible, at least in part, to rationalise α-helix formation in peptides as well as partially folded states.

In this paper we applied the established sequential assignment method (Wüthrich, 1986) to ^{15}N filtered 3D spectra of TFE denatured hen lysozyme. The conformational preferences of hen lysozyme in this partially ordered state are assessed by a number of complementary NMR parameters, including perturbation of chemical shifts from random coil values, coupling constants, the pattern of NOEs and protection against hydrogen exchange. The dynamic behaviour of the polypeptide chain in this TFE denatured state is assessed in detail in Chapter 8 by analysis of the NMR relaxation behaviour of the backbone ^{15}N nuclei.

Materials and Methods

Collection of NMR data

Hen egg white lysozyme was expressed in *Aspergillus Niger* using $^{15}NH_4Cl$ as the sole nitrogen source and purified from filtered culture medium (Roberts et al., 1992). NMR samples were prepared to contain ca. 3 mM protein in 70% TFE-d_2-OH/30% H_2O (v/v) and the pH was adjusted to pH 2.0 with phosphoric acid buffer. NMR experiments were performed at 37°C and 47°C on NMR spectrometers belonging to the Oxford Centre for Molecular Sciences with ^1H operating frequencies of 500.2, 600.2, and 750.2 MHz. The ^1H and ^{15}N and assignments were made using data recorded from two sets (37°C and 47°C) of 3D NOESY-HSQC, HSQC-NOESY-HSQC and HSQC-TOCSY experiments (Driscoll et al., 1990; Frenkiel et al., 1990; Ikura et al., 1990; respectively) with mixing times of 200 ms in the NOESY- and 62 ms in the TOCSY-experiments, and a recycle delay of typically 1.3 s. The data set comprised 128, 32, 512 complex points in t_1, t_2, t_3 for the NOESY-HSQC and TOCSY-HSQC experiments recorded with a digital resolution of 38.3, 25.9, 13.2 Hz/complex point respectively at 600 MHz. Forward linear prediction as implemented in Felix2.3 (Hare Research & Biosym Inc.) was used to extend the data in the ^{15}N dimension from 32 to 48 complex points and all 3D data sets were zero filled to yield a final matrix size of 256, 128, 2048 real data points.

$^3J(H^N,H_\alpha)$-coupling constants were measured using HMQCJ and HNHA experiments and were determined from lineshape simulations and crosspeak intensities (Kay & Bax, 1989; Vuister & Bax, 1994 respectively). The HMQCJ experiment was performed on a 600 MHz spectrometer. The data set comprised 432, 1024 complex points in t_1 (^{15}N), t_2 (N^H) with a resolution of 1.9 and 5.5 Hz/complex point. Squared sine-bell apodisation (72°) was used in t_1 and the data matrix was zero filled to 1024, 1024 real data points. Coupling constants were fitted to doublets in ω_1, after summing over the cross peak in ω_2 and four times zerofilling, by optimisation of the coupling constant and linewidths for each of the doublet components using home-written software. The HNHA experiment were performed on the 500 MHz spectrometer. The data set comprised 48, 30, 1024 complex points in t_1 (^{15}N), t_2 (H_α), t_3 (N^H) with a resolution of 28.8, 117.2 and 5.5 Hz/complex point. The constant time delay to defocus the $^3J(H^N,H_\alpha)$-coupling constant was set to 26.1 ms. The row data was mirror image linear predicted in the ^{15}N-dimension to 96 complex points, the final matrix size was zero filled to 128, 128, 512 real data points. The ratios of cross peak to diagonal peak intensities were determined using the volume integration routine in Felix2.3.

Direct measurement of amide hydrogen exchange

Lyophilised enzyme was dissolved in 70% TFE-d_3/D_2O (v/v) to a final concentration of 3 mM and the pH rapidly adjusted to 2.0. A series of twelve fast 2D ^{15}N-^{1}H HSQC spectra, each taking 23 min, were recorded at 37°C. Single exponential decays for each residue were fitted to amide occupancy, measured by peak heights over the time course of spectra. The resulting rates (k_{ex}) were compared to intrinsic exchange rates (k_{int}) predicted for amides in an unstructured polypeptide chain, in order to account near neighbour inductive effects on acid and water catalysis (Bai et al., 1993). The effect of TFE on the intrinsic exchange rates was found to be small (within a factor of 3) for model peptides at pH 2 (Chapter 4; Bolin, 1994). and protection factors (k_{int}/k_{ex}) uncorrected for TFE cosolvent effects are reported.

Results

Resonance Assignment

Figure 2 shows the ^{15}N-^{1}H HSQC spectrum at 750 MHz ^{1}H frequency for hen lysozyme denatured in 70% (v/v) TFE cosolvent. This spectrum is typical of those obtained for denatured proteins. As a result of the disruption of fixed tertiary interactions, the ^{1}H chemical shift dispersion characteristic of native folds, is dramatically reduced. Thus, the mainchain amide protons resonate over a range of 2.6 p.p.m. in the native state of hen lysozyme, whereas this dispersion is reduced to around 1 p.p.m. in the TFE denatured state at 37°C. The same observation holds true for the sidechain protons, for which only small deviations from random coil chemical shift values (Wüthrich, 1986) are observed. ^{1}H resonance overlap therefore severely complicates the resonance assignment of lysozyme in TFE in homonuclear spectra. However, the crosspeaks are well dispersed through their correlation with the attached ^{15}N nuclei. Thus, it is apparent that in excess of 85% of the cross peaks are fully resolved in the ^{15}N-^{1}H HSQC spectrum of this denatured state recorded at 750 and also at 600 MHz ^{1}H frequency.

In addition to good resolution of the ^{15}N-^{1}H crosspeaks, the NMR spectra of the TFE denatured state of hen lysozyme were found to contain a large number of intense interproton NOEs. This made possible the application of the sequential assignment approach based on the identification of spin systems, and their connection through sequential NOEs using the amino acid sequence information (Wüthrich, 1986). TOCSY transfer was found to be poor for residues in α-helices because $^{3}J(N^{H},H_{\alpha})$ coupling constants are small.

Furthermore, in the case of the TFE state resonances arising from residues in α-helices were often observed to have larger than average linewidths. Nevertheless, on a preliminary level, around 30% of resonances could be assigned to spin system types in TOCSY spectra of the TFE state, serving as starting points for further assignment (Figure 3a). Resonances could be assigned to certain amino acid types by the proximity of their ^{15}N and particularly sidechain ^{1}H chemical shifts to those expected in a random coil (Wishart et al., 1992a,b and Wüthrich, 1986 respectively). Some of the ^{15}N resonances show particularly clearly such amino acid specific clustering into regions of the HSQC spectrum. Thus, for example, all 12 glycine resonances are found in the region between 106 and 112 p.p.m..

Figure 3b shows ω_1, ω_3-"strips" at the respective ^{15}N-frequency in ω_2 from the 3D NOESY-HSQC spectrum showing intra- and inter-residue NOEs to the amide protons in the stretch of residues Tyr53 to Ile58. Analysis of the NOE data confirms that the great majority of NOEs from the amide protons throughout the protein are to protons belonging to residues within a distance of +3/-2 residues in sequence, as would be expected for a denatured state which lacks extensive tertiary interactions. Thus, the observed NOEs involve protons belonging to residues close in sequence. The relative intensity of NOEs between an amide and protons of the residue itself or to those of the preceding residue is generally higher than to others, suggesting possible assignments of the most intense NOEs should first be sought involving protons within a residue and to immediate residue neighbours. Furthermore, the 3D HSQC-NOESY-HSQC experiment often provided unambiguous links between adjacent amides (Figure 3c).

From these experiments, it was possible to assign sequentially the 63 N-terminal residues and the 12 C-terminal residues, with the exception of residues Phe34 and Val120, in the set of 3D spectra of the TFE denatured state recorded at 37°C. Assignment of resonances for residues in the C-terminal half of the molecule was complicated by the reduced number of NOEs observed in the 3D NOESY experiments and their apparently weaker intensity. These resonances correspond to ^{15}N-^{1}H correlations which have increased linewidths in the HSQC spectrum (Figure 2). In order to attempt to enhance the rates of the processes which give rise to the line-broadening, such as conformational exchange between several substates with rates approaching the intermediate NMR timescale, and to attempt to resolve the overlap of certain resonances, the NMR data were also acquired at 47°C, under otherwise identical experimental conditions. Indeed, under these conditions many of the weak

^{15}N-^{1}H correlations became more intense and a few additional resonances can be detected. Overlap of several ^{15}N-^{1}H correlations was resolved, as the chemical shift of the amide proton in particular is highly sensitive to temperature for a number of resonances. By contrast the chemical shifts of aliphatic sidechain protons were found to be nearly identical at the two temperatures. It was possible to assign resonances for all but two of the expected 126 mainchain amides in the spectra acquired at 47°C. The ^{15}N-^{1}H correlations for Lys97 and Lys116 remained unassigned, probably because their linewidths are still too large for their detection. However, NOEs from neighbouring amide protons to the Hα and to some of the sidechain protons of these two residues were observed. The assignments made at the higher temperature aided the assignment of spectra at the lower temperature for all but a few resonances of residues 65, 75-78, 90 and 120 which were presumably too weak to be seen. It is interesting to note that all of the residues whose resonances could not be assigned, except for the two C-terminal residues, are located in close sequence proximity to some of the cysteines involved in the four disulphide cross-bridges which remain intact in this denatured state. This suggests that the disulphide bridges may be responsible for generating dynamic events on the millisecond time scale leading to the resonance broadening (e.g. see Redfield et al. 1992; and Chapter 8). Figure 4 shows the summary of the short and medium range NOEs, grouped into strong, medium and weak. To take into account the variation of cross-peak intensity found in the ^{15}N-^{1}H correlation, NOE intensities to a given amide proton were scaled to the intensity of the autocorrelation peak in the NOESY spectra.

In addition to the mainchain resonances the ^{15}N-^{1}H HSQC spectrum (Figure 2) contains also ^{15}N-^{1}H correlations arising from sidechain nitrogens of the 11 arginine, the 6 tryptophan and 14 asparagine and 3 glutamine residues of the protein at pH 2.0. No appreciable NOEs were detected in the 3D NOESY spectrum involving sidechain protons bonded to the arginine $^{15}N^{\varepsilon}$ nuclei. However, several NOEs involving tryptophan indole-ring protons and asparagine/glutamine sidechain protons are seen in the 3D NOESY-HSQC spectrum. Assignment of sidechain NOEs to the sidechain ^{15}N-bound atoms was possible for all 6 tryptophan indole and 13 of the 17 asparagine/glutamine NH_2 groups. In order to assign the remaining protons of the tryptophan rings, not all of which were detected in ^{15}N edited spectra, high resolution 2D ^{1}H-^{1}H NOESY and TOCSY experiments were recorded in 70% TFE-d_2-OH/ 30% H_2O and 70% TFE-d_2-OD/ 30% D_2O on an unlabelled sample. As with the mainchain, the great majority of NOEs involving the sidechain NH or NH_2 groups, but also aromatic ring protons, are to protons of residues located

within +3/-2 residues in sequence. A significant number of NOEs could not be assigned within this range, however, suggesting that they may arise from longer range contacts in the TFE state, such as those made as a consequence of the disulphide bridges, which remain intact in this denatured state. However, these NOE assignments could not be made unambiguously due to the overlap of ^1H resonances. The resonance assignments made in this study are listed in Table 1.

Analysis and interpretation of NMR data

The NMR parameters obtained for the TFE denatured state, i.e. the NOE patterns and intensities, H_α chemical shift perturbations, $^3J(H^NH_\alpha)$ coupling constants, and amide hydrogen exchange rates, reflect to different extents the population of structures present in the TFE denatured state. The information from the different NMR parameters is, however, complementary, so that the problems in the interpretation of these parameters associated with their dynamic averaging can be largely overcome by the combination of the various results. A further complication arises, because the majority of the data sets are incomplete in certain regions of the protein due to the severe overlap of resonances or their poor intensity.

The *pattern and intensities of short and medium range NOEs* of the mainchain are summarised in Figure 4. Apart from marginal changes in the dynamic behaviour, which improved the quality of spectra at the higher temperature, there is no evidence for significant structural changes in the TFE state over this small range of temperature from either CD or NMR spectroscopy and so the data at the two temperatures were combined to give the overall NOE pattern. Short range NOEs are shown to indicate the most prominent connections used in sequential NOE assignments. Intrinsic to the study of proteins under denaturating conditions is the incompleteness of the analysis of short and medium range NOEs, principally as a result of the limited chemical shift perturbations in 3D NOESY spectra of these states, as noted above. If two resonances overlap, the presence or absence of the NOE between them cannot be ascertained. Nevertheless, in comparison with similar patterns of medium range NOEs derived from ^{15}N edited 3D NOESY spectra for native hen lysozyme (data not shown), it is noticeable that a far greater number of medium range NOE are observed in spectra of the TFE denatured state. It is possible that spin diffusion may influence the exact NOE intensities of nearby protons, but, this is an unlikely explanation for the observation of a large number of medium range $\alpha N(i,i+3)$ and $\alpha N(i,i+4)$ NOEs, because they are clustered only in certain of the regions of the polypeptide. Thus they are a clear

indication for helical structure. Their coexistence with αN(i,i+2) and NN(i,i+2) NOEs is anticipated from the distance of 4.4 Å for these protons in helices, although in this case, spin diffusion may also make their signal more intense. Alternatively, these NOEs could arise from a coexistence of several conformational states, such as those involving helical structure and turns. This would result in a larger number of NOEs. Such conformational averaging is evident in the case of some of the NOEs involving sidechains near helix termini, where scN(i,i+3) or scN(i,i+4) NOEs coexist with scN(i,i-3) and scN(i,i-4) NOEs, suggesting the sidechains may explore a number of alternative conformations. These effects would not be anticipated for a polypeptide chain which is entirely fixed in an α-helical conformation. Because of the $1/r^6$ dependence of the NOE-cross peak intensity, relatively low populations of conformations may be detected, as long as protons are brought close enough in space a fraction of time.

Despite the considerable overlap in the NOESY spectrum, it is possible to outline the approximate regions which have particular structural preferences in the TFE denatured state. Several regions are characterised by an extensive pattern of αN(i,i+3), αN(i,i+4) NOEs, indicative of α-helical structure (Table 2A). Most of these regions correspond to those which are helical in the native state, but the correspondence with the native helix termini is limited in detail. Moreover, a number of regions of the polypeptide chain appear to have adopted helical structures that are not formed in the native state (see discussion below). Even for helices well defined by the other NMR parameters (see Figures 5, 6 and 8; Table 2A) the expected NOE patterns appear to be incomplete, notably towards the C-termini. Resonance overlap in the latter part of the helices, as seen for the H_α resonances particularly of residues 7 to 14 and residues 31 to 38 (Table 1), accounts in part for this observation. Because NOE detection in the present experiments relies primarily on the correlation with NH resonances, it is possible that different dynamic behaviour of helices towards their C-terminus may also be responsible for these observations. As a result of the hydrogen bonding pattern in helices, regions with significant hydrogen exchange protection and chemical shift perturbations are located towards the C-terminus of each helix. These correspond at least in part to those regions defined poorly by medium range αN(i,i+3) and αN(i,i+4) NOEs, thus providing complementary evidence for α-helical structure (in the case of residues 33 to 38 and 111 to 115 for example). Several regions are characterised by NOEs which are few in number and weak. They are unlikely to be part of well defined structures for a substantial fractions of time (Table 2A) and

support the observation from CD spectroscopy that TFE is unable to stabilise the entire polypeptide chain in a fully helical conformation.

Chemical shift perturbations of Hα resonances from random coil values (H_α^{obs}-$H_\alpha^{random\ coil}$) (Wüthrich, 1986) and the chemical shift index as defined by Wishart et al., (1991) are shown in Figure 5a,b. The higher than average perturbation to the chemical shifts of several protons can be associated with local features of the polypeptide sequence. Thus two of the most significantly perturbed protons are located adjacent to proline residues 70 and 79. Others, succeed tryptophan residues or are close to phenylalanines in sequence. H_α resonances which experience significant upfield perturbations (> -0.1 p.p.m.) are clustered in six regions of the protein sequence (Table 2A), an observation which is also reflected in the chemical shift index, suggesting the involvement of these residues in helical structure. Comparison with the corresponding data for the native state is complicated by the effects of tertiary structure such as those arising from proximity to aromatic rings, which are superimposed on the H_α chemical shift perturbation due to helical or sheet-like secondary structure. However, good overall agreement exists between the helix limits inferred from the H_α shift data in the TFE state and those defined on the basis of the crystal as well as the solution structure of the native protein. There are several exceptions in detail, however. These include the chemical shift perturbations at the C-terminus of the B-helix and the N-terminus of the D-helix as well as those in a region corresponding to the final turn and strand of the β-sheet in the native state. The implication that non-native helical structure exists in these regions is consistent with the other NMR parameters (Table 2A). The shift perturbations also suggest that the first 3_{10}-helix (residues 80-84 in the native state) is not well structured. Interestingly, glycine H_α resonances experience no or slightly downfield chemical shifts in the TFE state, suggesting they are not part of highly populated helix or turn-like structures.

$^3J(N^H,H_\alpha)$ Coupling constants measured at 58 mainchain sites are shown in Figure 6. It is apparent that the majority of these are characteristic of helical structures. A coupling constant of 3.3 Hz, inferred from the comformation of an ideal helix (Wüthrich, 1986), can be used to estimate the population of the helical part of ϕ space, on a simple two state model using 8.5 Hz for an extended conformation. The estimated populatiuons are in excess of 80% in the regions for which NOEs and shift perturbations characteristic of helix structure are detected and 30-40% elsewhere. However, this is undoubtedly a conservative estimate, as coupling constants calculated on the basis of helical regions of crystal structures are 4.8 Hz on average (Smith, Bolin, CMD in preparation).

Several residues have coupling constants > 5 Hz suggesting they are not part of well populated helical structure. Indeed, coupling constants of up to 7 Hz are detected in regions corresponding to the N-terminal part of the β-sheet (residues 40-48) and the loop region (residues 64-72) in the native state, suggesting that these regions are largely unstructured and conformationally averaged in the presence of TFE (Table 2A).

Direct detection of hydrogen exchange from the TFE denatured state

In Chapter 3 and 4 it was described how hydrogen exchange rates could be measured site specifically from the TFE denatured state of hen lysozyme in 50% TFE (v/v) at 27°C. This procedure allowed exchange to occur for defined periods of time from the TFE-state prior to refolding of the protein which quenched the exchange at 67 mainchain sites by the protection ~~protection~~ of amides in the native fold of the protein (Chapter 2). The extent of amide exchange could then be followed by the measurement of the amide occupancy in spectra of the native state. Bulk hydrogen exchange experiments in presence of 50% (v/v) TFE-d_2-OD/D_2O using fully and partially protonated protein molecules lead to the conclusion that the great majority of the amides with significant exchange protection in the TFE state are also those protected strongly in the native state. This implied that regions which are not protecting amides in the native fold could not be affording significant protection to them in the TFE state. Thus non-native helix structures induced by TFE would not be anticipated to have considerable persistency (see Chapter 3)

The assignment of the spectrum of the TFE denatured state, albeit under slightly different conditions, allowed us to test these conclusions by direct detection of hydrogen exchange kinetics. Amide proton occupancy was determined in a series of fast ^{15}N-^{1}H HSQC spectra recorded at 23 min intervals after the protonated protein was dissolved in 70% TFE-d_2-OD/D_2O, pH 2.0 at 37°C. Two examples of such spectra recorded 10 and 100 min after the initiation of exchange are shown in Figure 7a,b. The intrinsic exchange rate for an alanine residue in an unstructured model peptide is 0.84 min^{-1} under these conditions (Bai et al., 1993). Around 60 amides have already nearly fully exchanged in the spectrum recorded after 10 min and are likely to have protection factors less than 10. These amides are indeed located at sites of the polypeptide chain which are considerably disordered in the TFE state. Amide exchange rates were fitted to the intensity data of the more slowly exchanging amides and protection factors were calculated by comparison of these rates to those predicted for an amide in an unstructured model peptide, taking into

account the temperature dependence and sidechain inductive effects (Bai et al., 1993). The protection factors of 53 such amides are shown in Figure 8.

Noteably, amides in regions of the TFE denatured state corresponding to the centres of the native helices have amongst the highest protection factors. Several breaks in this pattern are introduced by amides whose resonances are very weak (residue 109, 123 and 125) or not observed (residue 34 and 96) in the HSQC spectrum of the fully protonated protein under these conditions. By contrast, measurements of resonances with high intensity (Figure 2) can be made with good signal-to-noise. The fact that the signal of these normally intense amides disappears from the spectrum quickly (Figure 2 and 7a, b), lends additional weight to the conclusion that these amides have exchange rates on the order of those predicted for these residues in disordered conformations. In agreement with the other NMR data, they are indeed located in regions of the protein which are predominantly unstructured in the TFE state. Bearing in mind that the experimental conditions are different, the data agree well with those reported previously (Chapter 4), both in magnitude and in the overall pattern. The most noticeable difference to the previous data is the high protection of a few amides sites close to the disulphides, involving cysteines 6, 30, 80 and 94. The majority of these sites are not, or only weakly, protected in the native state and therefore, were not monitored in the previous indirect experiments. Similarly residues Met105, Ala107 and Gln121 which are marginally protected in the TFE state by non-native extensions to the α-helices, are not protected in the native state.

The protection factors in regions which are structured in non-native α-helical conformations, such as part of the native-state β-sheet, are around 10 and their magnitude implies that these non-native structures have some marginal stability in the TFE denatured state of hen lysozyme. However, the data presented here confirm the previous conclusions that the α-helical structures are most persistent in regions of the protein corresponding to those forming helical structure in the native state.

Discussion

Analysis of conformational preferences in a partially ordered state by NMR of proteins

The structural characterisation of unfolded and partially folded states has recently become possible at a site-specific level through the additional resolution of resonances possible in heteronuclear NMR (Wüthrich, 1994). The combination of significant residual chemical shift dispersion of resonances, particularly of ^{15}N nuclei, and of resonance linewidths sufficiently narrow to yield intense signals has been central to this study using ^{15}N labelled hen lysozyme. Thus it has been possible to assign nearly completely resonances of a 129 residue protein by the use of a conventional assignment strategy relying on spin system, sequence and NOE connectivity information (Wüthrich, 1986). Hen lysozyme in TFE is the second example of a denatured state of a small water soluble protein, which contains extensive secondary structure but few tertiary interactions, that has been characterised in detail by heteronuclear NMR methods; the first being ubiquitin in methanol (Harding et al., 1991; Stockman et al., 1993), a protein of almost half its size. The results of the present study suggest that the detailed description of structure in states with characteristics which approach those of "molten globules" is becoming feasible by NMR.

The quality of NMR spectra of α-lactalbumin, which has a structure in its native state homologous to lysozyme, has been shown to improve considerably when TFE is added to the molten globule state of the protein (Alexandrescu et al., 1994b). It was suggested that the effect of TFE was to disrupt a significant proportion of the long-range tertiary, albeit often non-specific contacts, which exist in the molten globule state and give rise to conformational averaging of structures on the millisecond timescale and broadening of NMR resonances. Indications that such long-range interactions are a principal cause of the increased linewidths are also apparent in the case of the TFE-denatured state of hen lysozyme, where such effects are observed in close proximity to the native disulphide bonds which remain intact in this state (see also Chapter 8). While putative long range interactions are most likely to involve sidechains in the proximity of disulphides, their unambiguous characterisation has not been possible from the data available on the ^{15}N labelled sample and, therefore the tertiary structure and overall topology of the TFE denatured state can not yet be described in detail (but see Appendix to PartII of this thesis for results of model calculations).

It has been possible to extract NMR parameters for the great majority of mainchain sites, providing information on the nature of the structures which exist in the TFE-denatured state of hen lysozyme. Thus, the aim of the present study is to provide a detailed description of the structural propensities of the polypeptide chain in the presence of TFE cosolvent, to assess the persistency of the structures involved, and to compare these results with predictions based on helix propensities.

The nature of structure in the TFE denatured state

The individual criteria for the inferral of features consistent with α-helical conformations in regions of the polypeptide chain and the results from these analyses are described above and are summarised in Table 2A. A significant finding of these analyses is that the description of the conformational features they provide are in excellent correspondence for almost all of the residues (Table 2A). The consensus of these results suggests the following classification of the mainchain structure of the TFE denatured state of hen lysozyme (Table 2B):

Native-like helices: The data suggest that all residues which are in α-helical structures in the native state are also involved in α-helical structure in the presence of 70% TFE. The NMR parameters, including protection against hydrogen exchange, imply that the centres of these native-like helices have a high degree of structural persistency, meaning that the population of any non-helical conformational states are very low for these regions.

Non-native conformations of helical character: These structures are formed as extensions to the native-like helices B, C, and D, and the C-terminal 3_{10}-helix, by up to one turn at either end. In addition residues 50-58, which form part of the β-sheet in the native state, are also stabilised in a helical conformation. None of these non-native helical structures or helix extensions are found to be as persistent as the native-like helices as judged by the hydrogen exchange and ^{15}N relaxation behaviour (see Chapter 8) of the amides involved in them.

Other preferences: Although not necessarily towards helical structures, other conformational preferences can be inferred for short regions of the polypeptide, which may, at least transiently, participate in turn-like structures. The regions comprising residues 17-20 and 60-64 are examples of such behaviour.

Highly unstructured regions of the polypeptide chain; The regions of residues 40-49 and 65-75 corresponding respectively to the N-terminal part of the β-sheet and part of the long loop in the native protein appear to be highly disordered in presence of TFE (Table 2A).

An overview of the secondary structure in the TFE denatured state inferred from the NMR data is shown in comparison with that of the native state in Figure 9a,b and see Table 2. Thus approximately 73 residues are involved in α-helical structure in the TFE denatured state, which compares closely to the 80 residues which are inferred to be in α-helical structure from the far uv CD signal under similar conditions (Chapter 4). This suggests that possible contributions due to helix regularity, disulphides and aromatic residues, (Woody, 1985) are not the principal components responsible for the increase of the far uv CD ellipticity in the TFE state over that of the native state, but that non-native α-helical structure is. For example, conversion of the two 3_{10}-helices, which are thought to contribute little to the far uv CD ellipticity at 222 nm (Gans et al., 1991) in the native state, to α-helical conformation and the formation of α-helical struture in part of the region corresponding to the β-sheet in the native state are likely to account for the majority of the increase in CD signal. Furthermore, the data on the location and extend of the α-helical structure in the TFE state are highly consistent with a recent study by circular dichroism of four peptide fragments, spanning the sequence of hen lysozyme in segments of 20 to 46 residues in length (Yang et al., in preparation). Is is not possible, for example, to stabilise α-helical structure in a peptide corresponding to the long loop region of the native protein (residues 61-82) even at a TFE concentration of 95% (v/v) (M. Kotik - unpublished data), By contrast, studies of a peptide corresponding to the β-sheet region of the native protein (residues 41-60) suggested that several residues may adopt α-helical conformation in presence of 50% TFE (Yang et al., 1994). Two peptides, residues 1-40 and 84-129 corresponding to the regions of the protein which form the major α-helices in the native state were both found to adopt extensive α-helical structure in the presence of TFE. Remarkably, the sum of the far uv CD spectra of these four peptides was found to be nearly identical to spectra obtained under identical solution conditions for the intact oxidised protein as well as for a derivative of lysozyme, in which the disulphides had been reduced and the cysteines blocked by methylation (Chapter 5). These results imply that the determinants which allow the polypeptide chain to adopt stable helical structures must be associated with highly localised features of the amino acid sequence.

Comparison with predictions for local structural propensities

Figure 9 summarises the nature of helical structures in the TFE state together with the intrinsic propensity for α-helix formation in the hen lysozyme sequence. The profile for helix propensity derived with the scale of Chou & Fasman (1978) is shown in Figure 9c and illustrates the features common to this and the profiles obtained using the helix propensity scales of Richardson &

Richardson (1988), O'Neil & Degrado, (1990), Wojcik et al., (1990), Horowitz et al., (1992), Chakrabartty et al., (1994) and of Blaber et al., (1994). Regions predicted to have high intrinsic α-helical propensity are found to be in α-helical conformations in both the TFE as well as in the native state. There is little difference in the helical propensity for regions which form α- and 3_{10}-helices in the native state, suggesting their structure in the native fold could be influenced by other interactions. Part of the region which forms the triple stranded β-sheet in the native state, has significant helical propensity as judged by higher than average propensity on 5 of the 7 propensity scales examined. This region has indeed been found to adopt helical structure in the TFE state of the intact protein. Thus the overall correspondence between the regions of the polypeptide chain which have adopted an α-helical conformation in TFE and those with considerable intrinsic propensity for helical structure is good overall. However, in a detailed comparison several exceptions to this correspondence are found particularly at helix termini. For example, the N- and C-terminal extensions to the B-helix are not anticipated on the basis of the predicted propensities of these residues. Furthermore, higher than average helical propensity is seen for residues 41-44 and 61-64 in patterns derived using the majority of the propensity scales, but is not accompanied by helix formation in the TFE- denatured state. Helical propensity in region of residue 80-84 is fulfilled by helix formation in the native state, but not in the TFE state where the helix begins around residue 85. These discrepancies may be due to the fact, that the propensity scales have been derived for individual residues located in the central regions of helices and do not take into account explicitly medium range interactions involving the sidechains of a helix (Scholtz & Baldwin, 1992).

Interactions between sidechains at the termini of helices and the mainchain, have, for example, been found to confer significant stability to helix formation (Serrano & Fersht, 1992; Serrano et al., 1992). The details of some of the capping interactions are, however, complicated (Milner-White, 1988; Harper & Rose, 1993; Harpaz et al., 1994). For example, aspartic and glutamic acid residues are preferred amino acids at the first or second position of a helix, reinforcing the capping effect of sidechains at the N-terminus (Dasgupta & Bell, 1993). A simplified profile of sites with N- and C-capping potential are plotted as a function of the protein sequence in Figure 9d and e using a scale derived from energetic analysis of site directed helix capping mutants (Munoz & Serrano, 1994). The pattern constructed using scales derived from structural statistics at helix termini is closely similar (Richardson & Richardson, 1988; Presta & Rose, 1988; Dasgupta & Bell, 1993). It is noticeable that while the A- and B-helices

have helix limits in the native state which coincide with the positions of amino acids suitable for capping, such correspondence seems to be much poorer for the termini of the C- and particularly the D-helix. In these cases alternative interactions, part of the native tertiary fold, appear to compensate for the loss of these favourable interaction in the native state. For example Ile98, Met105 and Trp108 are buried in the hydrophobic core of the protein, which is likely to result in conformations unsuitable for helix propagation.

A remarkable observation is, therefore, that the limits of α-helical structure as identified by the NMR parameters in the TFE state closely coincide with the positions of residues which can participate favourably in helix capping interactions. For example, the N-terminal extension of the D-helix to around residue 103 would allow for many capping interactions not possible at residue 108. The same is true for the second 3_{10} helix. Gly117 is a good capping residue whose position is reinforced by an aspartic acid at position 119. Equally residues 125-128 could be involved in C-caps. Another N-cap motif involving aspartic acid at position N+2, may be responsible for initiation of the first 3_{10}-helix near residue 85. The N- and C-terminal extension of the B-helix are also consistent with the location of alternative capping residues. Gly49 and Asp52 are favourable candidates for N-capping of the non-native α-helix which has been induced in part of the region corresponding to the β-sheet of the native state. It is noticeable, however, that these helices do not extend further, for example residues 40, 43, 44, 46-49 are candidates for N and C-caps for further extension of the B-helix. However, the helix propensity present in the first part of this region is primarily due to the presence of Ala41. As seen also for the region corresponding to the long loop of the native state, propensities to form favourable helix capping interactions appear to be insufficient on their own for helix formation in TFE.

In addition to the intrinsic helical propensity of individual residues and the interactions at helix termini, medium range interactions within a helix, for example between sidechains on adjacent turns, are also known to contribute to its structural stability. Moreover, high propensity for β-strand or turn formation would complete with the tendency of a polypeptide to adopt α-helical structure. A number of secondary structure prediction programs have been designed to include most, if not all, these contributions and were used to examine further the propensity of the hen lysozyme sequence to adopt helical structure (Figure 10). It is apparent that while the results of these predictions are similar for several regions of the polypeptide, for example for the region corresponding to the A-helix, they do not agree in detail with each other and

with the structure in the TFE or native state. This may be in part because ALB (Finkelstein et al., 1991), FUGUE (Rooman et al., 1992) and AGADIR (Munoz & Serrano, 1994), aim to predict structural propensity of the polypeptide as an isolated chain in aqueous environments, whereas the method by Chou & Fasman (1978) and GOR (Gibrat et al., 1987) are based on the statistics of structures in folded proteins. The correspondence of the experimental results with those of the latter two methods are poorer, as the propensity for the formation of β-structure is apparently higher than that for α-structure even in the regions corresponding to the helices in the native protein. Notably, the preference for helical structure in part of the β-sheet (Figure 9c) was based on the Chou & Fasman scale for helix propensity alone, while an even stronger preference for β-structure appears to exist in this and other regions of the polypeptide. This suggests that TFE has the potential to stabilise regions in helical structure even when these have only weak propensity. The formation of β-structures may be disfavoured in TFE because interactions are usually non-local and are often stabilised by interactions involving hydrophobic sidechains.

Overall, the data on hen lysozyme reinforces the notion that TFE stabilises helical structures only in regions of the polypeptide which have considerable local propensity to adopt helical conformation. This conclusion is in agreement with previous studies of peptides and protein fragments (Dyson et al., 1992; Lehrman et al., 1990; Sönnichsen et al., 1992). Additionally, it appears from other studies that it is often the native helix capping interactions which persist in the cosolvent and act as helix stop signals (Nelson & Kallenbach, 1986, 1989; Bruch et al., 1991; Storrs et al., 1992; Shin et al., 1993; Toumandjie & Johnson, 1994).

A "conformational switch" from β -to α-structure is observed on addition of TFE in the region of hen lysozyme corresponding to the β-sheet in the native state. This suggests that the final conformation in the native state may not be the one which is preferred on the basis of the structural propensities involving only local interactions. However, the present findings support the conclusions from other studies that such β- to α-transformation can be rationalised on the basis of local chain propensities towards α-structure (Sönnichsen et al., 1992; Lui et al., 1994, Y. Goto - personal communication).

The study on the TFE denatured state of α-lactalbumin, however, suggests that such a correspondence between prediction and the observed structures in TFE may not be general, however. Alexandrescu et al., (1994b) assigned around 25%

of the homonuclear spectra of α-lactalbumin under solution conditions similar to those of the present study. The data implied that residues 35-41, following the region which constitutes the B-helix in the native state of α-lactalbumin, adopts a non-native α-helical conformation in presence of TFE. This is despite the fact that no appreciable propensity can be detected for this region of the amino acid sequence by any of the scales and prediction programmes used in the present study. Another region of he protein, residues 101-110, was not found to assume α-helical conformation, even in high TFE concentrations (Smith et al., 1994; Alexandrescu et al., 1993, 1994b), even though high helical propensity is predicted for this region of the polypeptide. It will be interesting to see in future studies, whether these observations are particular to α-lactalbumin or whether they extend to other systems. However, it is likely that propensity scales and prediction methods are still inadequate in many cases and that the structures which can be formed purely by local interactions in particular environments are still incompletely understood.

Cosolvent denatured states and the role of trifluoroethanol

The nature of the interaction of TFE with a polypeptide chain are not understood in detail (Nelson & Kallenbach, 1986; Stoors et al., 1992). However, from its effect on the conformation of hen lysozyme and numerous peptides, TFE may be regarded as an agent which disrupts long range tertiary interactions, presumably by weakening of the hydrophobic effect relative to hydrogen bonding interactions (Thomas & Dill, 1994). As a consequence the TFE denatured states of hen lysozyme and α-lactalbumin do not resemble classical molten globule states in several of their global properties (Alexandrescu et al., 1994b; Chapter 4 and 5). For example, the TFE denatured states of these proteins do not appear to bind the hydrophobic surface probe, 1-anilinonaphthalene-8-sulphonate, known as ANS. In the case of hen lysozyme, low angle X-ray scattering data suggest that the protein has a non-native shape and is more expanded than would be anticipated for a molten globule (Y. Goto - personal communication). The far uv CD data and the results of the present study show that α-helical structure has been preferentially stabilised, possibly accounting for the fact that the protection factors against amide hydrogen exchange are higher than those seen in other partially structured states (Chapters 3 and 4). Disruption of longer range contacts, however, appears to allow the polypeptide to adopt conformations which involve only local interactions. This is thought to introduce a bias towards helical structures because β-sheet structures are often stabilised by non-local hydrophobic interactions between sidechains (Dill et al., 1993). However, there is recent evidence that TFE can also stabilise certain peptides in turn-like

conformations including β-turns and β-hairpins (Sönnichsen et al., 1992; Shin et al., 1993; Smith et al., 1994; Blanco et al., 1994).

Implications for protein folding

The previous finding (Chapter 4) that regions which are organised as helices in one subdomain of the native state of hen lysozyme also form the most persistent helical structures in a partially ordered state of the protein in the absence of extensive and specific tertiary contacts, is confirmed in the present direct and high resolution NMR study of its conformation in TFE. This supports the notion that if an equivalent state were to exist on the folding pathway productive folding may be accomplished by the simple coalescence of such preformed elements of secondary structure (Chapter 4). Intriguingly, model calculations on the arrangement of these helices in the presence of the native disulphide bonds, but no other long range interactions, suggest that these restraints may be sufficient to generate with a considerable probability, regional topologies in the TFE denatured state which have characteristics of the native fold (see Appendix).

A significant finding of the structural analysis presented here is the confirmation that a significant proportion of the polypeptide chain has non-native conformational preferences in the partially ordered state of hen lysozyme in presence of TFE. Similar observations have been made in a number of other studies on proteins and peptides under equilibrium conditions (Sönnichsen et al., 1992; Kemmink & Creighton, 1993; Lui et al., 1994; Alexandrescu et al., 1994a; Arcus et al., 1994). Non-native structures are likely to arise because the specific interactions which contribute to their formation and stability are entirely local, while in some cases additional non-specific interactions are also provided by the rest of the polypeptide or by cosolvents or denaturants. By contrast, the formation of the corresponding native structures must therefore involve a significant proportion of specific long range interactions. Depending on the environment that structures find themselves in at a given stage of the folding process, these latter interactions may not be fully or specifically formed. Local non-native structures may, therefore, precede and indeed compete with the higher order structures formed later in folding. The notion that environments equivalent to those provided by TFE, exist in loosely folded proteins or as part of membranes (Norwood et al., 1992; Rizo et al., 1993), implies that the structures detected in this cosolvent may have some relevance to non-native structures which may exist during the folding process. An observation made in kinetic studies of the refolding behaviour of hen lysozyme, for example, is that the formation of the native

structure in the β-subdomain of the protein proceeds much slower than the formation of stable structure in the α-subdomain for the majority of molecules, when refolded form a variety of denatured states (Radford et al., 1992). The possibility that a transient population of non-native structure in the β-sheet region may be responsible of these observations and that these structures may be similar to those formed in TFE is an intriguing one, but requires experimental verification. However, the observations made here and previously, that the majority of the most persistent structures in this denatured state are native-like supports a model of protein folding which is based predominantly on a hierarchical process of structure formation and stabilisation (see Chapter 9).

Conclusions

The present study has demonstrated that it is possible to examine in considerable detail the structural properties of a partially ordered state of hen lysozyme, a protein of 129 amino acid residues, by means of recently developed heteronuclear NMR methods. The assessment of conformational features of a protein in such a state is greatly enhanced by the consideration of a number of complementary NMR parameters. The conformational preferences of the polypeptide chain have been found to be in close correspondence with the structural propensities for helix formation anticipated from secondary structure predictions. This suggests that the features which determine structures in this denatured state are largely intrinsic to the amino acid sequence and are highly localised.

The aim of studies on models for proteins in states intermediate between native and fully unfolded conformations, is not just to define the structures which are present in them, but also to characterise their stability and thus their importance for folding. Such information may in part be provided by a characterisation of the dynamic behaviour of the structures and is described in the following chapter.

References

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

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

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

Alexandresu, A.T. & Shortle, D. (1994c) J.Mol.Biol. 242, 527-546.

Arcus, V., Vuilleumier, S., Freund, S.M.V., Bycroft, M. & Fersht, A.R. (1994) Proc.Natl.Acad.Sci. (USA) 91, 9412-9416.

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

Barrick, D. & Baldwin, R.L. (1993) Protein Science 2, 869-876.

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

Blaber, M., Zhang, X-J., Lindstrom, J.D., Pepiot, S.D., Baase, W.A. & Matthews, B.W. (1994) J.Mol.Biol. 235, 600-624.

Blanco, F.J., Jimenez, A., Pineda, A., Rico, M., Santoro, J. & Nieto, J.L. (1994) Biochem. 33, 6006-6014.

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

Bruch, M.D., Dhingra, M.M. & Gierasch, L.M. (1991) Proteins: Struct.Funct. & Genetics 10, 130-139.

Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994) Protein Science 3, 843-852.

Chou, P.Y. & Fasman, G.D. (1978) Ann.Rev.Biochem. 47, 251-276.

Creighton, T.E. (1988) Biophys.Chem. 31, 155-162.

Creighton, T.E. (1990) Biochem.J. 270, 1-15.

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

Dagsgupta, S. & Bell, J. (1993) Int.J.Pep.Prot.Res. 41, 499-511.

Dill, K.A. & Shortle, D. (1991) Ann.Rev.Biochem. 60, 795-825.

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

Dobson, C.M., Evans, P.A. & Williamson, K.L. (1984) FEBS Lett. 168, 331-334.

Dobson, C.M. (1991) Curr.Opin.Struct.Biology 1, 22-27.

Dobson, C.M. (1994) Curr.Biol. 4, 636-640.

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

Driscoll, P.C., Clore, G.M., Marion, D., Wingfield, P.T. & Gronenborn, A.M. (1990) Biochem. 29, 3542-3556.

Dyson, H.J. & Wright, P.E. (1991) Ann.Rev.Biophy.Biophy.Chem 20, 519-538.

Dyson, H.J., Merutka, G., Waltho, J.P., Lerner, R.A. & Wright, P.E. (1992) J.Mol.Biol 226, 795-817.

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

Evans, P.A., Topping, K.D., Woolfson, D.N. & Dobson, C.M. (1991) Proteins: Struct.Funct.& Genetics 9; 248-266.

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

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

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

Finkelstein, A.V., Badretdinov, A.Y. & Ptitsyn, O.B. (1991) Proteins: Struct.Funct. & Genetics. 10, 287-299.

Frenkiel, T., Bauer, C., Carr, M.D., Birdsall, B. & Feeney, J. (1990) J.Magn.Reson. 90, 420-425.

Gibrat, J.-F., Garnier, J. & Robson, B. (1987) J.Mol.Biol. 198, 425-443.

Gans, P.J., Lyu, P.C., Manning, M.C., Woody, R.W. & Kallenbach, N.R. (1991) Biopolymers 31, 1605-1614.

Handoll, H.H.G. (1985) D.Phil.Thesis, University of Oxford.

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

Harpaz, Y:, Elmasry, N., Fersht, A.R. & Henrick, K. (1994) Proc.Natl.Acad.Sci. (USA) 91, 311-315.

Harper, E.T. & Rose, G.D. (1993) Biochem. 30, 7605-7607.

Horovitz, A., Matthews, J.M. & Fersht, A.R. (1992) J.Mol.Biol. 227, 560-568.

Hua, Q-X., Kochoyan, M. & Weiss, M.A. (1992) Proc.Natl.Acad.Sci. (USA) 89, 2379-2383.

Ikura, M., Kay, L.E. & Bax, A. (1990) J.Magn.Reson. 86, 204-207.

Jeener, J., Meier, B.H., Bachmann, P. & Ernst, R.R. (1979) J.Chem.Phys. 71, 4556-4553.

Jennings, P.A. & Wright, P.E. (1993) Science 262, 892-900.

Kay, L.E. & Bax, A. (1989) J.Magn.Res. 86, 110-126.

Kay, L.E., Ikura, M., Tschudin, R. & Bax, A. (1990) J.Magn.Reson. 89, 496-514 .

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

Kemmink, J. & Creighton, T.E. (1993) J.Mol.Biol. 234, 861-878.

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

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

Lehrman, S.R., Tuls, J.L. & Lund, M. (1990) Biochem. 29, 5590-5596.

Logan, T.M., Olejniczak, E.T., Xu, R.X. & Fesik, S.W. (1993) J.Biomol.NMR 3, 225-231.

Logan, T.M., Theriault, Y. & Fesik, S.W. (1994) J.Mol.Biol. 236, 637-658.

Liu, Z-P., Rizo, J. & Gierasch, L.M. (1994) Biochem. 33, 134-142.

Matthews, C.R. (1993) Ann.Rev.Biochem. 62, 653-683.

Mayo, S.L. & Baldwin, R.L. (1993) Science 262, 873-876.

Milner-White, E.J. (1988) J.Mol.Biol. 199, 503-511.

Munoz, V. & Serrano, L. (1994) Nature: Struct.Biol. 1, 399-409.

Nelson, J.W. & Kallenbach, N.R. (1986) Proteins: Struct.Funct. & Genetics 1, 211-217.

Nelson, J.W. & Kallenbach, N.R. (1989) Biochem. 28, 5256-5261.

Neri, D.,Wider, G. & Wüthrich, K. (1992a) Proc.Natl.Acad.Sci.(USA) 89, 4397-4401.

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

Norwod, T.J., Crawford, D.A., Steventon, M.E., Driscoll, P.C. & Campbell, I.D. (1992) Biochem. 31, 6285-6290.

O'Neil, K.T. & Degrado, W.F. (1990) Science 250, 646-651.

Peng, Z. & Kim, P.S. (1994) Biochem. 33, 2136-2141.

Ptitsyn, O.B. (1992) in "Protein Folding", T.E. Creighton Ed., Freeman, N.Y.

Presta, L.G. & Rose, G.D. (1988) Science 240, 1632-1641.

Radford, S.E., Dobson, C.M. & Evans, P.A. (1992) Nature 358, 302-305.

Redfield, C., Boyd, J., Smith, L.J., Smith, R.A.G. & Dobson, C.M. (1992) Biochem. 31, 10431-10437.

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

Richardson, J.S. & Richardson, D.C. (1988) Science 240, 1648-1652.

Rizo, J., Blanco, B.K., Bruch, M.D. & Gierasch, L.M. (1993) Biochem. 32, 4881-4894.

Roberts, I.N., MacKenzie, D.A., Jeenes, D.B., Radford, S.E., Robinson, C.V., Aplin, R.T. & Dobson, C.M. (1992) Biotechnology letters 14, 897-902.

Roder, H. (1989) Meth.Enzym. 176, 446-473.

Rooman, M. J., Kocher, J.-P. A. & Wodak, S. J. (1992) Biochem. 31, 10226-10238.

Scholtz, J.M. & Baldwin, R.L. (1992) Ann.Rev.Biophy.Biomol.Struct. 21, 95-118.

Segawa, S-S., Fukundo, T., Fujiwara, K. & Noda, Y. (1991) Biopolymers 31, 497-509.

Serrano, L. & Fersht, A.R. (1989) Nature 342, 296-299.

Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A.R. (1992) J.Mol.Biol. 227, 544-559.

Shin, H-C., Merutka, G., Waltho, J.P., Wright, P.E. & Dyson, H.J. (1993) Biochem. 32, 6348-6355.

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

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

Sönnichsen, F.D., Van Eyk, J.E., Hodges, R.S. & Sykes, B.D. (1992) Biochem. 31, 8790-8798.

Staley, J.P & Kim, P.S. (1990) Nature 344, 683-688.

Stockman, B.J., Euvrard, A. & Scahill, T.A. (1993) J.Biomol.NMR 3, 285-296.

Storrs, R.W., Truckses, D. & Wemmer, D.E. (1992) Biopolymers 32, 1695-1702.

Thomas, P.D. & Dill, K. (1993) Protein Science 2, 2050-2065.

Toumadje, A. & Johnson, W.C. (1994) Biopolymers 34, 969-973.

Vuister, G. & Bax, A. (1993) J.Am.Chem.Soc. 115, 7772-7777.

Wider, G., Neri, D. & Wüthrich, K. (1991) J.Biomol.NMR 1, 93-96.

Wilson, K.P., Malcolm, B.A. & Matthews, B.W. (1993) J.Biol.Chem. 267, 10842-10849.

Wishart, D.S., Sykes, B.D. & Richards, F.M. (1992a) Biochem. 31, 1647-1651.

Wishart, D.S., Sykes, B.D. & Richards, F.M. (1992b) J.Mol.Biol. 222, 311-333.

Wojcik, J., Altman, K.-H. & Scheraga, H.A. (1990) Biopolymers 30, 121-134.

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

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

Woody, R.W. (1985) in "The Peptides", Vol 7, pp. 15-114. Academic Press.

Wüthrich, K. (1986) "NMR of Proteins and Nucleic Acids", John Wiley & Sons, N.Y..

Wüthrich, K. (1994) Curr.Opin.Struct.Biol. 4, 93-99.

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

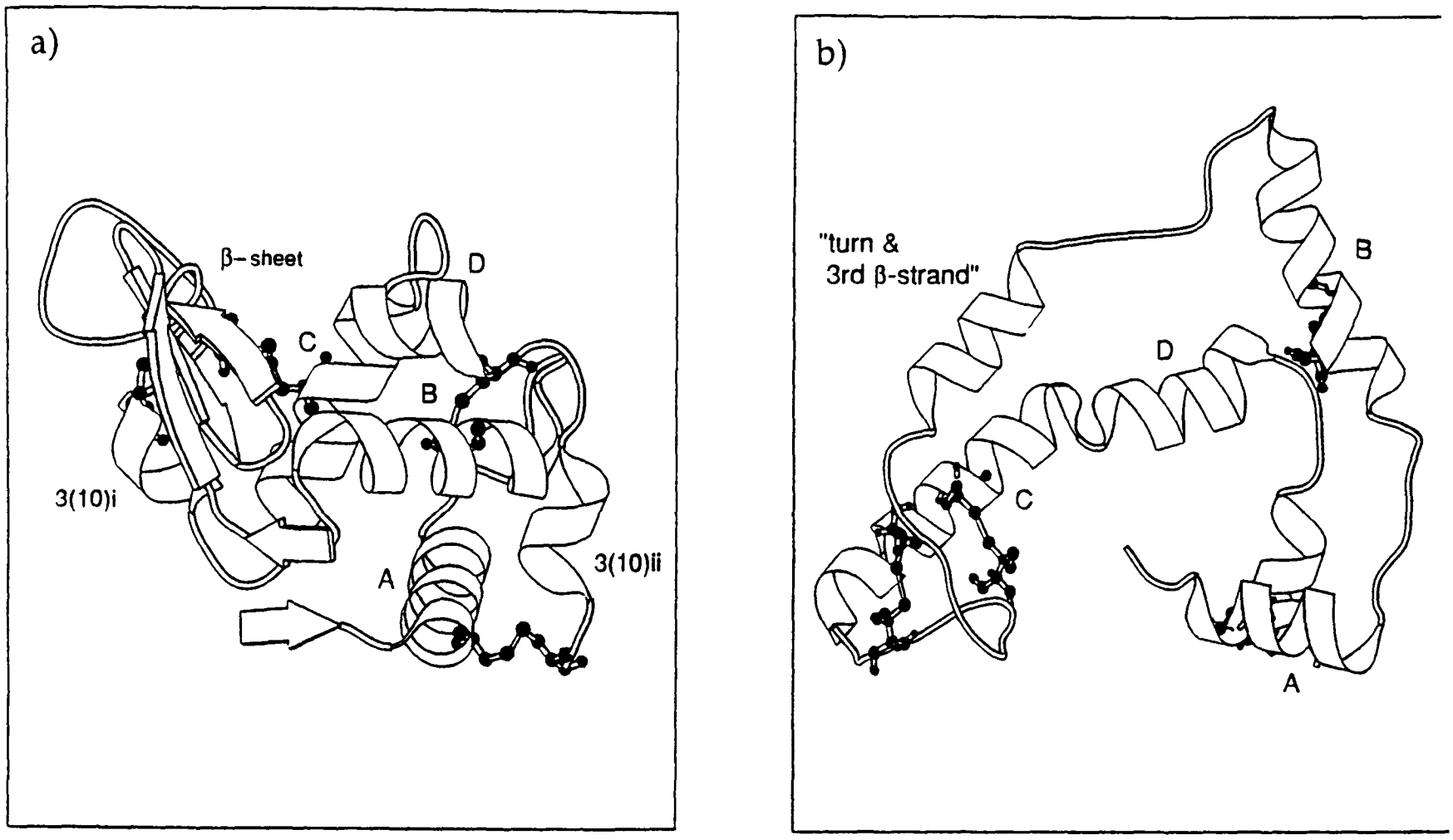


Figure 1
Schematic outline of the fold of hen lysozyme drawn using Molscript (Kraulis, 1991) in a) the native state, in b) as a representation of structure consistent with the NMR data for the TFE denatured state. Note that this is only one of many possible representations, at least as far as overall topology of the conformation is concerned, as this has not yet been defined by teritary contacts except by those arising from the native disulphides (see Appendix II for model calculations). The main secondary structural units are labelled and the locations of the four disulphide bonds are indicated in a) and b).

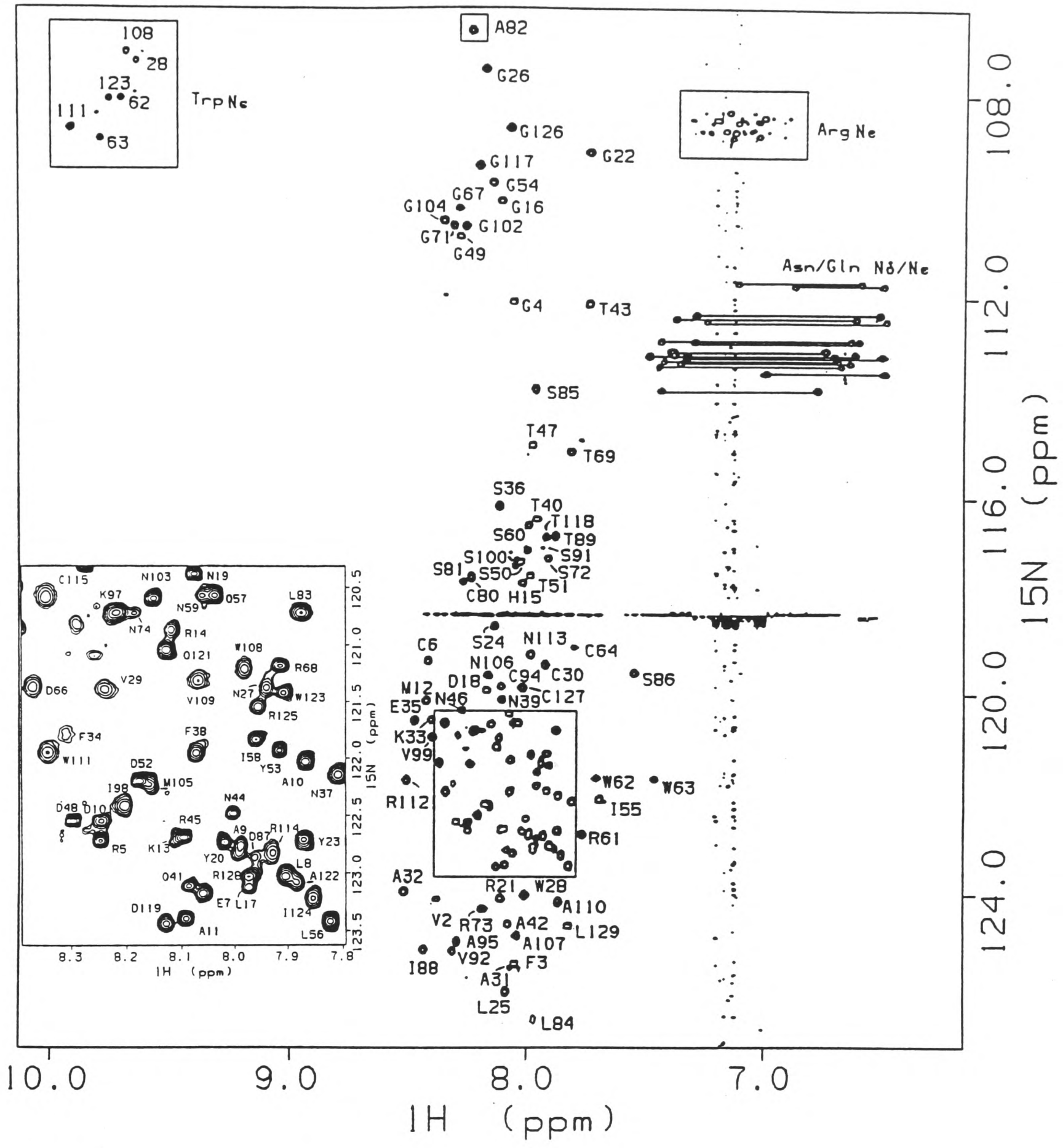


Figure 2

^{15}N-^{1}H HSQC spectrum of the TFE denatured state at 37°C using a spectrometer with 750 MHz ^{1}H spectrometer frequency. 117 out of the 126 expected mainchain ^{15}N-^{1}H correlations are assigned. Peaks labelled * could not be assigned at this temperature. Cross-peaks in boxes around 108 p.p.m. ^{15}N chemical shift are resonances which are folded in ω_1.

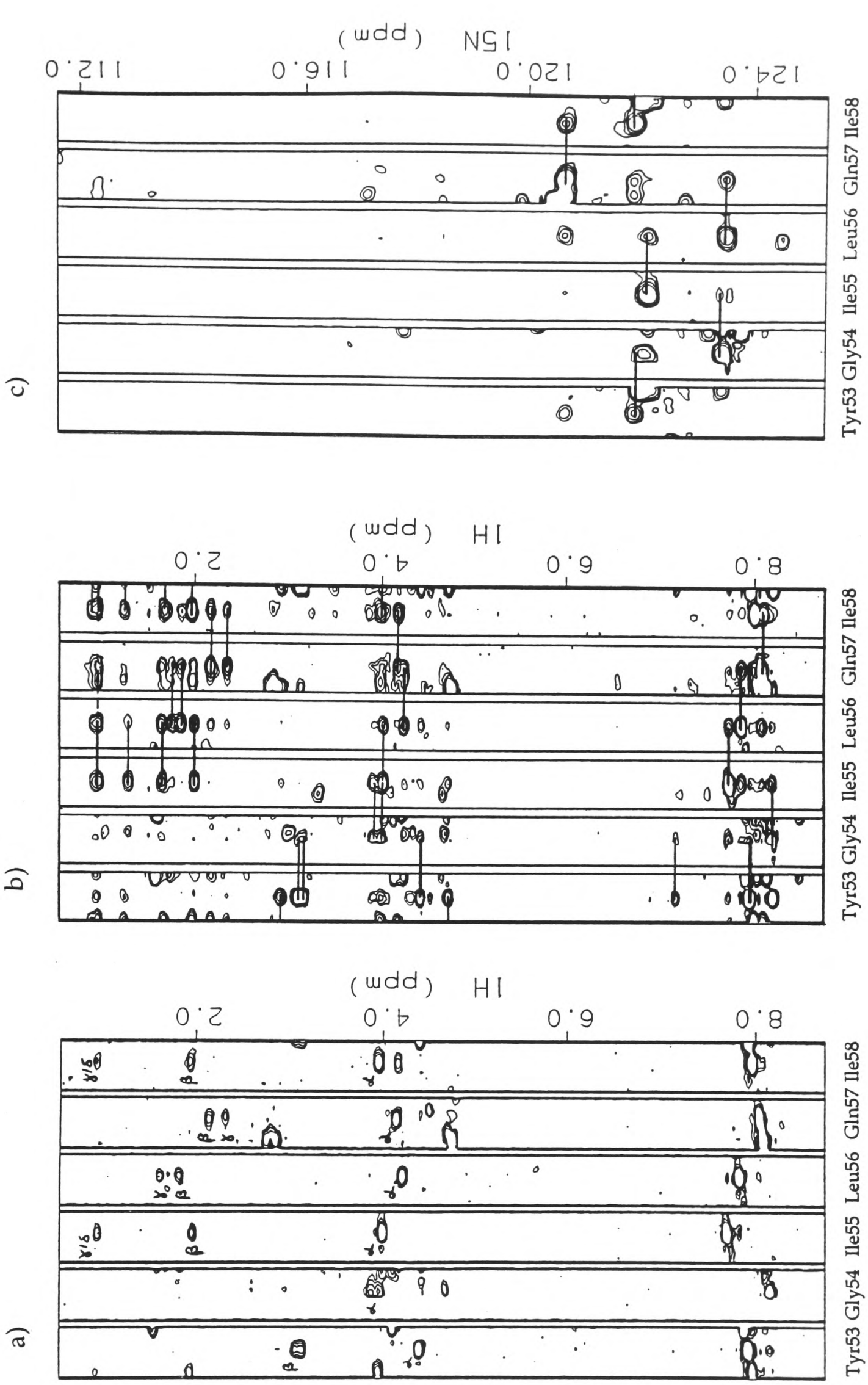


Figure 3

Representative regions of spectra of the TFE denatured state at 37°C used in the assignment process, shown as 3D strip plots. The region of residues 53 to 58, corresponding to part of the region which forms the main β-sheet in the native state of hen lysozyme, is shown in a) the 3D TOCSY-HSQC b) the 3D NOESY-HSQC and c) the 3D HSQC-NOESY-HSQC spectrum. Sequential mcN,$N_{i,i+1}$ and sc,$N_{i,i+1}$ NOEs are indicated.

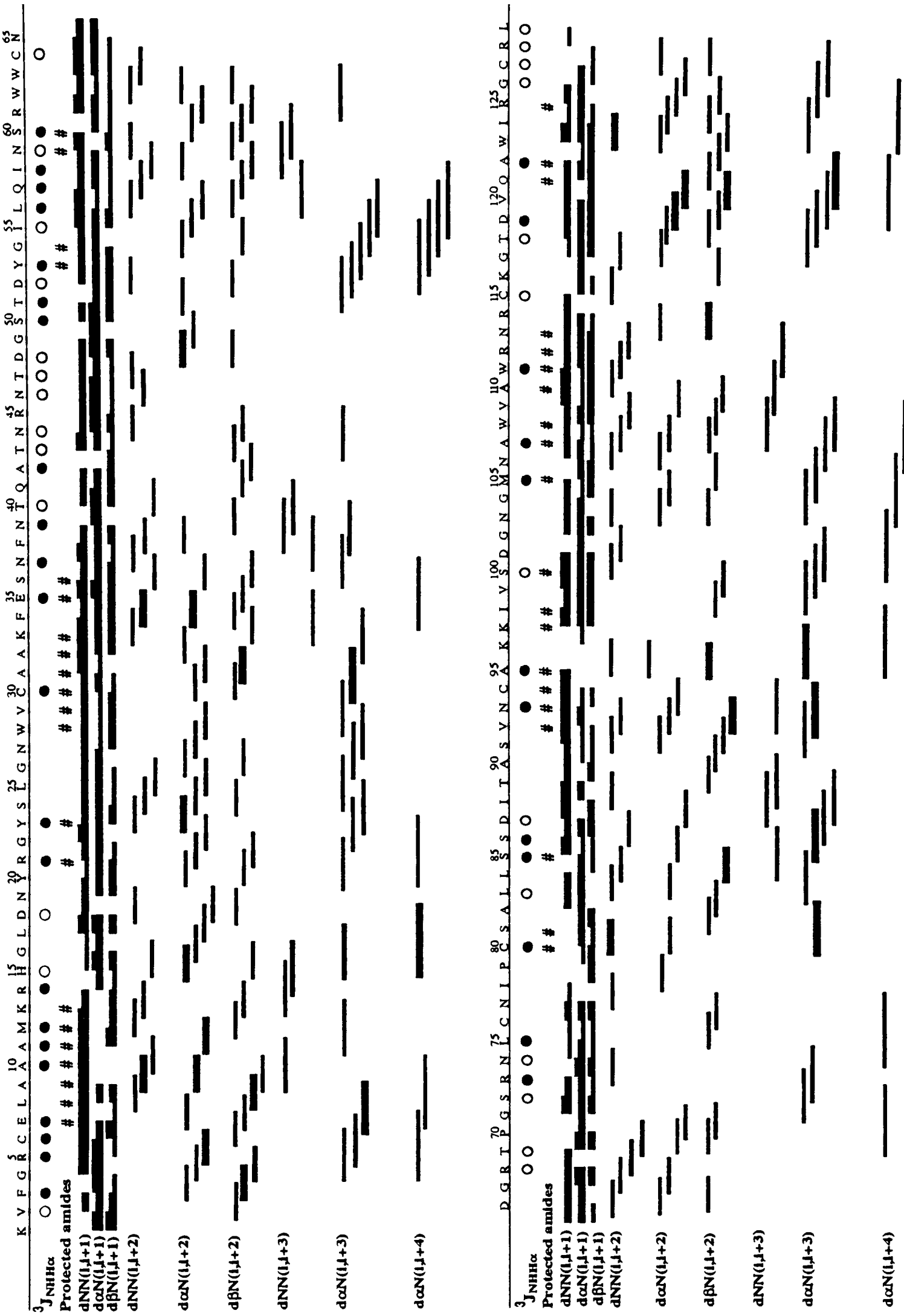


Figure 4

Overview of short and medium range NOEs in the TFE state. The data is the sum of the NOEs seen in spectra at 37°C and 47°C. Relative intensities are indicated as strong, medium and weak. Sites with amide protection factors exceeding 10 are indicated. Filled and open circles indicate sites for which $^3J(N^H,H\alpha)$ coupling constants were measured as less than 5.0 and between 5.0 and 8.0 Hz, respectively.

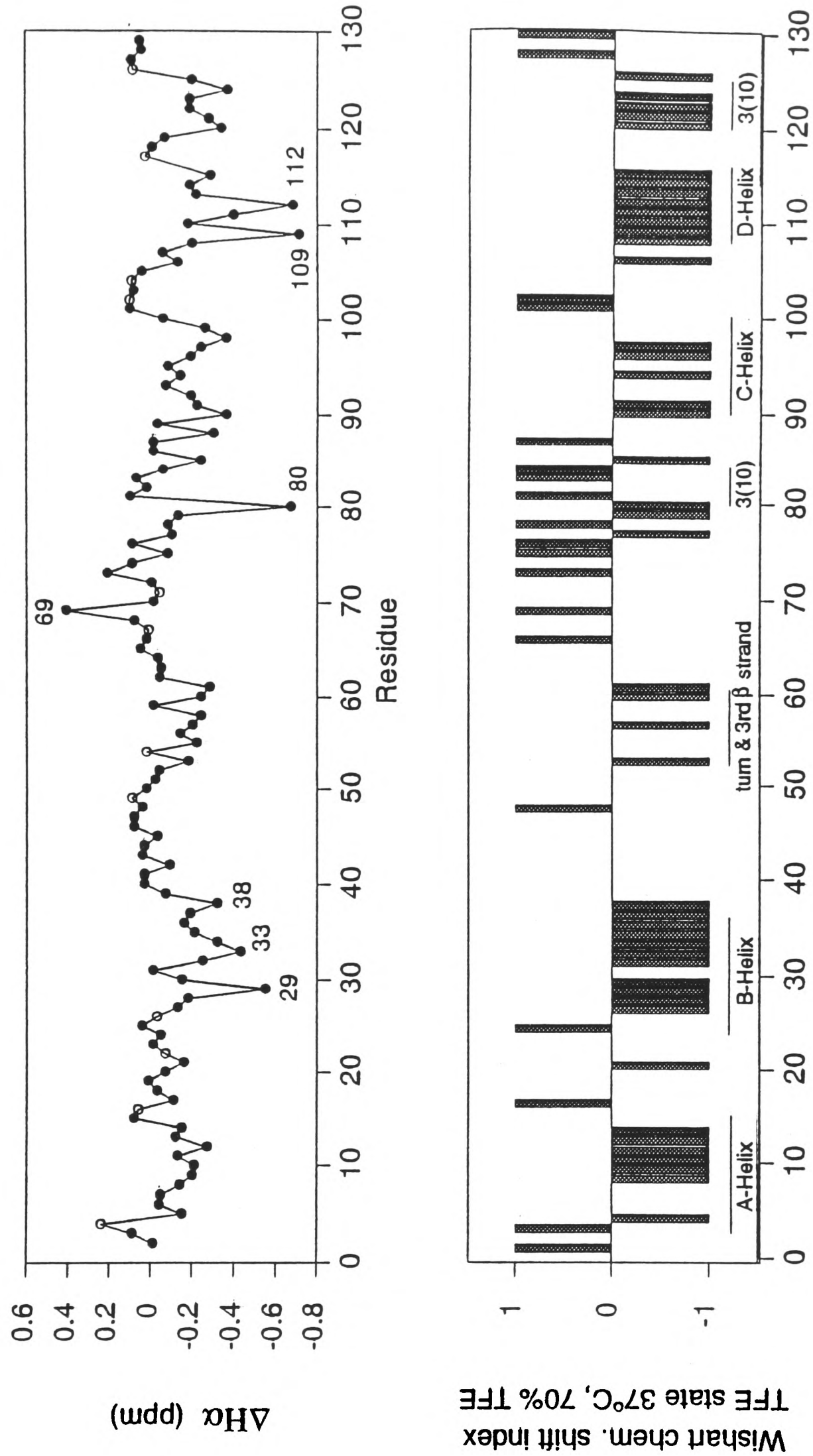


Figure 5

a) H_α chemical shift perturbations from random coil values of Wüthrich (1986) and b) the Chemical Shift Index of Wishart et al., (1992a, b) plotted as a function of protein sequence. The data for glycines is plotted as open circles to indicate their position. Regions corresponding to principal elements of secondary structure of the native state are marked.

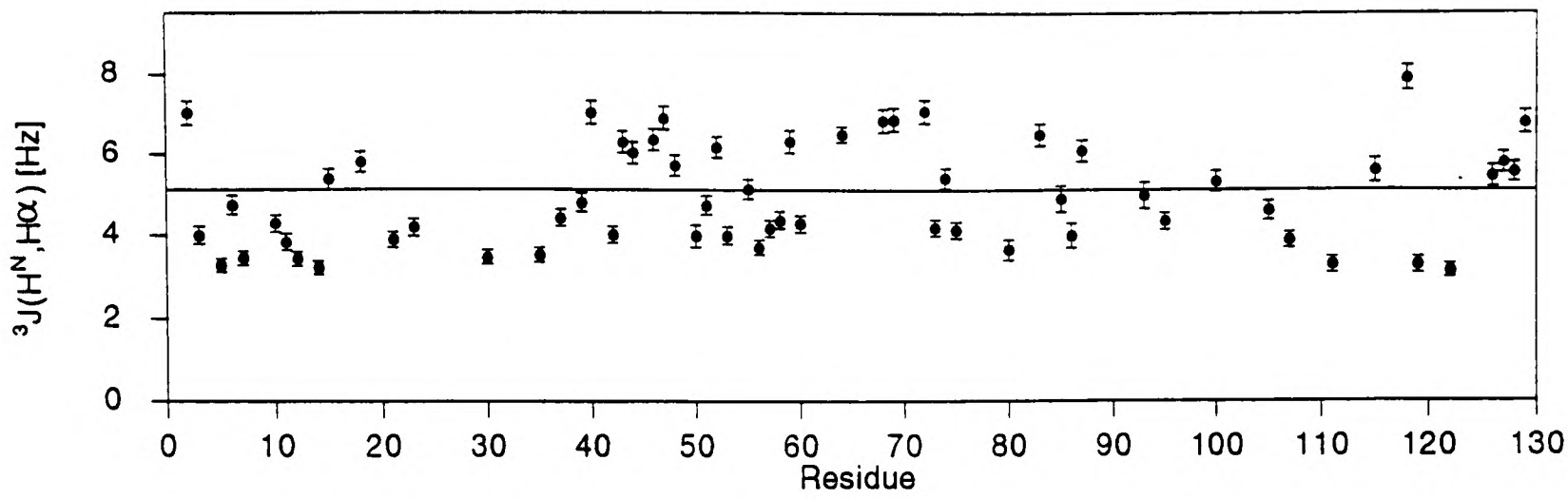


Figure 6

58 $^3J(H^N,H_\alpha)$ coupling constants could be measured reliably in HMQC-J (Kay et al., 1989) and 3D-HNHA (Vuister et al., 1993) experiments. The agreement between values measured in both experiments was within 0.5 Hz. Coupling constants with values below 5 Hz suggest that the conformational preferences of the residue lie entirely in the helical region of ϕ space.

Figure 7 (see following page)

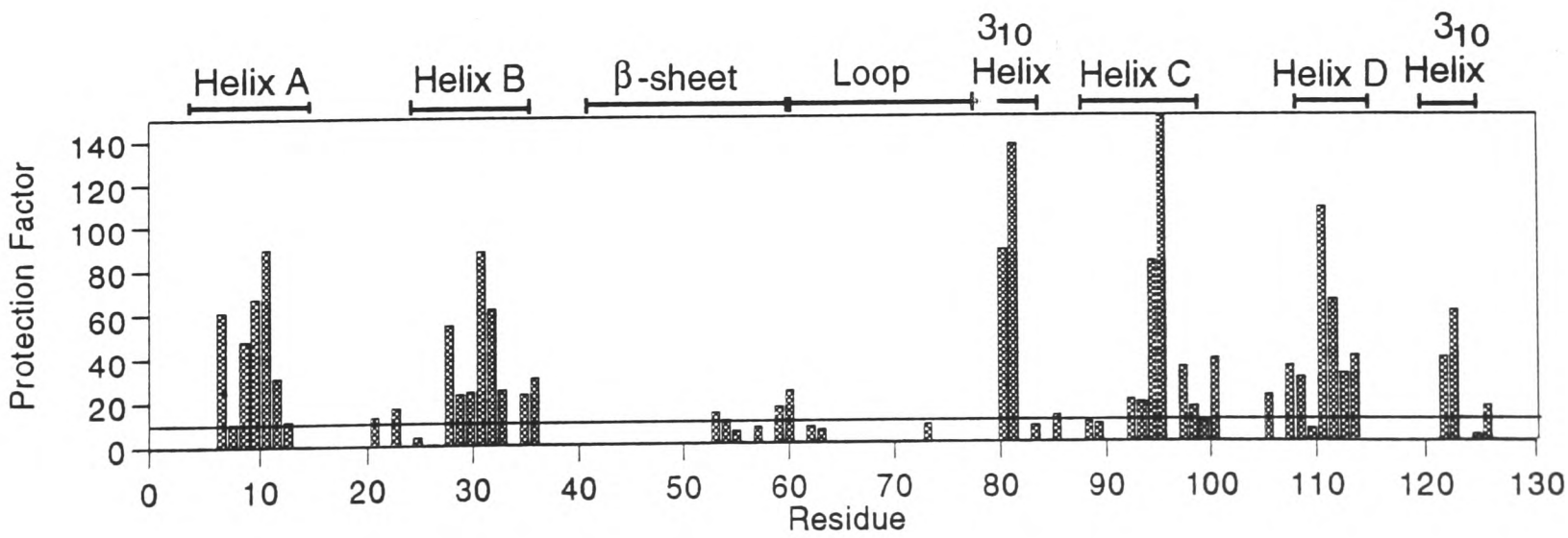


Figure 8

Protection factors (uncorrected for TFE effects) of amides against solvent exchange in the TFE state. Regions of secondary structure in the native state are shown.

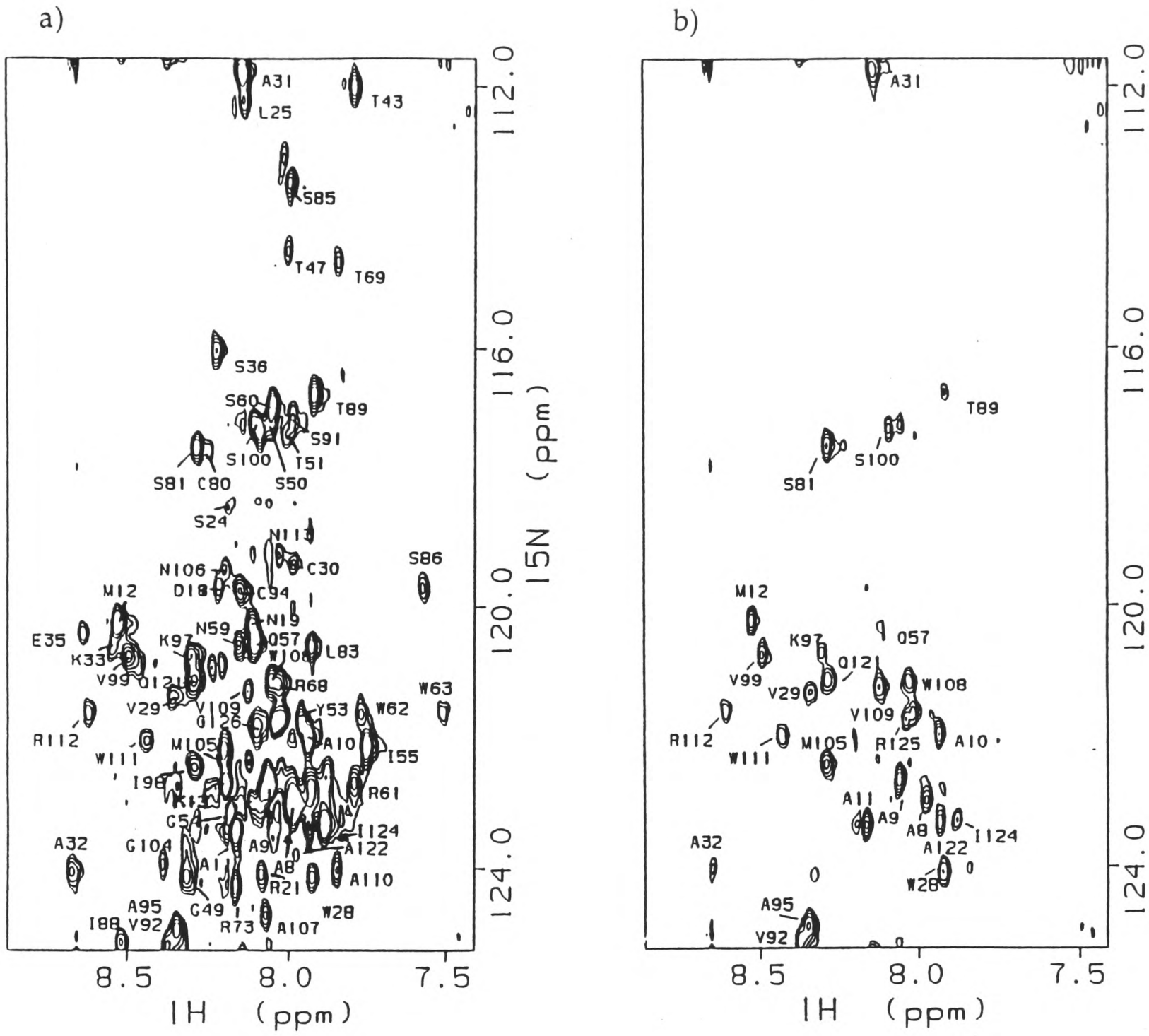


Figure 7

^{15}N-^{1}H HSQC spectra recorded a) 10 and b) 100 mins after dissolution of the protein in 70%/30% TFE-d_2-OD/D_2O (v/v), pH 2.0, 37°C. Amides with significant residual proton occupancy are labelled.

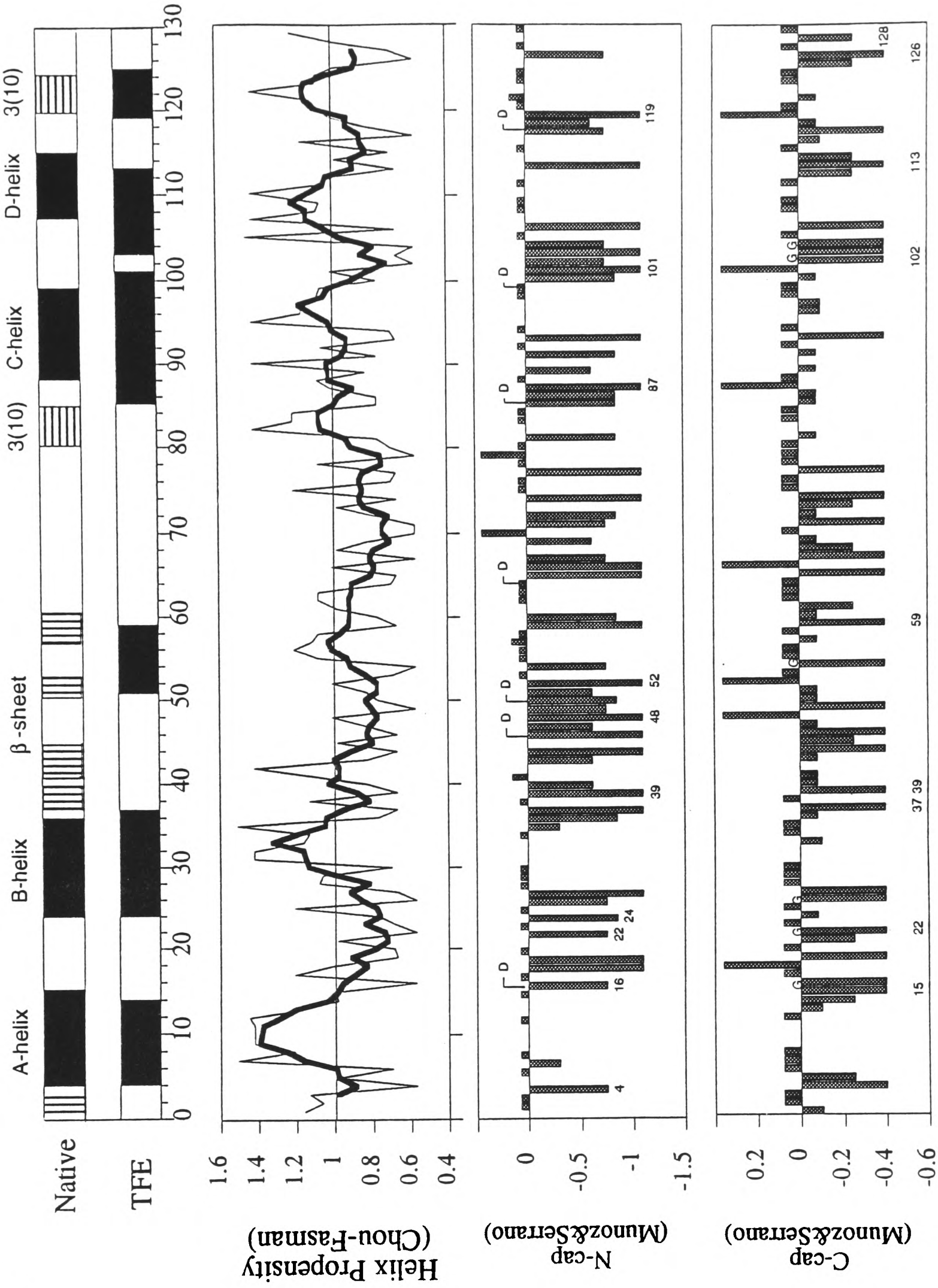


Figure 9

Secondary structural units are indicated as defined in a) the crystal structure (Handoll, 1985) b) in the TFE denatured state of hen lysozyme (Table 2A). c) Helix propensity derived using the scale of Chou & Fasman (1978), shown as individual data points and averaged over a sliding window of 5 residues. d) and e) show the N and C capping potential, respectively, of amino acids according to the scale of Munoz & Serrano (1994).

```
          1               15              30              45              60
sequence: KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRT
Native  : BBBHHHHHHHHHHHH         HHHHHHHHHHHHHHBBBBBBBBB     BBB    BBB

FUGUE   :    AAAAAAAAAAACG       PCGB                BBPB      BBPBBC BBBBCCGCP
GOR     : HBH HHHHHHHHHHH        B   BBBH HHH    H  B       B BBBBBBB BBBB
ChouFasman     HHHHHHHHH             BBBBHHHHHH BBBBB       BBBBBBBBBBBBB
AGADIR  :    HHHHHHHHHHHH            HHHHH
ALB     :       HHHHHHHHH          &&&&&&HHH&&&          TTTT  TTBBB&&&&&&  TTT

TFE state    HHHHHHHHHH          HHHHHHHHHHHHHH              HHHHHHH


            75              90              105                     129
sequence:GSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDVQAWIRGCRL
Native  :         GGGGG    HHHHHHHHHHH        HHHHHHHH    GGGGG

FUGUE   :C                          BBBBBC    AAAAAAAA
GOR     :             B      H B    BBB     H HBBHH          BBBBB BB
ChouFasman  BBBBBB  HHHH HBBBBBBBBHHHHH      BBBBBB       BBBBBBBBBBB
AGADIR  :                          HH         HHHHHHHH         HH
ALB     :TT     &&&&HHHHH&&&&&&&&&&&&BB T TT   H&&&HHHH&TTTT &&&&&&

TFE state              HHHHHHHHHHHHHHH   H HHHHHHHHH     HHHHHH
```

B = β-strand
H = α-Helix
G = 3_{10} Helix

Figure 10

Secondary structural units are indicated as defined a) in the crystal structure (Handoll, 1985) and in the TFE denatured state of hen lysozyme (Table 2A). Secondary structural predictions for hen lysozyme using programmes AGADIR (Munoz & Serrano, 1994), ALB (Finkelstein et al., 1991), GOR (Gibrat et al., 1987), Fugue (Rooman et al., 1992) and the method of Chou & Fasman, (1978). The programmes were used as follows: AGADIR: H denotes that the predicted helix population is greater 5% at 1°C, pH 2.0. ALB: T denotes turns or bends are probalby at these positions H denotes helix predicted with certainty; & denotes helix is probable; all for unfolded chain in water at pH 2.0, 50mM ionic strength at 20°C. GOR and FUGUE used standard input parameters, however, lysozyme and homologues were excluded in the derivation of parameter sets. Symbols in the seven state assignment of FUGUE (A and C), (B and P) and (G, E and O) are equivalent to H- (helix), B- (beta) and C (coil) in the others.

Residue		Chemical Shifts (ppm)				
		^{15}N	NH	Hα	Hβ	sidechains, including ^{15}N
Lys	1				1.96	γCH_2 1.49
Val	2	124.14	8.38	4.17	2.10	γCH_3 0.99, 0.97
Phe	3	125.37	8.06	4.75	3.12, 3.25	δCH 7.18; εCH 7.36; ζCH 7.44
Gly	4	111.90	8.08	4.09, 4.11		
Arg	5	122.69	8.24	4.23	1.93	γCH_2 1.75; δCH_2 3.21
Cys	6	119.30	8.42	4.65	2.93, 3.24	
Glu	7	123.08	8.06	4.24	2.20, 2.56	
Leu	8	123.01	7.93	4.24	1.74	γCH 1.60; δCH_3 0.95
Ala	9	122.76	8.01	4.15	1.59	
Ala	10	122.02	7.89	4.14	1.59	
Ala	11	123.30	8.11	4.22	1.61	
Met	12	120.08	8.45	4.25	2.18, 2.31	γCH_2 2.59; εCH_3 2.15
Lys	13	122.68	8.13	4.24	2.02	γCH_2 1.53; δCH_2 1.73
Arg	14	120.98	8.15	4.23	1.96	γCH_2 1.71; δCH_2 3.24
His	15	117.60	8.03	4.71	3.26, 3.56	δCH 8.66; εCH 7.46
Gly	16	110.09	8.12	4.03, 4.13		
Leu	17	123.02	7.98	4.27	1.71	δCH_3 0.96
Asp	18	119.73	8.17	4.73	2.93	
Asn	19	120.30	8.09	4.76	2.80, 2.89	
Tyr	20	122.59	8.02	4.53	3.14	δCH 7.17; εCH 6.90
Arg	21	124.00	8.11	4.22	1.65	γCH_2 1.84, 1.90
Gly	22	109.13	7.75	3.90, 3.95		
Tyr	23	122.65	7.88	4.59	3.12	δCH 7.16; εCH 6.90
Ser	24	118.48	8.15	4.45	4.08	
Leu	25	125.66	8.09	4.42	1.79	δCH_3 0.98
Gly	26	107.50	8.18	3.94		
Asn	27	121.35	7.97	4.62	2.84	δNH_2 6.55, 7.37 113.16
Trp	28	124.00	8.02	4.52	3.49, 3.62	NH1 9.65, 129.15; C2H 7.31; C4H 7.66; C5H 7.15; C6H 7.22; C7H 7.49
Val	29	121.21	8.23	3.63	2.15	γCH_3 1.07, 0.98
Cys	30	119.37	7.94	4.54	2.87, 3.34	
Ala	31	125.47	8.07	4.34	1.33	
Ala	32	123.83	8.53	4.10	1.33	
Lys	33	120.50	8.41	3.93	1.99	γCH_2 1.49; δCH_2 1.68
Phe	*34	122.20	8.05	4.34	3.02, 3.29	δCH 7.00 εCH 7.12; ζCH 7.21
Glu	35	120.48	8.50	4.08	2.23	
Ser	36	116.08	8.14	4.34	4.06	
Asn	37	122.20	7.85	4.56	2.49, 2.60	δNH_2 6.55, 7.05 113.49
Phe	38	122.07	8.10	4.34	2.80	δCH 7.19; εCH 7.29; ζCH 7.29
Asn	39	119.97	8.14	4.68	2.90	δNH_2 6.65, 7.41
Thr	40	116.56	8.02	4.38	4.23	γCH_3 1.30
Gln	41	123.07	8.10	4.40	2.10, 2.19	γCH_2 2.45; εNH_2 6.52, 7.28 112.43
Ala	42	124.50	8.10	4.26	1.42	
Thr	43	111.92	7.76	4.39	4.29	γCH_3 1.29
Asn	44	122.50	8.02	4.78	2.90	δNH_2 6.66, 7.19

Table 1

^{15}N and ^{1}H Assignments made in this study for the TFE denatured state in 70% TFE/H_2O (v/v), pH 2.0 and 37°C. Assignments for residues labelled (*) are only available at 47°C.

Arg	45	122.64	8.10	4.35	1.89, 1.99	γCH_2 1.74; δCH_2 3.26
Asn	46	120.22	8.28	4.83	2.90, 2.95	δNH_2 6.70, 7.47 113.35
Thr	47	114.72	7.99	4.43	4.31	γCH_3 1.27
Asp	48	122.45	8.31	4.80	3.02	
Gly	49	110.82	8.29	4.06		
Ser	50	117.18	8.03	4.52	4.28	
Thr	51	117.39	8.00	4.33	4.04	γCH_3 1.26
Asp	52	122.41	8.19	4.72	2.92	
Tyr	53	121.91	7.93	4.42	3.10	δCH 7.16; εCH 6.91
Gly	54	109.73	8.16	3.91, 3.99		
Ile	55	122.03	7.71	4.01	2.00	γCH_2 1.30; $\gamma CH_3/\delta CH_3$ 0.98
Leu	56	123.40	7.84	4.24	1.77, 1.86	γH 1.65; δCH_3 0.97
Gln	57	120.55	8.05	4.17	2.17	δCH_2 2.35; $\varepsilon NH2$ 6.93, 6.54 111.78
Ile	58	121.71	7.97	3.99	1.94	γCH_2 1.68, 1.26; $\gamma CH_3/\delta CH_3$ 0.96
Asn	59	120.40	8.09	4.74	2.84, 2.86	δNH_2 6.55, 7.33 112.33
Ser	60	116.98	8.02	4.26	3.98	
Arg	61	122.81	7.78	4.10	1.63, 1.70	$\gamma CH_2$1.50; δCH_2 3.10
Trp	62	121.72	7.72	4.66	3.32	NH1 9.92, 130.41; C2H 7.23; C4H 7.68; C5H 7.13; C6H 7.23; C7H 7.49
Trp	63	121.66	7.48	4.65	3.09	NH1 9.81, 130.67; C2H 7.25; C4H 7.29; C5H 7.05; C6H 7.15; C7H 7.51
Cys	64	118.99	7.81	4.66	2.96, 3.29	
Asn	*65	121.10	8.31	4.80	2.81, 2.95	δNH_2 6.81, 7.48 113.86
Asp	66	121.27	8.37	4.78	2.96	
Gly	67	110.25	8.30	3.98		
Arg	68	121.20	7.91	4.46	2.00, 1.86	γCH_2 1.69; δCH_2 3.25
Thr	69	114.96	7.83	4.76	4.34	γCH_3 1.29
Pro	70			4.43	2.15	γCH_2 2.00
Gly	71	110.60	8.32	3.93, 4.12		
Ser	72	117.11	7.93	4.50	4.06	
Arg	73	124.16	8.19	4.59	1.89, 1.75	γCH_2 1.64; δCH_2 3.20
Asn	74	120.68	8.23	4.84	2.93	δNH_2 6.72, 7.52 113.14
Leu	*75	123.24	7.90	4.30	1.67	γH 1.67; δCH_3 0.92
Cys	*76	116.70	7.99	4.78	3.08, 3.42	
Asn	*77	121.35	8.24	4.65	2.79, 3.01	δNH_2 6.63, 7.33 112.86
Ile	*78	122.80	7.83	4.15	1.88	γCH_2 1.56, 1.11; $\gamma CH_3/\delta CH_3$ 1.01
Pro	79			4.31	2.02, 2.24	γCH_2 2.02; δCH_2 3.83
Cys	80	117.53	8.25	4.02	3.03, 3.15	
Ser	81	117.57	8.27	4.60	3.83	
Ala	82	125.94	8.23	4.37	1.50	
Leu	83	120.65	7.89	4.45	1.74	δCH_3 0.99
Leu	84	126.48	7.98	4.32	1.75	δCH_3 0.98
Ser	85	113.68	7.98	4.26	4.02	
Ser	86	119.42	7.58	4.49	4.08	
Asp	87	122.91	8.00	4.75	3.06, 3.13	
Ile	88	125.01	8.44	3.93	1.98	γCH_2 1.72, 1.24; $\gamma CH_3/\delta CH_3$ 0.92
Thr	89	116.60	7.89	4.32	3.96	γCH_3 1.33
Ala	*90	125.50	8.09	3.99	1.35	
Ser	91	116.80	7.93	4.28	3.83	
Val	92	125.10	8.32	3.99	2.21	γCH_3 1.02

Asn	93	120.30	8.02	4.68	2.87	
Cys	94	119.64	8.12	4.55	2.80, 2.96	
Ala	95	124.95	8.31	4.27	1.57	
Lys	*96	121.70	8.03	4.17		εCH_2 2.93
Lys	97	120.68	8.25	4.12	1.78, 1.96	δCH_2 1.59; εCH_2 3.00
Ile	98	122.36	8.22	3.87	2.10	γCH_2 1.56; $\gamma CH_3/\delta CH_3$ 0.97
Val	99	120.70	8.41	3.92	2.24	γCH_3 1.03, 1.12
Ser	100	117.17	8.06	4.44	4.14	
Asp	101	122.45	8.26	4.85	3.11	
Gly	102	110.60	8.28	4.07		
Asn	103	120.50	8.16	4.83	2.97	δNH_2 6.66, 7.46 112.82
Gly	104	110.50	8.36	4.06		
Met	105	122.18	8.17	4.56	2.22	γCH_2 2.68; εCH_3 2.11
Asn	106	119.53	8.18	4.62	2.93	δNH_2 6,74, 7.23
Ala	107	124.78	8.06	4.29	1.60	
Trp	108	121.16	7.99	4.50	3.50	NH1 9.71, 129.04; C2H 7.22; C4H 7.64; C5H 7.15; C6H 7.28; C7H 7.48
Val	109	121.28	8.07	3.47	2.20	γCH_3 0.99, 1.14
Ala	110	124.12	7.89	4.17	1.57	
Trp	111	121.97	8.35	4.30	3.47	NH1 9.72, 129.88; C2H 7.23; C4H 7.70; C5H 7.13; C6H 7.22; C7H 7.50
Arg	112	121.70	8.52	3.70	1.59, 1.69	γCH_2 1.47
Asn	113	119.17	8.03	4.53	2.82, 2.93	δNH_2 6.63, 7.41 113.06
Arg	114	122.80	7.95	4.19	1.87	γCH_2 1.58
Cys	115	120.50	8.37	4.40	2.83, 3.19	
Lys	116					
Gly	117	109.41	8.21	3.99		
Thr	118	116.64	7.93	4.34	4.24	γCH_3 1.28
Asp	119	123.32	8.14	4.69	3.00	
Val	120	122.90	8.03	3.84	1.94	γCH_3 1.04
Gln	121	120.92	8.13	4.09	2.18	δCH_2 2.45, 2.54; εNH_2 6.62, 7.12 111.70
Ala	122	123.03	7.90	4.16	1.53	
Trp	123	121.26	7.93	4.51	3.48	NH1 9.77, 129.90; C2H 7.25; C4H 7.62; C5H 7.15; C6H 7.27; C7H 7.48
Ile	124	123.13	7.88	3.86	2.00	$\gamma CH_3/\delta CH_3$ 0.93, 1.03
Arg	125	121.80	7.98	4.18	1.87	γCH_2 1.71; δCH_2 3.24
Gly	126	108.64	8.08	3.96, 4.05		
Cys	127	119.77	8.04	4.78	2.92, 3.16	
Arg	128	122.91	7.99	4.42	1.87, 1.96	δCH_2 3.21
Leu	129	124.36	7.84	4.43	1.73	δCH_3 0.94

Table 2A

Designation	Native State	TFE- state HX (Nterm. +4)	NMR parameters ΔHα	^{3}JNα	NOE (overlap)	^{15}N relax	Consensus (X-PLOR)
A-helix	4-15	7-13	5-14	3-14	3-8 (7-14)	6-12	4-14
B-helix	24-36	28-36	27-38	20 ? -39	19-33 (31-38)	23-38	24-37
C-helix	89-99	92-100	90-99	(90->	90-102	90-101?	+89-99
D-helix	108-115	105-113	107-115	<-111)	102-109 (111-115)	105-114	103-113
3_{10} (i)	80-84	*85?*	84-89	(85,86)	84-89	80-89?	85-89+
3_{10} (ii)	120-124	121-125	120-125	119,121	117-128	117-124	119-125
β2,t2,	50-53, 54-57,	*53-60*	53-61	50-58	50-58	53-61	50-58
β3	58-60						

Table 2A

Limits and nature of secondary structure as defined by NMR parameters in the TFE state and secondary structure defined in the native state crystal structure. The following criteria were used for the inferral of helical structure: Hydrogen exchange: 3 out of 4 consecutive sites with protection factors exceeding 10; ΔHα chemical shift perturbations, using criteria of Wishart et al., (1992); ^{3}JNα; regions uninterrupted by residues with coulping constants > 5 Hz; NOE pattern: regions with several αN(i,i+3) or αN(i,i+4) (or evidence for resonance overlap); ^{15}N relaxation: regions with S^2 > 0.80 (see Chapter 8); X-PLOR consensus: ϕ, ψ angles in helical region for majority of structures in ensemble derived from NMR restraints (Appendix II).

Table 2B

Classsification of secondary structure as defined by NMR parameters in the TFE state (see main text).

Table 2B

Native-like Helix	Extension and non-native Helix	Unfolded and other
A: 4-14		N-term: 1-3 15-19
	20-23?	
B: 24-36	37-39	40-49
	50-59	60-84
C: 89-99	85-88	100-103?
D: 108-113	103-107	114-116
3_{10}(ii): 120-124	125	126-129

Chapter 8

A Heteronuclear NMR Study of a Partially Ordered State: Mainchain Dynamics of Hen Egg White Lysozyme in Trifluoroethanol.

Summary

^{15}N NMR relaxation measurements have been used to characterise the dynamic behaviour of the mainchain of hen egg white lysozyme in a partially ordered state, which is formed by the denaturation of the protein in presence of trifluoroethanol cosolvent (TFE). The NMR relaxation parameters were interpreted by a quasi-mapping of the spectral density function and by order parameters (S^2) derived according to the formalism of Lipari & Szabo. The observed variation in mainchain dynamics suggests that the structure in regions of the polymer chain has widely different levels of persistency. The polypeptide chain is involved in extensive motions ($S^2 < 0.6$, with internal correlation times of the order of 1 ns) at the N- and C-termini and in the regions corresponding to a long loop and to the N-terminal half of the β-sheet in the native state. TFE has been found to stabilise non-native α-helical structure in the denatured protein as extensions to native-like helices and in a region corresponding to C-terminal part of the β-sheet in the native fold (Chapter 8). However, the majority of these structures possess considerable flexibility ($0.6 < S^2 < 0.8$, with internal correlation times ranging from 200 ps to 1 ns). By contrast, several regions of the polypeptide chain are highly restricted in their motion ($S^2 > 0.8$, with internal correlation times < 200 ps) but are also found to experience conformational exchange on a millisecond timescale. Their location closely corresponds to the regions which are α-helical in the native state. Overall the mobility of the mainchain in TFE, although increased in magnitude, is remarkably similar to that seen in the native state. Local features of the polypeptide chain, including propensities of residues for helix formation, sidechain hydrophobicity and the position of the disulphide bonds could be used to rationalise the wide variation of the dynamical behaviour of the polypeptide chain in this partially ordered state.

Introduction

There is now considerable evidence for residual structural preferences which persist in denatured states of proteins and in polypeptides in aqueous solution, ranging in character from local clusters of sidechains to highly ordered subdomains (Dill & Shortle, 1991; Dyson & Wright, 1991; Ptitsyn, 1992; Dobson, 1994). The extent and order of the structures detected in these states, intermediate between the natively folded conformation and that anticipated for a fully unfolded chain, have become central to many investigations of protein folding (Kim & Baldwin, 1990; Creighton, 1990). The characterisation of these structures, which are presumed to be less extensive than those found in the native state of proteins, and the interactions which give rise to their stability are likely to provide important insights into the interactions which are responsible for their formation and their possible roles in the folding process (Fersht, 1993; Kemmink & Creighton, 1993).

A principal characteristic of the conformations detected in denatured or partially ordered proteins, as well as peptide fragments, is their often highly reduced stability compared to the structures found in native proteins. As a consequence, these structures are usually in equilibria with less ordered states or alternative conformations on a time-scale which is fast compared to that sampled by NMR. Thus the common NMR parameters, such as chemical shifts, coupling constants and NOE intensities are averaged over an ensemble of conformations which are often widely different in character. Whilst these parameters provide an indication of the nature of the more ordered structures and their relative populations, they can not usually be used directly to assess the time-scales and the extents of motions involved.

Recently developed heteronuclear NMR techniques have made it possible to probe with high sensitivity and site specificity the dynamics of N-H and C-H bond vectors in proteins by the relaxation behaviour of the ^{15}N and ^{13}C nuclei. Order parameters (S^2) derived from this data in conjunction with the model free approach of Lipari & Szabo (1982) yield a measure of the geometrical restriction of the bond vectors on a time scale of picoseconds to nanoseconds, which is faster than the stochastic rotational tumbling of the molecule. ^{15}N and ^{13}C relaxation measurements have recently become an attractive means of studying mainchain dynamics of proteins in solution (Kay et al., 1989; Clore et al., 1990; Peng & Wagner, 1992), but can also be used to probe the dynamic behaviour of the sidechains of certain amino acids (Stone et al., 1992; Palmer et al., 1993 and Chapter 6). Several conclusions have emerged from these

investigations. ^{15}N mainchain relaxation studies on native proteins have suggested, for example, that backbone motions are usually highly constrained on the picosecond to nanosecond timescale, while mobility of the residues located in loops and near chain termini of the protein can be considerable. The analysis of order parameters in the context of structural features of proteins implies that hydrogen bonding and burial from the protein surface are likely to be the principle determinants of the mobility of the N-H groups (e.g. Schneider et al., 1992; Kay et al., 1989 and see Chapter 6).

In contrast to the intensive work carried out on native states, there have been few studies of the NMR relaxation behaviour in denatured and partially ordered proteins (van Mierlo et al., 1993; Orekhov et al., 1994; Redfield et al., 1994; Logan et al., 1994; Alexandrescu & Shortle, 1994b). This is because the principal problems associated with the study of denatured proteins by NMR have only recently been overcome by the development of heteronuclear experiments, allowing increased resolution and further connectivity of ^{1}H NMR signals by their correlation with those of ^{15}N and ^{13}C nuclei in 2D and 3D experiments. Thus, the site-specific characterisation of the conformational behaviour of denatured proteins is now, in favourable cases, possible to the extent common for native proteins or peptides (for example see studies by Neri et al., 1992; van Mierlo et al., 1993; Logan et al., 1994; Alexandrescu et al., 1994a; Arcus et al., 1994 and Chapter 7). Studies on the dynamics of the polypeptide backbone provide valuable complementary information to that provided by other NMR parameters.

Recently we have examined a state of hen lysozyme which is formed by denaturation in trifluoroethanol (TFE) as a model for understanding the nature of partially ordered conformations of proteins. Standard 2D and 3D heteronuclear NMR experiments have allowed nearly complete resonance assignment for a ^{15}N labelled sample in presence of 70% TFE (v/v) at pH 2.0, 37°C. This has made possible a detailed characterisation of the secondary structure which is present in this state of the protein by consideration of the pattern of medium range NOEs, coupling constants, Hα chemical shift perturbations and amide hydrogen exchange rates. The results of this study are presented in Chapter 7 and can be summarised as follows: All residues, which are in α-helical structure in the native state are also involved in α-helical structure in presence of TFE. Significant protection against hydrogen exchange and low values for $^{3}J(H^{N},H_{\alpha})$ coupling constants imply that the centres of these native-like helices have a high degree of structural persistency. Extension of some the native helices by up to one turn at either end is inferred from the

NMR data in several regions of the polypeptide chain. In addition a region which forms part of the β-sheet in the native state, is also stabilised in non-native α-helical conformation by TFE. There are however, several regions which are partially organised in structures, other than those corresponding to helical conformations and regions for which the NMR data indicated no well defined structural preferences. Examples for the latter are the sequences corresponding to part of the main β-sheet and the long loop in the native protein. (These results are summarised schematically by a pictorial representaion of the structure in the TFE denatured state of hen lyoszyme in Figure 7 of Appendix II.)

This chapter describes the analysis of the dynamic behaviour of the polypeptide as part of a further investigation of the mainchain conformation in a partially ordered state of hen lysozyme formed in presence of TFE. The relationship between the conformational preferences which have been detected in this state of the protein (and described in Chapter 7) and their dynamic behaviour is examined. Comparison is made with the results obtained in a similar study for the native protein (Chapter 6). Moreover, possible determinants for the dynamic behaviour are investigated in the context of features of the primary structure of the protein and the position of the native disulphide bonds.

Materials and Methods

^{15}N NMR relaxation measurements and analysis

Hen egg white lysozyme was expressed in *Aspergillus Niger* using $^{15}NH_4Cl$ as the sole nitrogen source and purified from filtered culture medium (Roberts et al., 1992). NMR samples were prepared to contain ca. 3mM protein in 70% TFE-d2-OH/30% H_2O (v/v) and the pH was adjusted to pH 2.0 with phosphoric acid buffer. NMR experiments were performed at 37°C and 47°C on two homebuilt NMR spectrometers belonging to the Oxford Centre for Molecular Sciences with ^1H operating frequencies of 600.2 and 750.2 MHz. Pulse sequences for the measurement of the longitudinal (T_1) and transverse (T_2) relaxation time and the heteronuclear nuclear Overhauser effect (NOE) of ^{15}N nuclei in NH groups have been described previously (Kay et al., 1989, Boyd et al., 1990, Palmer et al., 1992). Suppression of the solvent resonance was achieved by low power presaturation or by application of trim pulses (Messerle et al., 1989). The T_1 and T_2 relaxation measurements used a series of 10 and 9 experiments with relaxation delays ranging from 20 to 1400 ms and 17.6 to 351 ms respectively. A recycle delay of 4.0 s was used for the T_1 relaxation and NOE experiments and

3.1 s for the T_2 experiments. A CPMG duty cycle delay of 0.5 ms was used in the T_2 experiments. The data sets were acquired using 128 complex data points in t1 with resolution of 9.5 Hz/point both at 600 MHz and for the heteronuclear NOE experiment at 750 MHz. The 2D spectra were zero filled once in F2 and twice in F1 and were processed with Felix (Hare Research and Biosym, Inc.). The heteronuclear NOE effect was calculated as the ratio of peak heights in spectra recorded with and without ^1H saturation. Rates R_1 ($=1/T_1$, longitudinal), and R_2 ($=1/T_2$, transverse relaxation), were fitted as single exponential decays to peak height data. The errors in the primary intensity data were taken as the standard deviation of spectral noise and uncertainties in the fitted rates were estimated using published procedures (Palmer et al., 1991, Stone et al., 1992). The ^{15}N relaxation data were analysed in terms of a reduced expression of spectral density function and according to the approach of Lipari & Szabo (1982). Selected aspects of both of these approaches are described below. Parameters calculated using the appropriate spectral density function were optimised by minimisation of the rmsd between the calculated and the experimental relaxation data using procedures described previously (Clore et al., 1990; Palmer et al., 1991; Constantine et al., 1993). The order parameters and their uncertainties were estimated by Monte Carlo sampling using "Modelfree3.1" (Arthur G. Palmer, Columbia University). (The theory of ^{15}N relaxation behaviour and data analysis is described in Appendix I).

Results

^{15}N Relaxation data

Relaxation data, R_1, R_2 and heteronuclear NOE, could be extracted for 97 of the 126 mainchain nitrogens from spectra measured at a ^{15}N spectrometer operating frequency of 60.8 MHz. A further heteronuclear NOE measurement was performed on a spectrometer with a ^{15}N operating frequency of 76.0 MHz. The data are plotted in Figure 1. The pattern seen in the data is similar at the two frequencies, with the offset between them varying as a consequence of the dynamic behaviour of the polypeptide chain and the different frequency contributions which are sampled (Lipari & Szabo, 1982; Redfield et al., 1992). The lower frequency data appears to be more sensitive to motions but the uncertainty in the data is greater. From figure 1, it is apparent that values for the NOE vary considerably in the TFE denatured state as a function of the polypeptide chain. Broad depressions in its value, for example below 0.4 at 60.8 MHz, suggest that considerable motions occur in several regions of the protein relative to others. These regions include at least two residues at each terminus

which have lower than average NOE and R_2 values, that rapidly decrease in magnitude as the termini are approached. Uncertainties in the relaxation data, especially for R_2 and NOE are particularly high for amides which give rise to weak ^{15}N-^{1}H correlations in the HSQC spectra. As noted from the other NMR parameters, these belong regions of the polypeptide chain close to 3 of the 4 disulphides of the protein. The variations in the R_2 data broadly mirror the heteronuclear NOE data as a function of sequence.

These data can be compared to the data recently obtained for the native state of hen lysozyme at 50.7 MHz (Chapter 6). In the native state the value for the heteronuclear NOE was found to be uniformly high with values greater or equal to 0.7, except at the C-terminus and for residues 85 and 101-104. It is interesting that the variation of the R_2 data along the sequence corresponds, at least in outline, to that seen in the native state when the R_2 values for the native state are multiplied by a factor of approximately 2. By contrast, the R_1 data, shows little correspondence with the native state profile. R_1 has been shortened considerably even when the difference in spectrometer field used in these two studies, is taken into account. Regions which have longer values of R_2 in the TFE state and higher than average NOEs, have typically shorter R_1 values. This is opposite to the trend observed for these parameters in the native state and suggests that the overall rotational correlation time in the TFE state is significantly longer than the 5.7 ns derived for the native molecule (see Figure 3 in Appendix I). The high viscosity of water/TFE mixtures, and a presumed increase in protein volume upon denaturation, are both likely to increase the correlation time significantly.

Quasi mapping of the spectral density function

The assumption of a single overall correlation time requires that the rotational tumbling of the entire molecule is isotropic. Data to test this assumption are generally not available for denatured proteins. Parameters can be devised which take into account significant deviation in protein shapes from that of a sphere, provided that a structure of the molecule and the orientations of NH vectors relative to the axes of symmetry are known (e.g. Barbato et al., 1992). The dimensions or the global topology of the TFE denatured state have not been established and therefore rotational tumbling has been assumed to be isotropic as a first approximation. Correlation times were optimised both globally and locally, to test indirectly this assumption. As a first step, however, the relaxation data were analysed using an approach which does not rely on an overall correlation time.

The three ^{15}N relaxation measurements R_1, R_2 and NOE are determined by the function spectral density function $J(\omega)$, as a combination of its value at five frequencies $J(0)$, $J(\omega_N)$, $J(\omega_H+\omega_H)$, $J(\omega_H)$, $J(\omega_H-\omega_N)$ (See Equations 1-3, Appendix I). It is not possible, given these 3 relaxation measurements to determine directly the value of the density function at these 5 frequencies. However, the spectral density function can be mapped at three of the of the frequencies, using a simplified relationship between the spectral density function and the relaxation parameters when the values of $J(\omega_H+\omega_N)$, $J(\omega_H)$, $J(\omega_H-\omega_N)$ are assumed to be approximately equal and R_{ex} makes negligible contributions to R_2 (Peng & Wagner, 1992). These assumptions are reasonable given that $\omega_H \sim 10^* \omega_N$ and $R_2 >> R_{ex}$, where Rex are exchange contributions not eliminated by CPMG pulse train in measurements of R_2. Explicit equations (Equations 4-6, Appendix I) can thus be used to derive the values for the spectral density function at the three frequencies 0, ω_N and ω_N using the values R_1, R_2 and NOE at a ^{15}N frequency of 60 MHz as input data. The calculations were repeated 500 times, mapping the standard deviations of the input data onto a normal distribution, in order to propagate the uncertainties.

The resulting average values and standard deviations, are shown in Figure 2 plotted versus the protein sequence. The results are found to reflect closely the variations in the input data. $J(0)$, $J(\omega_N)$ and $J(\omega_H)$ follow the pattern of R_2, R_1 and 1-NOE respectively, for the TFE as for the native state (data not shown). The relative magnitude of the values of the spectral density function at the three frequencies is indicative of the motional behaviour of NH vectors in the following manner (Peng & Wagner, 1992). In cases of amides which are motionally confined in persistent secondary structure, low frequency components, $J(0)$, make the most significant contributions, whereas those of $J(\omega_H)$ are negligible. As the NH vector becomes more mobile, higher frequency components contribute significantly to $J(\omega)$, and because the area under the spectral density function must remain constant, low frequency contributions are decreased, resulting in an overall flattening of the spectral density function (London, 1989; Peng & Wagner, 1992). This is evident from the data in Figure 2, when the value of $J(0)$ is compared with that of $J(\omega_H)$ for residues 40-50; which are in a region a region that appears to form little, if any ordered structure in the TFE denatured state. By contrast, the decay of the spectral density function is steeper for residues 25-35, consistent with their location in a well ordered region of the polypeptide chain. Relaxation of the backbone ^{15}N nuclei of these residues is accomplished primarily through low frequency contributions to the density function. In the extreme, such relaxation would occur entirely through the overall rotational tumbling of the molecule in solution, as all internal

motions of NH vectors are fixed (S^2=1), and the value of $J(\omega_H)$ would approximate 0. Interestingly, the lowest values of $J(\omega_H)$ at 600 MHz for NH vectors in the TFE denatured protein (around $5*10^{-12}$ $rad^{-1}s$) are similar to those of the sites in the native state which are most restricted in their mobility, with values for $J(\omega_H)$ at 500 MHz of around $7*10^{-12}$ $rad^{-1}s$. This suggests that at the mainchain motion of at least some of the residues is highly restricted in TFE.

Interpretation of the relaxation data using the modelfree approach of Lipari & Szabo: Modelling of the spectral density as an explicit decay function involving one or several adjustable parameters (Lipari & Szabo, 1982; Clore et al., 1990) has become a popular approach to the analysis of ^{15}N relaxation data. The simplest model for a denatured state is one which treats each NH vector as a rigid rotor with motions completely independent of the remainder of the polypeptide (Alexandrescu & Shortle, 1994a). In the case of this "free flight" particle model, a local correlation time is derived for the reorientation of the NH vector for each residue. Generally, the model was found to fit the experimental data very poorly. Interestingly, the values derived for the local correlation time are in the range 8-14 ns for all but a few residues. By contrast correlation times anticipated for this "free flight" particle model (Allerhand et al., 1971; Torchia et al., 1975; Alexandrescu & Shortle, 1994a), are in the range of 0.5-2 ns and are observed only at the termini and for residues 47-50. It follows that the effect of higher order structures must be taken into account for the motional behaviour of the majority of residues of the polypeptide chain.

Another extreme model, which considers the protein as a fixed rotor with no internal motions is equally inappropriate, except possibly in the case that the protein shape is highly elongated resulting in global motions being highly anisotropic. However, in this case the regional variation of relaxation parameters would not be smooth as a function of sequence but would depend on, in case of a helix, the periodic variation of angles of the NH vector with the long axis of the molecule. Furthermore, such variation is unlikely to be entirely responsible for the good correspondence which exists in the majority of the regions of the molecule between the magnitude of mobility inferred from the relaxation data and that suggested from the other NMR data (see below). Therefore, a fast internal motion must be modelled in the presence of the slower tumbling motion which affects the polypeptide chain either as a whole or as segments. In the following analysis the global rotational motion of the molecule was assumed to be isotropic and is characterised by a single overall correlation time. The extent of internal motion is on the simplest

model described by an order parameter S^2 (equation 12 in Appendix I), assuming that the effective correlation time for these internal motions is short (< 20 ps). This model predicts an NOE value of greater than 0.80 at 60.8 MHz ^{15}N frequency, for example, and is clearly incompatible with the experimental data for all but a few residues. A model which allows for internal motions of lower frequency, described by an internal correlation time, τ_e, and of greater amplitude was found to fit adequately the relaxation data for 44 of the residues. An overall correlation time of 10.0 ± 0.1 ns was optimised in the process of fitting the relaxation data for all the residues simultaneously. Large discrepancies between the experimental data and the relaxation parameters which are predicted on the basis of this model, however, remained for the majority of residues. There were three types of discrepancy: i) data for residues with experimental R_2/R_1 ratios above 12 could not be matched in R_2 by the model; ii) residues with predicted NOE values lower than the experimental values; iii) residues for which the predicted frequency dependence of the NOE was opposite to that found experimentally. These remaining discrepancies implied that a significant proportion of the data could not be modelled adequately by a spectral density function with a single internal motion.

A second internal motion, introduced to model motions on the μ-ms timescale contributes as an exchange term (R_{ex}, equation 2 in Appendix I) to measured R_2 values and was used for 36 residues having a R_2/R_1 ratio greater than 12, given a global correlation time in the region of 10 ns. By contrast, fits for 31 residues with an R_2/R_1 ratio of less than 10 were improved by inclusion of an additional component of fast motion, described by a second order parameter S^2_{fast} ranging from 0.61 to 0.93 for the present data (Clore et al., 1990; equation 5 in Appendix I). Inclusion of this motion yielded the correct frequency dependence of the NOE for a number of residues (see iii, above). The implementation of these more complicated models did not appreciably change the value of the global correlation time on re-optimisation, but resulted in a better agreement with the experimental data as measured by the statistical significance of the residual discrepancy (Table 1). Adequate fits to the data could not be obtained for 10% of the residues even with these extended models. The majority of these residues are located in regions of the polypeptide which are presumed to undergo the most extensive motions and it is likely that even more complicated models are required to describe their motion. For example, the assumption that a second internal motion is either slow, in the μ-ms range, or has a fast effective correlation time (< 20 ps) may not be valid here (Clore et al., 1990). However, in order to model this second internal motion explicitly further experimental data are required, which could in principle be obtained by

sampling of the dynamics of the TFE denatured state at a number of different spectrometer frequencies. From the present analysis it is apparent, however, that the formalism of Lipari & Szabo (1982) with the extension of Clore et al. (1990) adequately modelled the great majority of the experimental data and that this appeared to be possible with one single optimised overall correlation time of 10.7 ± 0.1 ns.

Optimisation of local rotational correlation times

In order to test the dependence of the derived parameters on the assumption of a single overall rotational correlation time, the polypeptide chain was treated as a series of segments (Barbato et al, 1992). These segments were taken to correspond to regions which have a near common value for S^2 and relate to the underlying structural units, previously identified in the TFE denatured state by NMR (see Chapter 7). Slowed hydrogen exchange in the helical regions implies, for example, that residues are not only highly restricted in their motions, but also that a large component of their motions are likely to be common or concerted as part of the structural unit. The rotational correlation times were then optimised regionally by fitting the relaxation data for each of these segments separately. The resulting correlation times and the order parameters, shown in table Table 1, are in good agreement with those derived with an optimised rotational correlation time of 10.7 ± 0.1 ns for the whole molecule. Segments corresponding to the A-helix and the C-terminal 3_{10}-helix in the native state, as well as the non-native α-helical structure, promoted by TFE in the region corresponding to part of the β-sheet in the native state, have rotational correlation times which are close to this global estimate. Interestingly, the value for the effective internal correlation times, τ_e, is decreased for the region corresponding to the A-helix, residues 6-12, to values approaching those found in other regions of helical character (see below). A number of segments have larger regional rotational correlation times with values of near 12.3 ns each. However, changes in the values of τ_e and R_{ex} were small when the data were fitted regionally rather than globally (Table 1). Values for τ_e are increased in regions with a local rotational correlation time greater than the global estimate and decreased otherwise. The opposite is true for exchange contributions, which are generally decreased in regions for which local rotational correlation times greater than that of the global estimate are found. An increase of 2 ns in τ_R is associated with a decrease of around 1.3 s^{-1} in R_{ex}. Interestingly, when the data were fitted locally without allowance for such an exchange contribution (Table 1), the resulting values for τ_R, S^2 and τ_e were changed only marginally, suggesting the inclusion of this term has no significant effect of the fitting of parameters. However, fitting the data with a

global correlation time fixed ot 13.2 ns, corresponding to the value derived in absence of possible exchange contributions, resulted in poorer fits for all sites, except for those located in the sections corresponding to the B, C and D-helices. The pattern and magnitude of the order parameters are hardly affected by these differences in τ_R, suggesting that the assumption of a global effective correlation time appears to be a reasonable one or that effects due to anisotropic rotational motion have only marginal effects on the relaxation data.

A plot of the fitted parameters (the order parameter S^2, the internal correlation time, τ_e, and the exchange term, R_{ex}), derived with a globally optimised correlation time of 10.7 ± 0.1 ns, are shown in Figure 3 and they are listed in Table 2. The value of the exchange term reflects its purpose in reducing R_2 to around 13 s^{-1}, the maximum values allowed for an overall correlation time of 10.7 ns.. However, the uncertainty in R_2 estimated on the basis of Monte Carlo sampling is large, up to 100% of its value, and it is therefore possible that none of the exchange terms exceed 5 s^{-1} in magnitude. It is noticeable that the sites which require an exchange contribution to R_2 are those associated with the broadest resonances in the HSQC spectrum and the majority of these residues are located in the region corresponding to the B and D helices, which are linked by a disulphide bond, Cys30-Cys115.

The data show an extensive variation of the value for the internal correlation time, τ_e, along the sequence. τ_e values of around 1000 ps are found in regions of residues 2-6, 14-20, 40-55 and for residues 64-74. These values reflect reorientations of the NH vectors which are often associated with motions of considerable amplitudes (i.e. having small values of S^2) on a time-scale much slower than the librational motions (20-200 ps), observed for amide vectors which are undergoing restricted motions. τ_e values typical of these restricted motions are found in the other regions of the protein for example, with few exceptions, in the entire C-terminal region of the polypeptide chain. Significantly, the majority of amides in regions with such relatively short values of τ_e also require an exchange contribution to model their relaxation.

The order parameter, S^2, is plotted in Figure 4a and is commonly interpreted as a measure of the geometrical restriction of the ^{15}N-^{1}H bond vector, assuming values between 1 and 0 for completely restricted internal and fully isotropic motion respectively. The lowest S^2 values of this study of 0.2-0.3 are observed in regions for which there is little evidence for ordered structure. By contrast other regions of the polypeptide have order parameters as high as 0.9. While

the variation in the pattern of S^2 values appears to be generally smooth there are several sites which have order parameters significantly reduced in value compared to those of their neighbours. It is noticeable that several of these, numbered in Figure 4a, are located at or adjacent to aromatic residues. This suggests that the dynamic processes associated with these aromatic rings, such as ring flipping, may provide contributions to relaxation of the ^{15}N nuclei, which are additional to relaxation mechanisms of ^{15}N nucleus considered in this study (London, 1989). Other "outliers", include residues close to disulphides, Ala95 and Cys115 as well as Gly102 and Gly104.

Remarkably, good correspondence exists between the variation of S^2 values and the other NMR parameters which have been used to characterise the structure and its persistency in the TFE denatured state of hen lysozyme (Table 3; Chapter 7). For example, regions which are defined from Hα chemical shift perturbation, hydrogen exchange, and NOEs to be in persistent α-helical conformation generally have order parameters greater than 0.8. (Table 3). Variation of S^2 towards the termini of helices and the dynamics in regions which are unstructured provide supporting evidence for their conformational assessment on the basis of the other NMR parameters.

Discussion

Dynamics of the mainchain in the TFE denatured state of hen lysozyme

The motion of maichain ^{15}N-^{1}H bond vectors has been characterised by parameters derived in a model independent manner according to the formalism of Lipari & Szabo (1982). While these models are only an approximation to the underlying spectrum of dynamic events, they provide a framework for the interpretation of the relaxation data. This enables comparisons to be made of the extent of motion of N-H bond vectors (order parameter S^2) and its time-scale (effective correlation time, τ_e) in different conformational states of proteins and in different proteins. Motions are evident on several timescales from the ^{15}N relaxation data for the mainchain in a partially ordered state of hen lysozyme formed in 70% TFE. The evidence for different classes of the motional behaviour and the regions of the polypeptide chain involved are summarised in the following sections and in Table 3.

Fast motions, described by internal correlation times, τ_e, of less than 200 ps are associated with small amplitude librational movements of NH vectors as part

of vibrational fluctuations typically within secondary structures. These motions are characterised by order parameters in the range of 0.8 to 1.0 and are commonly seen in the most structured regions of native folds (e.g. Clore et al., 1990; Redfield et al., 1992; Chapter 6). They are also evident in the most ordered regions in partially structured states (van Mierlo et al., 1993, Alexandrescu & Shortle, 1994a). In the case of the TFE state, these fast motions are observed in the centres of highly persistent native-like helixes; including the C-terminal 3_{10}-helix, and the A-helix, provided that a locally optimised correlation time is used for the latter. The detection of considerable protection of amide hydrogens against exchange with solvent in these regions, implies that the motions do not involve the breaking of hydrogen bonds.

Motions of increased amplitude, $0.6 < S^2 < 0.8$, are associated with increased values of the internal correlation time, typically in the range of 200-800 ps. This is consistent with the notion that fluctuations which explore a larger region of conformational space in a diffusive and random manner are accompanied by a reduced frequency for the reorientation of the NH vector. Such dynamics are inferred for regions at the termini of several helices and for residues 82-87 (Table 1B). Hydrogen bonds are likely to break and reform frequently as part of these helix fraying motions and therefore no significant protection against hydrogen exchange is observed at these sites. An exception to this is the region comprising residues 53 to 58, which has been shown to adopt a non-native α-helical conformation with marginal stability against amide hydrogen exchange, indicated by protection factors which are as high as 10 for several amides. However, it is possible that the fluctuations in this region may not necessarily involve the frequent exposure of NH groups to solvent, given that this region of the polypeptide chain is relatively hydrophobic and that less ordered conformations are likely to retain considerable residual clustering of sidechains.

Some of the most extensive and slowest motions occur in regions of the protein which, at least for the majority of the time, have been shown to be unstructured in presence of TFE cosolvent (Chapter 7). It is interesting to note that while the rate of reorientation, indicated by τ_e is maximally around 1000 ps for any of the N-H vectors, the amplitude of the motion, reflected in the value of S^2, appears to be closely related to the conformational space which is accessible, not just through motion of the residue itself but also through those of neighbouring residues. A model for this behaviour is proposed below. Similar values of τ_e and S^2 have been associated with the presence of long disordered loops in native proteins (e.g. Clore et al., 1990; Redfield et al., 1992)

but also appear to be characteristic for regions which are largely disordered in denatured or partially folded states (van Mierlo et al., 1993; Alexandrescu & Shortle, 1994).

The stochastic rotational tumbling of the molecule presents an upper limit for the internal motions which can in principle be modelled by the spectral density function (Lipari & Szabo, 1982). Nevertheless, internal motions have been found to affect the relaxation measurements in the great majority of NMR studies (e.g. Clore et al., 1990; Grasberger et al., 1993; Chapter 7). A manifestation of these motions, which can be attributed to result from chemical exchange processes on the μ-ms timescale, is the observation of increased values for R_2 in certain regions of the protein, with values over and above those which are compatible with a single rotational correlation time. However, these motions cannot be modelled explicitly at present, except that their contribution is removed by introduction of an exchange term to offset their effect on the R_2 data. The average value of this term, for 36 residues in the TFE state is 4.7 s^{-1}, compared to an average of 0.8 s^{-1} required for 32 residues in the native state. However, exchange contributions, similar in magnitude to those of the TFE state were observed for a number of sidechains in the native state of thioredoxin and hen lysozyme (Stone et al., 1992; Chapter 6) and for several mainchain sites in the denatured state of staphylococal nuclease (Alexandrescu & Shortle, 1994a).

These slow motions are of particular interest in relation to the study of partially ordered states of proteins. Assignment of NMR spectra and the characterisation of partially ordered proteins is made difficult because the exchange processes result in an increase of resonance linewidths and reduced resonance intensities (e.g. Baum et al., 1989; Alexandrescu et al., 1994b). It is notable that the resonances which are affected by such conformational exchange broadening in spectra of the TFE denatured state arise from parts of the protein in close proximity to three of the four disulphide bonds which persist in this denatured state. (Cys30-115, Cys64-80 and Cys76-94). This suggests that the conformational exchange could arise from slow reorientation of parts of the protein structure, for example as motions of the helices relative to each other. Indeed, similar resonance broadening close to disulphides has been reported in interleukin-4 and BPTI (Redfield et al., 1992; Szyperski et al., 1993). Furthermore, preliminary data suggests that resonance linewidths are decreased considerably upon removal of the disulphides, while the characteristics of the secondary structure remain similar (see Chapter 5).

Alternatively, it is possible that these slower motions are associated with helix unfolding transitions. In principle, this is not entirely inconsistent with the high degree of protection against amide hydrogen exchange which is observed in several of the regions, given that the exchange rates are determined by a combination of the fluctuations and pH in the EX_2 regime (Roder et al., 1985). However, the chemical exchange of protons itself is too slow to contribute to resonance broadening under these conditions. Other, fluctuations which involve amides may be responsible, provided that they result in states of significant population, the majority of which still protect amides against exchange. It has been suggested from molecular dynamics simulations, for example, that an α- to 3_{10} helix transitions may be one of the first steps in helix unfolding (Brooks, 1993). It is, at least in principle, possible to determine the magnitude of the exchange contribution more carefully and to derive a rate for the underlying exchange processes from experiments, either using the field dependence of R_2 and/or a variation of the length of the CPMG duty cycle or spin lock power (Orekhov et al., 1994; Szyperski et al., 1993). Similarly, the pH dependence of amide exchange rates could be used to estimate the kinetic parameters which characterise the unfolding processes required for hydrogen exchange (e.g. Pedersen et al., 1993).

Structural determinants for the dynamics in the TFE denatured state

Protein dynamics are encoded, just like protein structure, in the primary sequence of amino acid residues. In states approaching random coil conformations, the characteristics of certain amino acids types are likely to be of particular importance in determining the intrinsic mainchain flexibility. Of particular importance are glycine and proline residues, and in the case of oxidised hen lysozyme, cysteine residues which cross-link the polypeptide chain. The latter can be treated by the following simple model: If each amino acid peptide unit may be found in n-states of different mainchain dihedral angles, the number of possible placements of a particular residue of the polypeptide chain m residues from the nearest disulphide is n^m (ignoring possible degeneracies of the conformations and the presence of a second disulphide constraining the other end of the loop). A measure of the entropy of a residue in this model is simply proportional to the distance m to the nearest disulphide as $S \sim \ln (n^m)$. These distances are plotted in Figure 4, together with the order parameters of the native and the TFE denatured state. Some of the residues with the lowest S^2 value in the TFE state are located at sites in exact agreement with those predicted to have the greatest "entropy" at the centre of loop regions bounded by the disulphides. This very crude model appears to work extremely well, at least for the residues in the N-terminal half

of the protein. A special case are the residues at the termini of the polypeptide chain which are expected to be more felixible than the others, because unlike the case of loops, only one end of the chain is is restrained by the continuing polypeptide. Significant deviations from the dynamic behaviour predicted by the above model are apparent near the disulphides themselves and in the C-terminal half of the polypeptide chain. Cys6, 30 and 94 are located near the centre of helices, residues N-terminal to Cys64, 115 and 127 are also involved in helical structure, as is the region residues 85-90. Therefore, clearly other interactions, for example those as consequence of the secondary structure of the protein stabilised in TFE, may overrule these propensities.

A particularly interesting observation is that several features of the pattern of S^2 values in the TFE state are shared with that recently derived for the native state (Figure 4a and b). For example, the location of several of the most mobile regions, in the N-terminal part of the β-sheet and in the long loop, coincide with their position in the native state, although the overall variation in S^2 is reduced in the native state. Regions involved in structure in the native state, such as residues 1-4 which are part of a small β-strand and residues 13-20, which have several sidechains involved in the hydrophobic core, are clearly more disordered in the TFE state, showing that such structures have unfolded substantially. Intriguingly, the pattern of S^2 of residues 13-20 appears to be similar in both states. Regions which appear to have adopted, at least in part, helical structure in the TFE state, but are part of loops or tight turns in the native state also have different local patterns of S^2 values. Interestingly, these regions are predominantly located in the C-terminal half of the molecule and it is apparent from the NMR data that the great majority of residues in this part are restricted in their mobility in the TFE state. This striking difference between the N- and the C-terminal halves of the protein cannot be rationalised fully by the consideration of helix propensities and possible capping interactions, such as those discussed in Chapter 7.

In addition to the lengths of loops between disulphides, the location of glycine residues is also likely to be of considerable importance in determining the dynamic properties of the polypeptide chain. It is noticeable, for example, that a number of glycines occur in regions which are particularly mobile in the N-terminal region. This is also true for two glycines in the C-terminal region. Clearly, Gly102 and 104 at the centre of the longest loop of the C-terminal part of the protein, show considerable mobility. The other two glycines (117, 126) in this part of the protein are, however, close to disulphide residues and are therefore in regions of the polypeptide chain with higher than average density

of residues. Here, volume exclusion effects together with the propensity of certain sidechains to cluster may limit the number of conformations available to the polypeptide chain. Hydrophobicity is a measure of both of these parameters and is plotted in Figure 4d using the scale of Kyte & Doolittle (1982). The pattern obtained with other hydrophobicity scales, for example those of Eisenberg & McLachlan (1986), Sharp et al. (1991) and Blaber et al., (1994), is closely similar. It is apparent that there is good correspondence between this hydrophobicity profile pattern and the variation of S^2. Such a correlation is not unanticipated for the native state since hydrophilic residues are likely to be located in parts of the native structure which form loops and solvent exposed turns, and therefore are likely to spend on average a greater proportion of time in contact with solvent affording them greater flexibility. However, the close correlation seen hydrophobicity and S^2 values for the TFE denatured state implies that the distribution of global topologies of this state has allowed at least some partitioning of hydrophobic and hydrophilic regions of the polypeptide chain into different environments. Regions which have hydrophobic character are likely to cluster into more compact conformations, than those with hydrophilic character. Helices are also usually associated with amino acid sequences which have at least partial hydrophobic character and often have a periodicity associated with amphipathic behaviour and packing around a protein core (Eisenberg et al., 1984). Local compactness of the TFE state is likely to arise predominantly as a result of the persistence of the native disulphides (Chapter 5). Higher than average hydrophobicity is found in the regions of residue 10, 30 and 56 in the N-terminal region of the polypeptide chain. Thus there is exact overlap with the position of Cys30, but the other two maxima are shifted relative to the location of cysteines, Cys6 and Cys64. Interestingly, the minima (hydrophilic residues) overlap closely with those sites which are furthest away from disulphides in the N-terminal region up to residue 70. This correlation is, however, less distinct in the C-terminal region. The average hydrophobicity over these residues is higher than that in the N-terminal region, -0.06 for residues 80-129 compared to -0.61 for residues 1-60,; the average values of S^2 are 0.79 and 0.67 respectively for these regions. The mobility of residues 102 and 104 agrees with a local minimum of hydrophobicity. A further minimum is also found around residue 112-114, close to Cys115, whereas the centre of the final loop, bounded by cysteines 115 and 127, is hydrophobic. Thus, by contrast to the N-terminal half of the polypeptide chain, the location of hydrophobic residues and disulphides is largely out of step in the latter half of the protein sequence. This could be of possible reason why the dynamic behaviour of the C-terminal chain more uniform, at least in presence of the disulphides. The overall agreement of the

pattern of S^2 with the linear sequence profile of hydrophobicity are reflected in a correlation coefficient 0.52 if the data is optimally smoothed over a sliding window of 5 residues. Similar observations were made for the dynamics of a partially structured fragment of staphylococal nuclease in aqueous solution by Alexandrescu et al. (1994a).

Conclusions

The results of ^{15}N studies of the dynamic behaviour of the mainchain of hen lysozyme in the partially ordered TFE state, have been shown to be highly complementary to the information obtained previously from hydrogen exchange, coupling constants, NOe information and chemical shift perturbations on the structure in this denatured state (Chapter 7). The relaxation data suggest that regions of the protein can be classified in to three groups by their dynamic behaviour: Regions which are largely disordered and resemble exposed flexible loops of native proteins; regions which are partially structured and regions which posses stable and persistent secondary structure.

In particular, further study of the non-native α-helical structures which are stabilised by TFE in this state of the protein confirm their enhanched flexibility compared to the structures which are native-like. Remarkable similarities exist between the dynamic behaviour of the mainchain in the TFE denatured state and that found for the native protein. Although, the fluctuations are greater in the denatured protein, the pattern of dynamic behaviour of the polypeptide chain is similar in both states, particularly in the N-terminal half of the protein. The dynamic behaviour could be rationalised almost fully be a consideration of helix propensity, chain hydrophobicity and the location of disulphide bonds.

References

Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J., & Dobson, C.M. (1993) Biochem. 32, 1707-18.

Alexandresu, A.T. & Shortle, D. (1994a) J.Mol.Biol. 242, 527-546.

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

Allerhand, A., Doddrel, D., Glushko, V., Cochran, D.W., Wenkert, E., Jawson, P.J. & Gurd, F.R.N. (1971) J.Am.Chem.Soc. 93, 544-549.

Arcus, V., Vuilleumier, S., Freund, S.M.V., Bycroft, M. & Fersht, A.R. (1994) Proc.Natl.Acad.Sci. (USA) 91, 9412-9416.

Barbarto, G., IkurA, M., Kay, L.E., Pastor, R.W. & Bax, A. (1992) Biochem. 31, 5269-5278.

Blaber, M., Zhang, X-J., Lindstrom, J.D., Pepiot, S.D., Baase, W.A. & Matthews, B.W. (1994) J.Mol.Biol. 235, 600-624.

Boyd, J., Hommel, U. & Campbell, I.D. (1990) Chem.Phys.Lett. 175, 477-482.

Brooks, C.L. (1993) Curr.Opin.Struct.Biol. 3, 92-98.

Clore, G.M., Driscoll, P.C., Wingfield, P.T. & Gronenborn, A.M. (1990) Biochem. 29, 7387-7401.

Constantine, K.L., Friedrichs, M.S., Goldfarb, V., Jeffrey, P.D., Sherriff, S. & Mueller, L. (1993) Proteins: Struct.Funct.Genetics 15, 290-311.

Dill, K.A. & Shortle, D. (1991) Ann.Rev.Biochem. 60, 795-825.

Dobson, C.M. (1994) Curr.Biol. 4, 636-40.

Dyson, H.J. & Wright, P.E. (1991) Ann.Rev.Biophy.Biophy.Chem 20, 519-538.

Eisenberg, D., Weiss, R.M. & Terwilliger, T.C. (1984) Proc.Natl.Acad.Sci 81, 140-144.

Eisenberg, D. & McLachlan, A.D. (1986) Nature 319, 199-203.

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

Grasberger, B.L., Gronenborn, A.M. & Clore, G.M. (1993) J.Mol.Biol. 230, 364-372.

Handol, H.H.G. (1985) D.Phil.Thesis, Univeristy of Oxford.

Kay, L.E., Torchia, D.A. & Bax, A. (1989) Biochem. 28, 8972-8979.

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

Kemmink, J. & Creighton, T.E. (1993) J.Mol.Biol. 234, 861-878.

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

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

Kyte, J. & Doolittle, R.F. (1982) J.Mol.Biol. 157, 105-132.

Lipari, G. & Szabo, A. (1982) J.Am.Chem.Soc. 104, 4559-4570.

Logan, T.M., Theriault, Y. & Fesik, S.W. (1994) J.Mol.Biol. 236, 637-658.

London, R.E. (1989) Meth.Enzym. 176, 358-375.

Messerle, B.A., Wider, G., Otting, G., Weber, C. & Wüthrich, K (1989) J.Mag.Reson. 85, 608-613.

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

Norwod, T.J., Crawford, D.A., Steventon, M.E., Driscoll, P.C. & Campbell, I.D. (1992) Biochem. 31, 6285-6290.

Orekhov, U.A., Pervushin, K.V. & Arseniev, A.S. (1994) Eur.J.Biochem. 219, 887-896.

Palmer, A.G., Rance, M. & Wright,P.E. (1991) J.Am.Chem.Soc. 113, 4370-4380.

Palmer, A.G., Skelton, N.J., Chazin, W.J., Wright, P.E. & Rance, M. (1992) Mol.Phys. 75, 699-711.

Palmer, A.G., Hochstrasser, R.A., Millar, D.P., Rance, M. & Wright, P.E. (1993) J.Am.Chem.Soc. 115, 6333-6345.

Ptitsyn, O.B. (1992) in "Protein Folding", T.E. Creighton Ed., Freeman, N.Y.

Peng, J.W. & Wagner, G. (1992) Biochem. 31, 8571-8586.

Pedersen, T.G., Thomsen, N.K., Andersen, K.V., Madsen, J.C. & Poulsen, F.M. (1993) J.Mol.Biol. 230, 651-660.

Redfield, C., Boyd, J., Smith, L.J., Smith, R.A.G & Dobson, C.M. (1992) Biochem. 31, 10431-10437.

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

Roberts, I.N., MacKenzie, D.A., Jeenes, D.B., Radford, S.E., Robinson, C.V., Aplin, R.T. & Dobson, C.M. (1992) Biotechnology letters 14, 897-902.

Roder, H., Wagner, G. & Wüthrich, K. (1985) Biochem. 24, 7396-7405.

Schneider, D.M., Dellwo, M.J. & Wand, J.A (1992) Biochem. 31, 3645-3652.

Sharp, K.A., Nicholls, A., Friedman, R. & Honig, B. (1991) Biochem. 30, 9686-9697.

Stone, M.J., Fairbrother, W.J., Palmer, A.G., Reizer, J., Saier, M.H. & Wright, P.E. (1992) Biochem. 31, 4394-4406.

Szyperski, T., Luginbuehl, P., Otting, G., Günert, P., & Wüthrich, K. (1993) J.Biomol.NMR 3, 151-164.

Torchia, D.A., Lyerla, J.R. & Quattrone, A.J. (1974) Biochem. 14, 887-900.

Wüthrich, K. (1994) Curr.Opin.Struct.Biol. 4, 93-99.

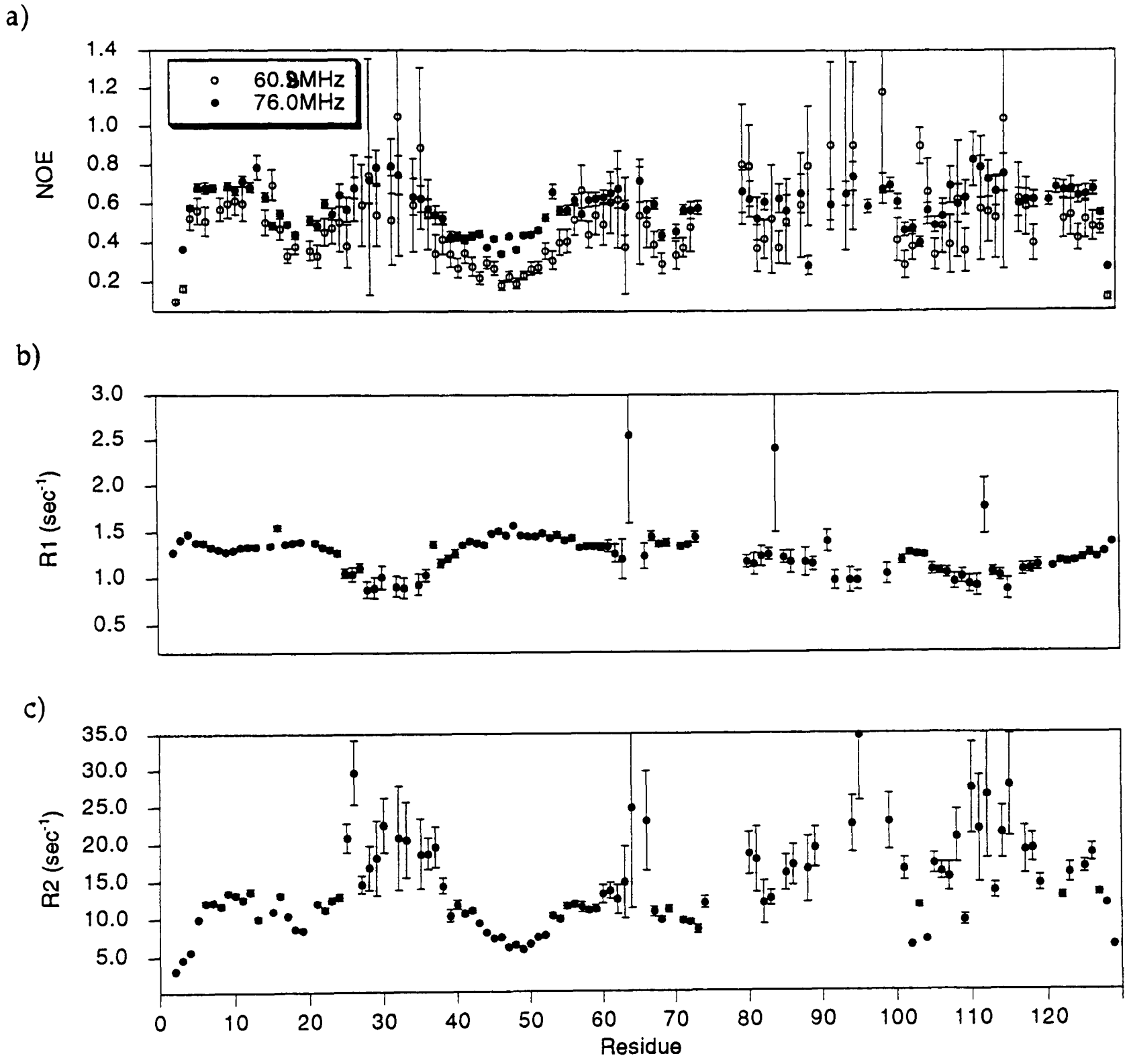


Figure 1

Relaxation data of mainchain ^{15}N sites in the TFE denatured state

a) heteronuclear NOE at ^{15}N spectrometer frequencies of 76.0 and 60.8 MHz,

b) Longitudinal $R_1 = 1/T_1$ and c) Transverse $R_2 = 1/T_2$ relaxation rates, both measured at 60.8 MHz.

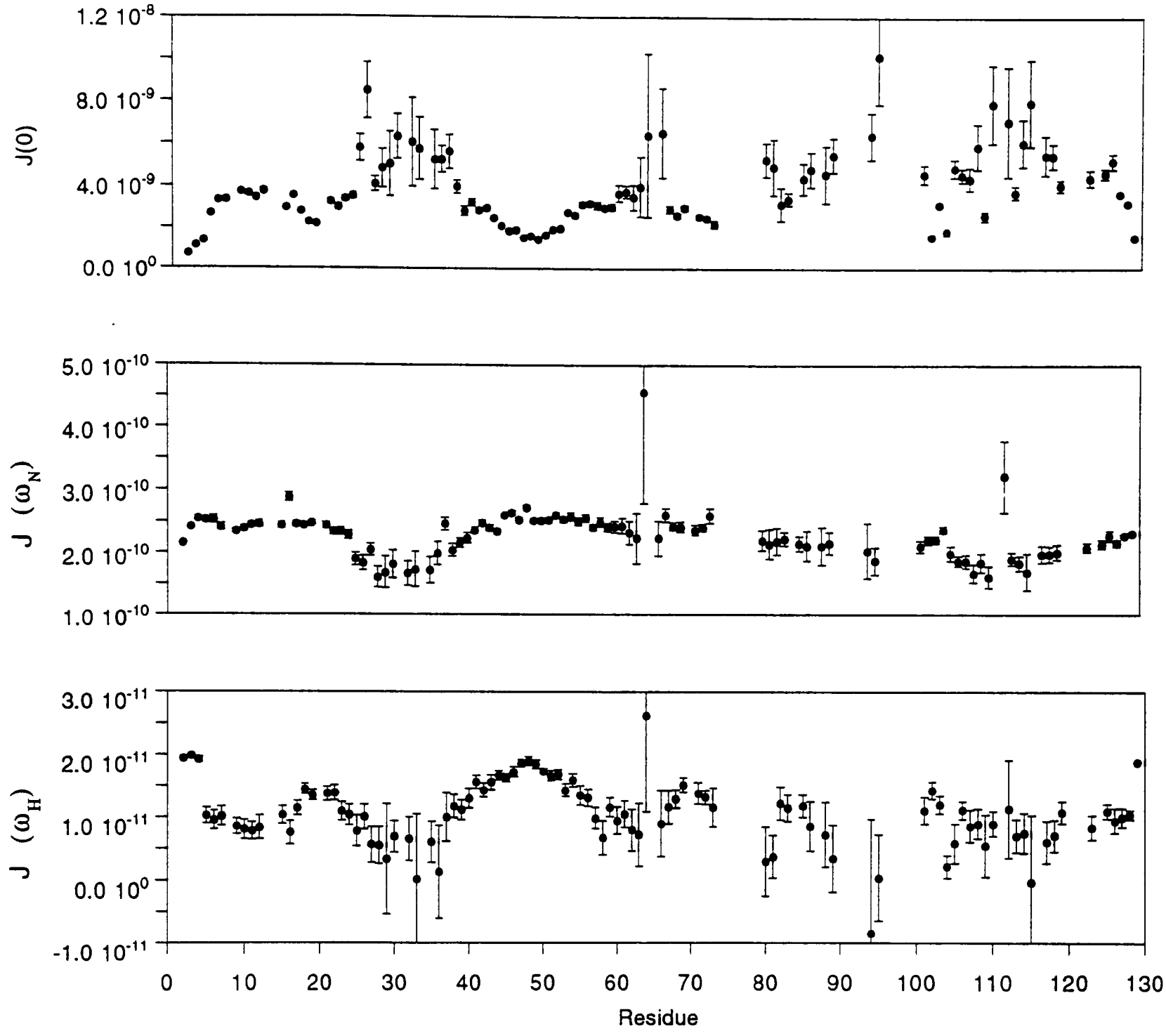


Figure 2

Values of the spectral density function (in rad s^{-1}) derived at frequencies of 0, ω_N and ω_H using the quasi spectral density mapping approach plotted as a function of protein sequence.

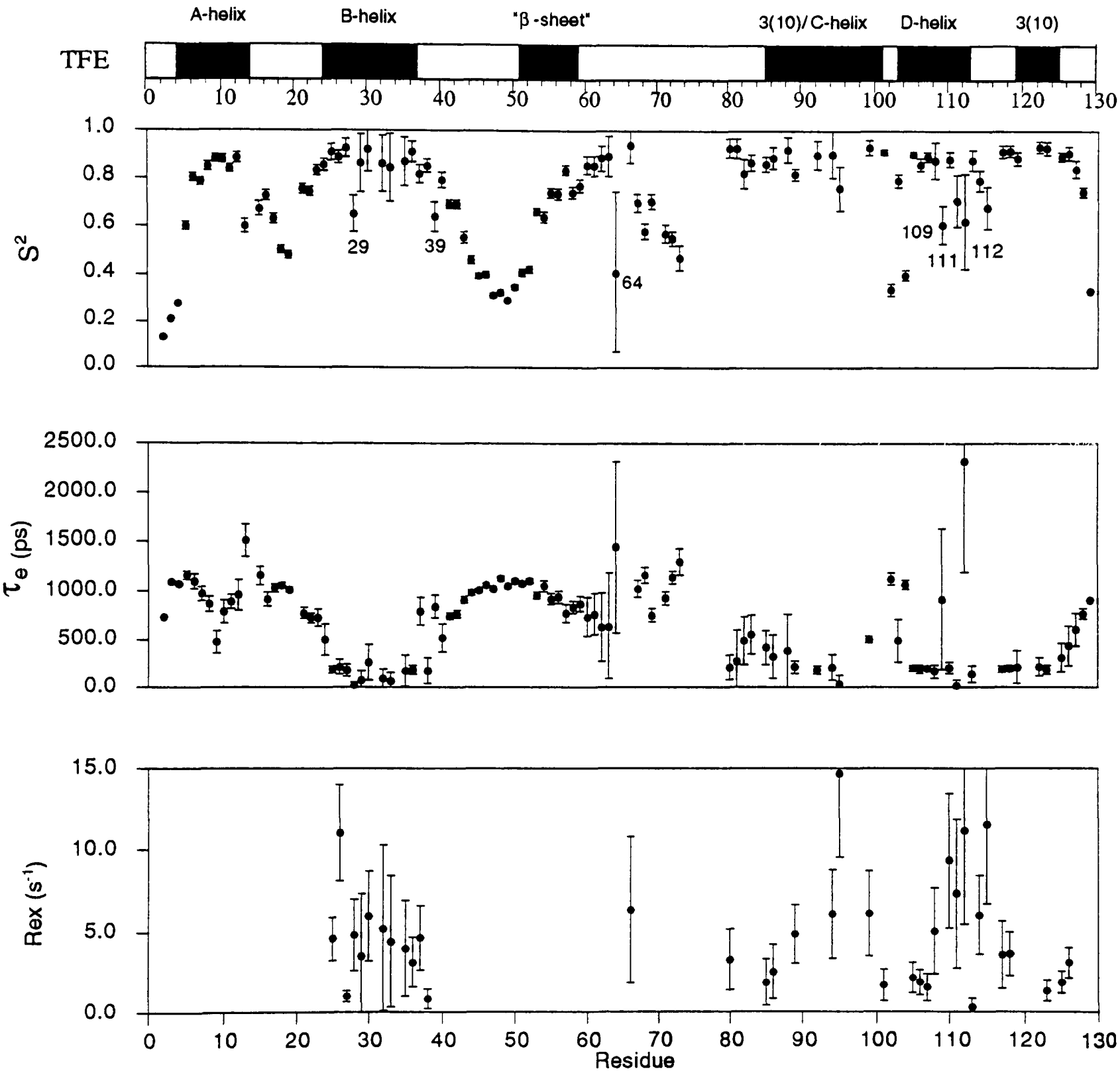


Figure 3

Order parameters for the mainchain in the TFE denatured state, derived from the relaxation measurements using an optimised rotational correlation time of 10.7 ± 0.1 ns. The location of helical secondary structure in the TFE state, defined by NOEs, hydrogen exchange protection and Hα chemical shift perturbations (Chapter 7), is indicated.

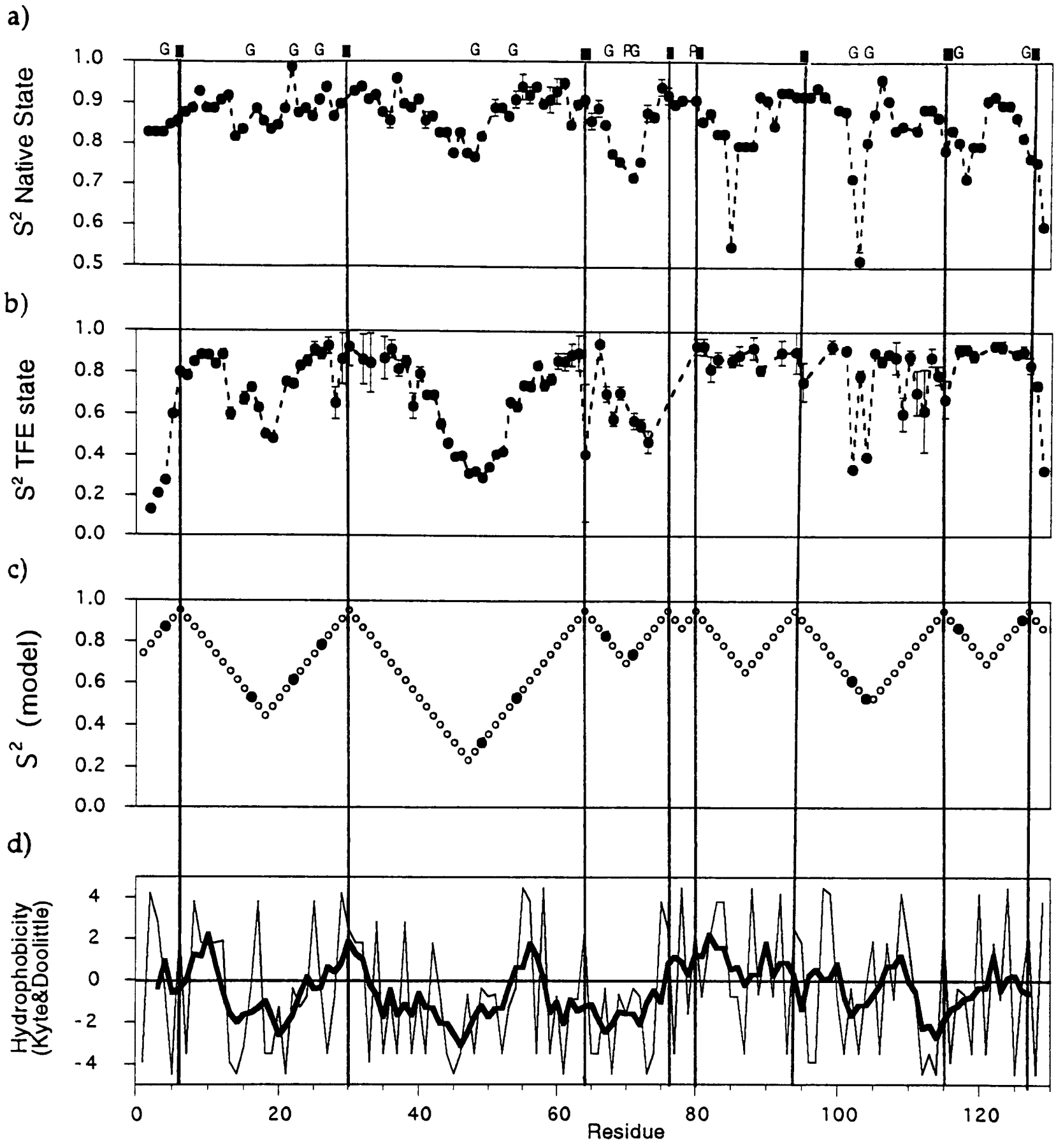


Figure 4

Order parameters (S^2) for a) the native state of hen lysozyme at pH 3.8, 35°C (Chapter 6), b) the TFE denatured state of hen lysozyme, c) plot of distances to the nearest disulphide, according to the model described in the text, and scaled linearly to fit the range of S^2, d) hydrophobicity profile for the sequence of hen lysozyme derived using the scale of Kyte & Doolittle (1982), the data points are shown individually and averaged over a sliding window of 5 residues.

Designation	Residues	S^2	τ_e (ps)	Rex (sec^{-1})	$\chi 2$	$\tau_{R,}$ loc (nsec)	S^2,loc	τ_e,loc (ps)	Rex,loc (sec^{-1})	$\chi 2$,loc
A-helix	6-12	.85+/-.04	870+/-200	0.0+/-0.0	11.0	10.0+/-0.2	.89+/-.03	510+/-270	0.0+/-0.0	8,7
B-helix	23-38	.86+/-.07	270+/-240	4.4+/-2.6	8.0	12.6+/-0.2 13.1+/-0.3	.86+/-.08 .86+/-.09	480+/-360 510+/-380	3.0+/-2.4 0.0+/-0.0	3.3 5.0
C-helix	90-101	.88+/-.07	240+/-130	7.1+/-5.4	5.3	12.4+/-0.4 12.8+/-0.4	.94+/-.01 .94+/-.01	350+/-130 360+/-140	5.4+/-5.2 0.0+/-0.0	3.6 5.1
D-helix	105-114	.80+/-.11	490+/-730	5.0+/-3.8	10.7	12.5+/-0.4 13.2+/-0.4	.80+/-.02 .82+/-.16	650+/-900 750+/-910	3.7+/-3.8 0.0+/-0.0	2.9 4.0
3_{10}(i)	80-89	.88+/-.05	360+/-130	3.1+/-1.3	2.6	12.1+/-0.5	.86+/-.06	650+/-150	1.8+/-0.6	1.8
3_{10}(ii)	117-124	.92+/-.02	210+/-12	2.9+/-1.3	12.9	11.5+/-0.3	.91+/-.02	330+/-160	2.1+/-1.3	9.5
"β-sheet"	53-61	.76+/-.08	870+/-100	0.0+/-0.0	9.2	9.96+/-0.2	.81+/-.07	680+/-180	0.0+-/-0.0	8.5

Table 1

Results from regional fitting of rotational correlation times and order parameters. The globally and locally estimated rotational correlation times are indicated, together with the derived values for S^2, τ_e and Rex. In the case of the B, C and D-helices the spectral density function was also modelled in absence of possible exchange contributions (second line of these entries).

Residue		S^2 (S^2s*S^2f)	S^2f	S^2s	τ_e (ps)	Rex (s^{-1})	χ^2
Lys	1						
Val	2	0.13 ±0.00	0.77 ±0.01	0.17 ±0.00	737.1 ±2.9		1949.7
Phe	3	0.21 ±0.00	0.77 ±0.00	0.27 ±0.00	1093.8 ±12.1		0.0
Gly	4	0.28 ±0.01	0.83 ±0.01	0.33 ±0.01	1068.8 ±11.9		9.7
Arg	5	0.60 ±0.02	0.90 ±0.01	0.67 ±0.01	1160.2 ±44.5		1.5
Cys	6	0.80 ±0.02			1098.4 ±75.4		8.7
Glu	7	0.79 ±0.02			978.2 ±69.8		18.6
Leu	8	0.85 ±0.02			872.8 ±76.3		29.2
Ala	9	0.88 ±0.02			480.4 ±117.1		1.4
Ala	10	0.88 ±0.02			795.9 ±115.7		7.3
Ala	11	0.84 ±0.02			894.4 ±70.0		10.8
Met	12	0.89 ±0.02			962.5 ±157.4		0.7
Lys	13	0.60 ±0.03	0.84 ±0.02	0.71 ±0.02	1510.9 ±163.5		0.0
Arg	14						
His	15	0.67 ±0.03	0.91 ±0.02	0.74 ±0.02	1159.2 ±93.5		0.2
Gly	16	0.73 ±0.02			917.6 ±69.6		43.0
Leu	17	0.63 ±0.02	0.91 ±0.01	0.69 ±0.01	1033.0 ±44.9		0.7
Asp	18	0.50 ±0.02	0.87 ±0.01	0.58 ±0.01	1057.6 ±28.1		0.0
Asn	19	0.48 ±0.01	0.87 ±0.01	0.55 ±0.01	1014.0 ±25.1		9.0
Tyr	20						
Arg	21	0.76 ±0.02			781.1 ±57.4		0.8
Gly	22	0.75 ±0.02			731.1 ±51.3		8.0
Tyr	23	0.83 ±0.02			728.8 ±86.9		5.5
Ser	24	0.86 ±0.03			505.4 ±162.8		2.1
Leu	25	0.91 ±0.03			193.9 ±39.6	4.6 ±1.3	17.6
Gly	26	0.89 ±0.03			217.3 ±77.6	11.0 ±3.0	9.9
Asn	27	0.93 ±0.04			185.0 ±63.2	1.0 ±0.3	12.6
Trp	28	0.65 ±0.07			27.7 ±34.8	4.8 ±2.2	0.0
Val	29	0.86 ±0.12			77.2 ±97.4	3.5 ±3.8	9.9
Cys	30	0.92 ±0.09			268.3 ±186.5	6.0 ±2.7	6.1
Ala	31						
Ala	32	0.86 ±0.12			96.8 ±99.6	5.2 ±5.1	10.4
Lys	33	0.85 ±0.14			66.2 ±94.1	4.4 ±4.0	10.9
Phe	34						
Glu	35	0.87 ±0.10			179.4 ±163.5	4.0 ±2.9	8.9
Ser	36	0.91 ±0.04			191.0 ±46.1	3.2 ±1.5	12.8
Asn	37	0.82 ±0.04			795.2 ±138.2	4.6 ±2.0	0.1
Phe	38	0.85 ±0.03			178.1 ±136.4	0.9 ±0.6	5.1
Asn	39	0.64 ±0.06	0.86 ±0.03	0.75 ±0.04	843.5 ±115.2		0.0
Thr	40	0.79 ±0.03			521.8 ±144.0		2.9
Gln	41	0.69 ±0.02			749.9 ±36.8		9.9
Ala	42	0.69 ±0.02			767.7 ±38.9		2.7
Thr	43	0.56 ±0.02	0.91 ±0.01	0.61 ±0.02	911.2 ±34.5		0.1
Asn	44	0.46 ±0.02	0.85 ±0.01	0.54 ±0.01	986.8 ±28.1		0.8
Arg	45	0.40 ±0.01	0.89 ±0.01	0.45 ±0.01	1011.0 ±15.7		23.5
Asn	46	0.40 ±0.01	0.89 ±0.01	0.45 ±0.01	1062.0 ±19.5		3.6
Thr	47	0.31 ±0.01	0.85 ±0.01	0.37 ±0.01	1019.9 ±15.8		14.3
Asp	48	0.33 ±0.01	0.88 ±0.01	0.37 ±0.01	1128.5 ±15.8		1.3
Gly	49	0.29 ±0.01	0.83 ±0.01	0.35 ±0.01	1051.0 ±14.0		9.7
Ser	50	0.35 ±0.01	0.83 ±0.01	0.42 ±0.01	1106.4 ±15.6		0.6
Thr	51	0.41 ±0.01	0.86 ±0.01	0.47 ±0.01	1073.0 ±23.3		1.0
Asp	52	0.42 ±0.01	0.88 ±0.01	0.48 ±0.01	1104.6 ±21.9		0.1
Tyr	53	0.66 ±0.01			957.0 ±32.7		17.1
Gly	54	0.64 ±0.02			1050.4 ±52.9		25.3

Table 2

Derived Lipari & Szabo parameters and goodness of fit.

Residue	S^2 (S^2_S*S^2_f)	S^2_f	S^2_s	τ_e (ps)	Rex (s^{-1})	χ^2
Lys 1						
Val 2	0.13 ±0.00	0.77 ±0.01	0.17 ±0.00	737.1 ±2.9		1949.7
Phe 3	0.21 ±0.00	0.77 ±0.00	0.27 ±0.00	1093.8 ±12.1		0.0
Gly 4	0.28 ±0.01	0.83 ±0.01	0.33 ±0.01	1068.8 ±11.9		9.7
Arg 5	0.60 ±0.02	0.90 ±0.01	0.67 ±0.01	1160.2 ±44.5		1.5
Cys 6	0.80 ±0.02			1098.4 ±75.4		8.7
Glu 7	0.79 ±0.02			978.2 ±69.8		18.6
Leu 8	0.85 ±0.02			872.8 ±76.3		29.2
Ala 9	0.88 ±0.02			480.4 ±117.1		1.4
Ala 10	0.88 ±0.02			795.9 ±115.7		7.3
Ala 11	0.84 ±0.02			894.4 ±70.0		10.8
Met 12	0.89 ±0.02			962.5 ±157.4		0.7
Lys 13	0.60 ±0.03	0.84 ±0.02	0.71 ±0.02	1510.9 ±163.5		0.0
Arg 14						
His 15	0.67 ±0.03	0.91 ±0.02	0.74 ±0.02	1159.2 ±93.5		0.2
Gly 16	0.73 ±0.02			917.6 ±69.6		43.0
Leu 17	0.63 ±0.02	0.91 ±0.01	0.69 ±0.01	1033.0 ±44.9		0.7
Asp 18	0.50 ±0.02	0.87 ±0.01	0.58 ±0.01	1057.6 ±28.1		0.0
Asn 19	0.48 ±0.01	0.87 ±0.01	0.55 ±0.01	1014.0 ±25.1		9.0
Tyr 20						
Arg 21	0.76 ±0.02			781.1 ±57.4		0.8
Gly 22	0.75 ±0.02			731.1 ±51.3		8.0
Tyr 23	0.83 ±0.02			728.8 ±86.9		5.5
Ser 24	0.86 ±0.03			505.4 ±162.8		2.1
Leu 25	0.91 ±0.03			193.9 ±39.6	4.6 ±1.3	17.6
Gly 26	0.89 ±0.03			217.3 ±77.6	11.0 ±3.0	9.9
Asn 27	0.93 ±0.04			185.0 ±63.2	1.0 ±0.3	12.6
Trp 28	0.65 ±0.07			27.7 ±34.8	4.8 ±2.2	0.0
Val 29	0.86 ±0.12			77.2 ±97.4	3.5 ±3.8	9.9
Cys 30	0.92 ±0.09			268.3 ±186.5	6.0 ±2.7	6.1
Ala 31						
Ala 32	0.86 ±0.12			96.8 ±99.6	5.2 ±5.1	10.4
Lys 33	0.85 ±0.14			66.2 ±94.1	4.4 ±4.0	10.9
Phe 34						
Glu 35	0.87 ±0.10			179.4 ±163.5	4.0 ±2.9	8.9
Ser 36	0.91 ±0.04			191.0 ±46.1	3.2 ±1.5	12.8
Asn 37	0.82 ±0.04			795.2 ±138.2	4.6 ±2.0	0.1
Phe 38	0.85 ±0.03			178.1 ±136.4	0.9 ±0.6	5.1
Asn 39	0.64 ±0.06	0.86 ±0.03	0.75 ±0.04	843.5 ±115.2		0.0
Thr 40	0.79 ±0.03			521.8 ±144.0		2.9
Gln 41	0.69 ±0.02			749.9 ±36.8		9.9
Ala 42	0.69 ±0.02			767.7 ±38.9		2.7
Thr 43	0.56 ±0.02	0.91 ±0.01	0.61 ±0.02	911.2 ±34.5		0.1
Asn 44	0.46 ±0.02	0.85 ±0.01	0.54 ±0.01	986.8 ±28.1		0.8
Arg 45	0.40 ±0.01	0.89 ±0.01	0.45 ±0.01	1011.0 ±15.7		23.5
Asn 46	0.40 ±0.01	0.89 ±0.01	0.45 ±0.01	1062.0 ±19.5		3.6
Thr 47	0.31 ±0.01	0.85 ±0.01	0.37 ±0.01	1019.9 ±15.8		14.3
Asp 48	0.33 ±0.01	0.88 ±0.01	0.37 ±0.01	1128.5 ±15.8		1.3
Gly 49	0.29 ±0.01	0.83 ±0.01	0.35 ±0.01	1051.0 ±14.0		9.7
Ser 50	0.35 ±0.01	0.83 ±0.01	0.42 ±0.01	1106.4 ±15.6		0.6
Thr 51	0.41 ±0.01	0.86 ±0.01	0.47 ±0.01	1073.0 ±23.3		1.0
Asp 52	0.42 ±0.01	0.88 ±0.01	0.48 ±0.01	1104.6 ±21.9		0.1
Tyr 53	0.66 ±0.01			957.0 ±32.7		17.1
Gly 54	0.64 ±0.02			1050.4 ±52.9		25.3

Table 2

Derived Lipari & Szabo parameters and goodness of fit.

Residue		S^2 (S^2s*S^2f)	S^2f	S^2s	τ_e (ps)	Rex (s^{-1})	χ^2
Trp	111*	0.71 ±0.11			22.2 ±56.4	7.3 ±4.5	0.0
Arg	112 *	0.62 ±0.20			2322.5 ±1132.3	11.2 ±5.7	0.2
Asn	113	0.88 ±0.04			144.3 ±89.7	0.3 ±0.6	17.9
Arg	114	0.79 ±0.04				6.0 ±2.4	7.6
Cys	115	0.68 ±0.09				11.6 ±4.9	1.0
Lys	116						
Gly	117	0.91 ±0.03			197.2 ±34.5	3.6 ±2.1	7.6
Thr	118	0.92 ±0.02			203.0 ±32.2	3.7 ±1.3	10.4
Asp	119	0.89 ±0.03			214.6 ±172.7		8.3
Val	120						
Gln	121						
Ala	122	0.93 ±0.02			220.4 ±95.7		22.8
Trp	123	0.93 ±0.03			190.5 ±46.1	1.4 ±0.7	15.4
Ile	124						
Arg	125	0.89 ±0.02			314.2 ±150.6	1.9 ±0.7	10.2
Gly	126	0.91 ±0.03			436.9 ±207.2	3.1 ±0.9	1.2
Cys	127	0.84 ±0.04	0.93 ±0.02	0.90 ±0.02	606.8 ±175.5		10.0
Arg	128	0.74 ±0.02	0.93 ±0.01	0.80 ±0.01	771.6 ±54.1		0.1
Leu	129	0.33 ±0.01	0.83 ±0.01	0.40 ±0.01	908.8 ±9.3		36.2

Table 3

$S^2 \geq 0.82$ (Centre of Helices)	$0.82 \leq S^2 \leq 0.60$ (Ext. & non-native Helix)	$S^2 \leq 0.60$ (Unfolded and other)
A: 8-12	5-7 13-17	N-term: 1-4 18-19
	21,22	
B: 24-38*	39-42, 53-59****	43-52
3_{10}(i): 80-88		71-73
C: 89-99**	103	102,104
D: 105-113***	114,115	
3_{10}(ii): 117-127	125	129

Except residue *) 28, **) 95, ***) 109, 111, 112, ****) 57

Table 3
Classification of the principal regions of the polypeptide chain in the TFE denatured state of hen lysozyme by their dynamic behaviour.

Chapter 9

Epilogue

"L'essentiel est invisible pour les yeux""

"Le petit Prince", Antoine de Saint-Exupéry

This section comprises a brief overview of the principal findings of the research described in this thesis presented in the context of interesting, but largely unresolved questions and suggests several hypotheses for future investigation.

Relationship between equilibrium and kinetic intermediates

Relationships between intermediates populated transiently during kinetic refolding and partially folded states of a protein which are stable under equilibrium conditions has traditionally been inferred from their similar spectral characteristics. For example, near and far uv CD suggested that the acid denatured state of α-lactalbumin is closely similar to a refolding intermediate (Ikeguchi et al., 1986). However, given that these probes report on the overall structure, such agreement may be coincidental. Similarly, the pattern of protection against amide hydrogen exchange of the TFE-denatured state of lysozyme resembles that found in the major intermediate, populated transiently during kinetic refolding (Chapter 3). However, in this case results obtained from other measurements, for example the absence of ANS binding, the presence of residual near uv CD and the negligible role of disulphides in the TFE state (Chapter 4) as well as the preliminary NMR structure (Appendix II), suggest the TFE state has at least some characteristics unlike those of the kinetic intermediate.

Additional evidence for the equivalence of kinetic and equilibrium intermediates could in principle be obtained by using the equilibrium intermediate as the starting or final state in kinetic refolding or unfolding experiments. In practice such experiments are complicated by the fact that solution conditions must necessarily be changed to initiate (un-)folding. Interestingly, while the extent of overall residual structure differs greatly in the guanidine HCl, DMSO and TFE denatured states of hen lysozyme, the refolding kinetics and the nature of the folding pathway is not significantly affected, but depends critically on the final

solution conditions (Kotik et al., 1994; Lu, Buck, Radford and Dobson - in preparation). This is despite the fact that the majority of secondary structure in the TFE-denatured state is native-like and could be anticipated to offer an advantageous starting point for refolding. However, it is probable that in the case of the TFE- or other denatured states, residual structures are stabilised to a significant extent by the presence of denaturant, which is diluted at the onset of the folding process. Thus the structures become relatively more unstable and interconvert easily, making it less likely that they would direct the folding pathway (Creighton, 1988; Kemming & Creighton, 1993).

The role of the equilibrium intermediates of cytochrome c in the folding process was probed by a number of similar experiments. Folding was found to be greatly accelerated when initiated from the molten globule state at low pH, while the refolding from a high pH form, which included a non-native heam ligation as part of the partially folded state, appeared to first require its substantial unfolding. This does not mean, however, that structure in equilibrium intermediates which is native-like, is *a priori* that which drives productive folding in kinetic intermediates. For example, the C-peptide of ribonuclease A is known to populate native-like α-helical conformations with high persistency when isolated in solution, but is apparently not involved in any of the early folding processes of the protein (Bierzynski et al., 1982; Brems & Baldwin, 1984; Udgaonkar & Baldwin, 1990). Conversely, site-directed mutagenesis suggests that non-native, but persistent clustering of a number of sidechains in the molten globule state of α-lactalbumin may be responsible for its formation and efficient folding to the native state (Alexandrescu et al., 1993; Brew - personal communication). Recently, in particular with the observation of parallel pathways in a number of systems, it has been questioned, whether kinetic intermediates populated transiently to the extent allowing their detailed structural characterisation, represent intermediates which lead to productive folding, or whether they are in fact transiently trapped and largely misfolded species (Radford et al., 1992; Sosnick et al., 1994; Creighton, 1994).

Protein folding studies at equilibrium

(i) Significance of structures in denatured and partially folded states.

Recently, the study of denatured and, in favourable cases, of partially structured proteins has become possible in site-specific detail by the use of heteronuclear multidimensional NMR (Wüthrich, 1994). Variation of the solution conditions, often necessary in order to overcome the stability of the native state and to generate denatured states, allows an assessment to be made of the propensity of the polypeptide chains to adopt structures in different solution environments.

Furthermore, studies of the structures and their stabilities as a function of chain length, as parts of truncated proteins, protein fragments or peptides, aims to provide a description of the alternative conformations available to a stretch of polypeptide chain and of the identity and interdependence of the competing interactions, whether they are within the polypeptide or with (co-)solvent. It is known from analysis of crystal structures that in some instances the same stretch of amino acids can assume different conformations, depending on its structural context (e.g. Argos, 1987) while peptide conformation may be largely determined by the solution conditions (Zhong & Johnson, 1992). Relevance to protein folding, therefore exists in that the folding mechanisms themselves rely on differential environments which are created by the polypeptide chain and the surrounding medium, possibly involving interactions with environments provided by auxiliary factors, such as chaperones. It has been suggested that TFE in particular may mimic the environment of membranes or of the interior of loosely folded proteins (Norwood et al., 1992; Rizo et al., 1993)

The majority of the structure described in the TFE denatured state in Chapter 7, as well as those detected in denatured and partially folded states of other proteins or peptides, is native-like in character, at least at the level of secondary structure (Barrick & Baldwin, 1993; Logan et al., 1994; Alexandrescu et al., 1994; Arcus et al., 1994). Although these structures are often present only in low population, compared to other more unstructured conformations, this finding may be of significance for the understanding of protein folding. Thus, although near-native like conformations in denatured states may not necessarily offer kinetic or thermodynamic advantages for the folding process *per se*, their presence reflects the nature of the potential landscape, which is explored during folding. This would support a hypothesis of folding as a hierarchical process, in which folding may be accomplished by a simple amplification of structures which are already present at the local level of a linear polypeptide chain. The folding process would then correspond to a raising of barriers of interconversion and stabilisation of structured conformations. This proposition is supported by a further intriguing observation made during preliminary structure calculations on the structure of the TFE denatured state. In presence of disulphides and persistent secondary structure a significant proportion the conformations in the ensemble of NMR structures have interhelical angles corresponding to those of the native state. While there are no molecules with a native-like overall topology, certainly several subdomains have near-native conformational preferences (see Appendix II). It is envisaged that such structures, with the right secondary structure (Chapter 4) and approximate native topology could coalesce through a diffusion-collision process (Karplus & Weaver, 1994) and thus direct the conformational search towards the native

state. In the case of the TFE state these topological preferences are likely to be dependent, at least in part, on the persistence of the native disulphides. However, this may not always be the case. It is possible that a number of largely non-specific contacts, for example those arising from the pattern of polar and non-polar residues in the polypeptide chain, are sufficient to specify both the preference for certain secondary structures as well as the overall topology of the protein in a partially structured state. Indeed, Peng & Kim (1994) have suggested that this may be the case for the molten globule state of the α-subdomain of α-lactalbumin. Experiments are in progress on a number of systems, including models for protein design, to elucidate the code for these topological preferences (Harbury et al., 1993; Kamtekar et al., 1993).

Examples of non-native structures have also be reported, including in the present work, and would at first sight appear to contradict such a hypothesis. However, non-native interactions in peptides can be largely explained by the absence of residues which are providing long range native interactions (e.g. Smith et al., 1994). In the case of chemically denatured states of proteins, it is likely that long-range interactions are disrupted by competing interactions with cosolvent or denaturant. Significantly, a large proportion of regions of the polypeptide chain found to adopt a non-native conformation are part of β-structures in the native state of the proteins (e.g. Sönnichsen et al., 1992; Fan et al., 1993; Liu al., 1994). Intriguingly, there is experimental evidence in the case of β-lactoglobulin, a protein largely comprised of β-sheet structure in the native state, that intermediates with extensive α-helical character are sampled on the folding pathway (Kuwajima et al., 1987). The reasons for such an α- to β-structural rearrangement in the folding process are not clear at present and from the few examples available it cannot be certain whether or not such a mechanism is part of the repertoire of protein folding processes. Advantages could, for example, arise in the transient formation of α-helical structure because this may entail a partial compaction of the protein early in folding and avoid intermolecular association reactions.

(ii) Significance of dynamics in denatured and partially folded states.

NMR can provide information not only about structural propensities in denatured states but also about the persistency and dynamics of the structures (See Appendix I). The low protection of amides against hydrogen exchange with solvent in the urea and other highly denatured states of hen lysozyme for example (Chapter 2), implies that residual structure which exists in these states (Schwalbe, Buck & Dobson - in preparation) is not very persistent and may therefore not be expected to play a significant part in folding.

The ^{15}N relaxation data for the mainchain and several sidechain sites in native lysozyme suggest that absence of van der Waals contacts is a major determinant for mobility (Chapter 6). This principle could be of possible significance for protein folding if it were to extend to models of partially folded or "molten globule" type states (Chapter 8 & Appendix I): once the main and sidechains atoms are buried by, or are in extensive contact with, other protein structures, they are likely to become conformationally less flexible. This would suggest that folding of autonomous units in proteins proceeds along a pathway involving a mechanism of sequential on-site assembly or propagation. Such a mechanism would require a "structural nucleus" as a centre, which may be formed by an all-or-none process; such as the collapse to a "molten globule" type state. The conformation of such a nucleus would be predicted to be largely native-like in character. If non-native interactions are formed at all in the folding process they have to be few and in exposed environments, since extensive reorganisation of the mainchain or of sidechains in even loosely structured environments is likely to be difficult and slow. Similar predictions for folding processes have recently been made from theoretical studies using lattice models (Abkevich et al., 1994).

A significant finding is that while the amplitude of fluctuations is much greater than that of the native state, the pattern of dynamic behaviour along the mainchain in a partially folded state in TFE (Chapter 8), and also in a highly denatured state in urea (Schwalbe, Buck & Dobson, in preparation) bears, at least in part, considerable resemblance to it. Depending on the extent of partial compactness of these states, for example as a consequence of the persistence of native-like disulphides, this observation would appear to lend support to the hypothesis above. It may also extend the hypothesis advanced in section (i) above. It is possible that, if one accepts that native-state protein dynamics are an integral part of the folding code, such dynamics should similarly be represented in an embryonic manner in denatured and partially folded states. Because structure and dynamics are intimately related, it is highly probable that the level of complexity of this code is similar at least to the stage of loosely folded molten globule states; indeed, lattice models containing only polar and non-polar residues are beginning to reproduce basic features of the folding process with some accuracy (Chan & Dill, 1991). Thus it is intriguing that there should exist a reasonably good correlation between polypeptide chain hydrophobicity and reduced mobility in several partially folded states (Alexandrescu et al., 1994b; and Chapter 8).

Sidechain felixibility is of particular interest in partially ordered states of proteins. The application of ^{15}N and ^{13}C relaxation methods to sidechains (eg. Palmer et al.,

1993; & Chapter 6) should allow a detailed assessment of the dynamics of the interior of incompletely packed "molten globule" type states. Following the argument made above, the detection of considerable heterogeniety in the conformational preferences of sidechains, such as their their tendency to cluster as part of embryonic cores, for example, would have significant implications for our understanding of the last stages of protein folding.

(iii) Molecular Dynamics and Protein Folding.

Increasingly the interpretation of the NMR data, particularly with respect to the elucidation of structures and the dynamics in non-native states, requires the derivation computational models. Such models are frequently tested against the experimental data by subjecting them to extensive molecular dynamics simulations (e.g. Daggett & Levitt, 1994; Caflisch & Karplus, 1994). The detail available on molecular processes from such simulations is unparalleled by experimental techniques and if used cautiously should provide new insights into the protein folding problem. Specifically, provided that the simulations can be run over longer periods of time or that efficient sampling methods are available, some of the above hypotheses could be tested by computational as well as by experimental means.

References for Epilogue

Abkevich, V.I., Gutin, A.M. & Shakhnovich, E.I. (1994) Biochem. 33, 10026-10036.

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

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

Alexandrescu, A.T. & Shortle, D. (1992b) J.Mol.Biol. 242, 527-546.

Arcus, V., Vuilleumier, S., Freund, S.M.V., Bycroft, M. & Fersht, A.R. (1994) Proc.Natl.Acad.Sci. (USA) 91, 9412-9416.

Argos, P. (1987) J.Mol.Biol. 197, 331-348.

Barrick, D. & Baldwin, R.L. (1993) Protein Science 2, 869-876.

Bierzynski, A., Kim, P.S. & Baldwin, R.L. (1982) Proc.Natl.Acad.Sci. (USA) 79, 2470-2474.

Brems, D.N. & Baldwin, R.L. (1984) J.Mol.Biol. 180, 1141-1156.

Caflisch, A., & Karplus, M. (1994) Proc.Natl.Acad.Sci. (USA) 91, 1746-1750.

Chan, H. & Dill, K. (1991) Ann.Rev.Biophys.Biophs.Chem. 22, 447-490.

Creighton, T.E. (1988) Biophys.Chem. 31, 155-162.

Creighton, T.E. (1994) Nature: Struct.Biol. 1, 135-138.

Daggett, V. & Levitt,M. (1994) Curr.Opin.Struct.Biol. 4, 291-295.

Harbury, P.B., Zhang, T., Kim, P.S. & Alber, T. (1993) Science 262, 1401-1407.

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

Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M. & Hecht, M.H. (1993) Science 262, 1680-1685.

Karplus, M. & Weaver, D.L. (1994) Protein Science 3, 650-668.

Kemmink, J. & Creighton, T.E. (1993) J.Mol.Biol. 234, 861-878.

Kotik, M., Radford, S.E. & Dobson (1994) Biochem., in press.

Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. & Nagamura, T. (1987) FEBS Lett. 221, 115-118.

Liu, Z-P., Rizo, J. & Gierasch, L.M. (1994) Biochem. 33, 134-142.

Logan, T.M., Theriault, Y. & Fesik, S.W. (1994) J.Mol.Biol. 236, 637-658.

Norwod, T.J., Crawford, D.A., Steventon, M.E., Driscoll, P.C. & Campbell, I.D. (1992) Biochem. 31, 6285-6290.

Palmer, A.G., Hochstrasser, R.A., Millar, D.P., Rance, M. & Wright, P.E. (1993) J.Am.Chem.Soc. 115, 6333-6345.

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

Peng, Z., & Kim, P.S. (1994) Biochem. 33, 2136-2141.

Radford, S.E., Dobson, C.M. & Evans, P.A. (1992) Nature 358, 302-305.

Rizo, J., Blanco, B.K., Bruch, M.D. & Gierasch, L.M. (1993) Biochem. 32, 4881-4894.

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

Sönnichsen, F.D., Van Eyk, J.E., Hodges, R.S. & Sykes, B.D. (1992) Biochem. 31, 8790-8798.

Sosnick, T.R., Mayne, L., Hiller, R. & Englander, S.W. (1994) Nature: Struct.Biol 1, 149-156.

Udagaonkar, J.B. & Baldwin, R.L. (1990) Proc.Natl.Acad.Sci. (USA) 87, 8197-8201.

Wüthrich, K. (1994) Curr.Opin.Struct.Biol. 4, 93-99.

Zhong, L. & Johnson, W.C. (1992) Proc.Natl.Acad.Sci. (USA) 89, 4462-4465.

Publications

Stefani, M., Cappugi, G., Pazzagli, L., Camici, G., Manao, N., Taddei, N., Buck, M. & Ramponi, G. (1991) "Preparation and properties of *des*-Tyr^{98} and *des*-Arg^{97}-Tyr^{98} acylphosphatase" Int.J.Pep.Prot.Res. 38, 278-284.

Radford, S.E., Buck, M., Topping, K.D., Dobson, C.M. & Evans, P.A. (1992) "Hydrogen exchange in native and denatured states of hen egg-white lysozyme" Proteins: Struct.Funct. & Genetics 14, 237-248.

Buck, M., Radford, S.E. & Dobson, C.M. (1993) "A partially folded state of hen egg-white lysozyme in trifluoroethanol: Structural characterisation and implications for protein folding" Biochem. 32, 669-678.

Buck, M., Radford, S.E. & Dobson, C.M. (1994) "Amide hydrogen exchange in a highly denatured state: hen egg-white lysozyme in urea" J.Mol.Biol. 237, 247-254.

Taddei, N., Buck, M., Broadhurst, R.W., Stefani, M., Ramponi, G. & Dobson, C.M. (1994) "Equilibrium unfolding studies of horse muscle acylphosphatase" Eur.J.Biochem. 225, 811-817.

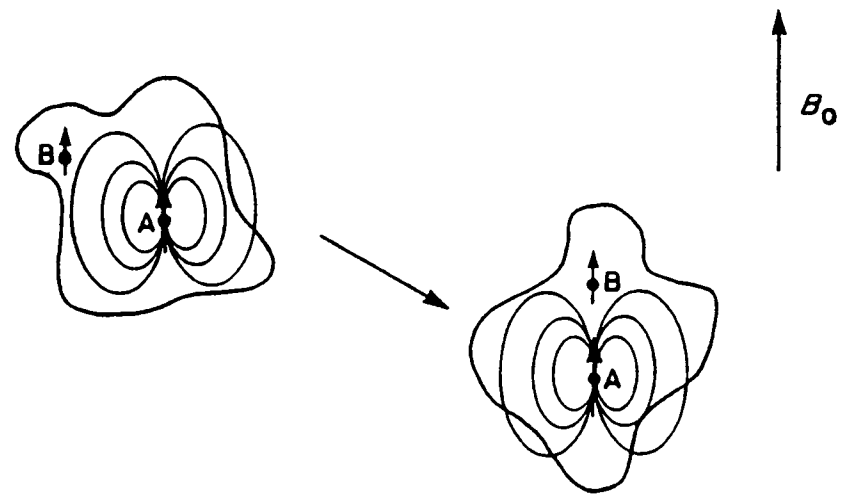
Buck, M., Boyd, J., Redfield, C., MacKenzie, D.A., Jeenes, D.J., Archer, D.B. & Dobson, C.M. (1995) "Structural determinants of protein dynamics: Analysis of ^{15}N relaxation measurements for maichain and sidechain nuclei of hen egg-white lysozyme" accepted for publication in Biochemistry.

Appendix I

Theory of ^{15}N relaxation behaviour and methods of data analysis

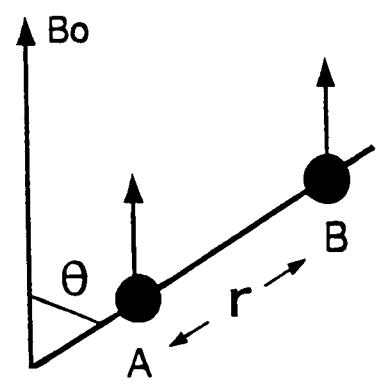
The process of NMR relaxation is intimately related to molecular motion. This relationship exists because transitions giving rise to relaxation in NMR are not spontaneous but must be stimulated by suitable changes of magnetic fields. These arise, for example, as a consequence of dipole-dipole interactions. The magnetic moment or dipole $\mu=\gamma J$; where γ is the gyromagnetic ratio of the nucleus and J its angular momentum, produces a small magnetic field which interacts with the static magnetic field B_0 of the spectrometer and with other dipoles in its neighbourhood. Figure 1 (adapted from London, 1989) illustrates, how changes in the magnetic field arise as a consequence of the interactions of the nuclear magnetic moments of nuclei and the motion of the molecule; here a global tumbling leads to a reorientation of of nucleus A relative to B.

Figure 1



Local magnetic fields are oriented by the external field (B_0). Their relative position depend on the location of nuclei in the molecule and the strength of their interaction depends on the relative orientation of the fields with respect to each other. So for example in the case of dipole-dipole interactions, the local field at B (H_L) arises from a neighbouring dipole, μ_A, at distance r as

$$H_L = \mu(3\cos^2\theta - 1)/r^3$$

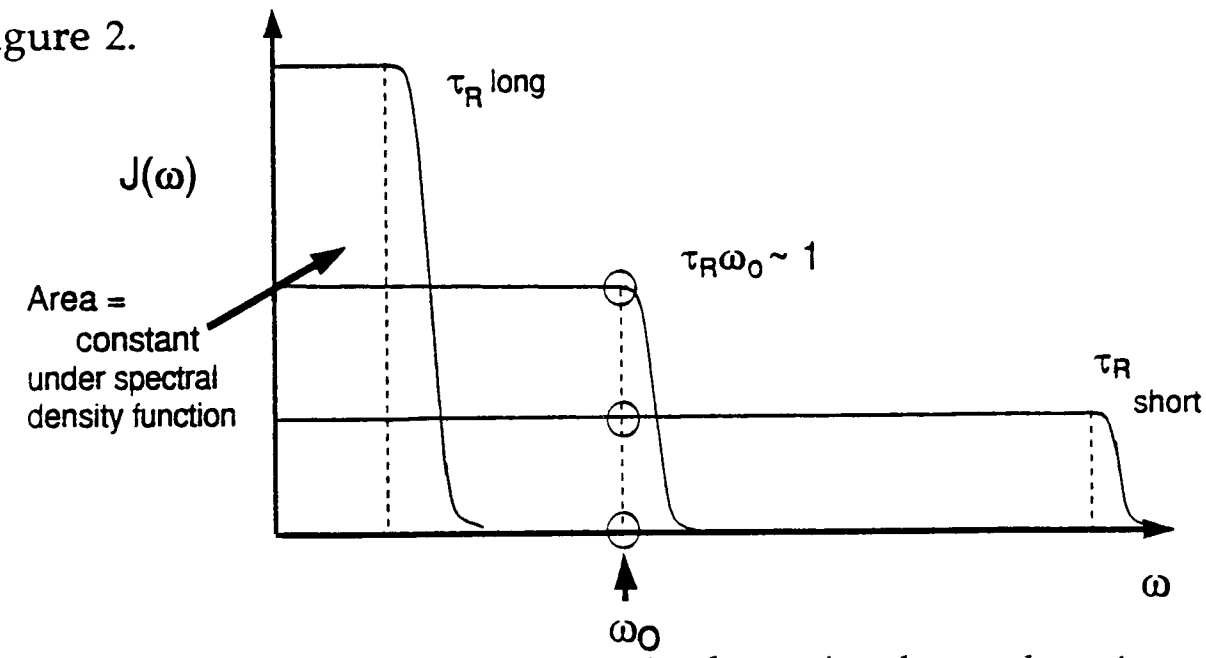


Central to relaxation is the concept of correlation. This arises from the need to model the random motions of molecules in solution. In the case above, the overall rotation, will lead to a change in the field observed at nucleus B. If r is fixed but θ changes with time, one can separate $\mathbf{H}_L = F[(t)]/r^3$. A correlation function C(t) determines how rapidly the vector connecting spins A and B changes orientation and can be calculated on the basis of models for the motion, as diffusion or site-jump processes. For relaxation through dipolar coupling, here, we consider the simplest model of global motion as stochastic tumbling of the molecules.

A correlation function $C(t) = \langle r^{-6} \rangle F(t)F(t+\tau)$, describes average rate of change of local field $\mathbf{H}_L$. In general C(0)=1 and C(t) -> 0 as t increases beyond some time τ_c; i.e. after a sufficient interval of time τ has elapsed there is no correlation at that time with the initial orientation (eg. $\mathbf{H}_L$). Fourier Transform of C(t) yields the spectral density function $J(\omega)$ which describes the frequency spectrum of the motion. So in the case of a rigid body tumbling isotropically the correlation and the spectral density functions are given by

$$C(t) = \exp(-t/\tau_R) \text{ and } J(\omega) = 2\tau_R/(1+\omega^2\tau^2_R) \text{ respectively.}$$

Thus the stochastic motions can be summarised by a single parameter - τ_R - the rotational molecular correlation time. τ_R corresponds to the shortest "waiting" time between reorientations (i.e. shorter rarely occur). The form of the function is shown in Figure 2.



So, if a field at frequency ω_0 is absolutely required to stimulate relaxation of a nucleus, we consider the magnitude of $J(\omega_0)$ which in this model depends on τ_R. Various scenario arise:

τ_R short	: $J(\omega_0)$ small, relax. slow
$\tau_R\omega_0 \sim 1$	: $J(\omega_0)$ large, fast relax.
τ_R long	: $J(\omega_0)$ v.small, relax v.slow.

From another point of view, a short τ_R means that very rapid motions (eg. tumbling of small molecules) average the dipolar interactions and lead to

inefficient relaxation. By contrast, τ_R long (e.g. for a large molecule) results in very efficient relaxation, provided τ_R is not too long that ω_0 falls outside the range in which molecular motion occurs (e.g. T_1 relaxation in solid state NMR is slow).

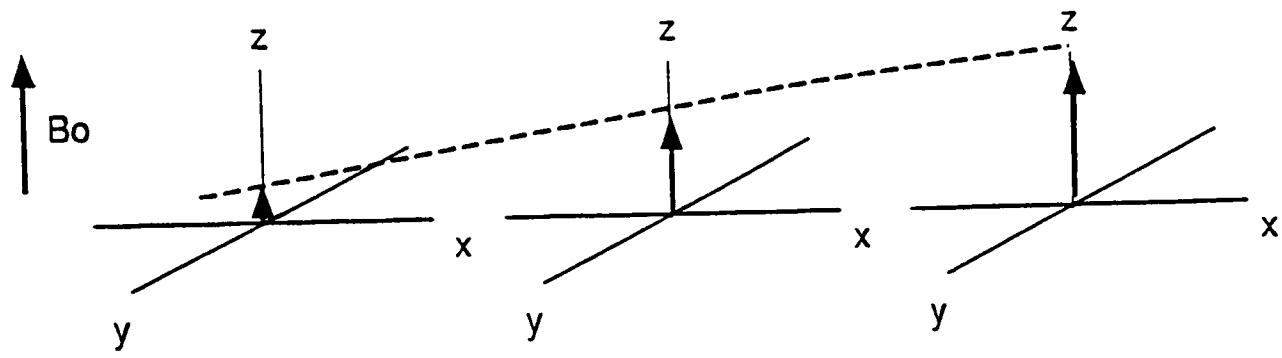
Relaxation mechanisms.
A number of mechanisms can contribute to relaxation (Abragam, 1961):
Dipolar interactions can be intramolecular (see above), but also intermolecular, for example arising from interactions with dissolved paramagnetic species.
Chemical shift anisotropy (CSA) arises as a consequence of the dependence of the chemical shift tensor of a particular nucleus on the molecular orientation in the applied field. This mechanism is important for some nuclei (^{31}P-^{1}H, ^{13}C-^{1}H), but for ^{15}N-^{1}H it contributions typically 20% to the overall relaxation.
Exchange effects arise as chemical (or hydrogen) exchange and "slow" physical exchange between different conformations involving a nucleus.
Scalar relaxation is possible between coupled spins by chemical exchange or spin diffusion.

Careful design of NMR experiments aims to minimise or account for the possible contributions to relaxation. Ideally the relaxation contributions should be expressed as a sum of rates for each possible relaxation interaction, as otherwise cross-correlations make relaxation multi-exponential (e.g. involving several spins).

This can be done in the case of **^{15}N relaxation** better than in the case of **^{1}H relaxation**. Firstly, relaxation of ^{15}N by dipolar interaction is predominantly with its attached ^{1}H, and the average distance <r> is assumed to be constant. By contrast, many neighbouring ^{1}H dipoles exist at variable distances and contribute to ^{1}H relaxation. Secondly, the interacting spins (i.e. ^{1}H) can be decoupled in ^{15}N relaxation experiments, by contrast to ^{1}H relaxation studies which are carried out with coupled spins. This results in significant cross-relaxation effects such as spin-diffusion, especially for proteins. In such cases relaxation is affected by the strength of interaction with the various ^{1}H coupling partners and their dynamics. This means relaxation is no longer necessarily a measure of motions at a single nuclear site.

Experimentally measurable relaxation parameters (most commonly T_1, T_2 and the heteronuclear NOE) are functions of $J(\omega)$, the Fourier Transform of $C(t)$, at five characteristic frequencies for a 2-spin system, $J(0)$, $J(\omega_H)$, $J(\omega_N)$, $J(\omega_H+\omega_N)$ and $J(\omega_H-\omega_N)$. Provided there are no contributions to relaxation due to cross-correlations, then the relaxation parameters can be expressed as the sum of the dipolar and CSA contributions.

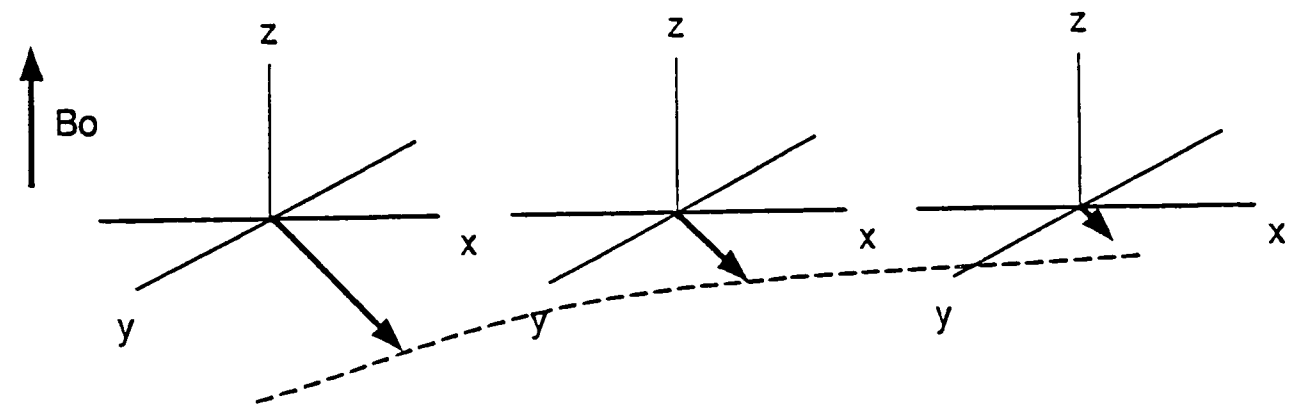
T_1 - **longitudinal relaxation** is the exponential recovery of magnetisation along z-axis (starting magnetisation $Iz = Iz_o$) with rate $1/T_1$.



$$1/T_1 = (d^2/4)[J(\omega_H-\omega_N) + 3J(\omega_N) + 6J(\omega_H+\omega_N)] + c^2J(\omega_N) \qquad (1)$$

Thus T_1 relaxation is dominated by $J(\omega_N)$ and is
sensitive to motions with frequency 10^8- 10^{12} s^{-1}.

T_2 - **transverse relaxation** is the exponential decay of magnetisation in the x-y plane with rate $1/T_2$.

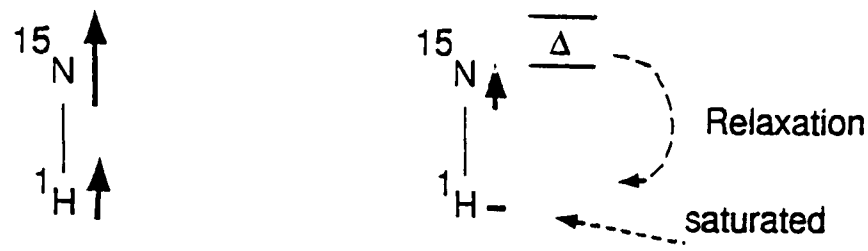


$$1/T_2 = (d^2/8)\,[4J(0) + J(\omega_H-\omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6(\omega_H+\omega_N)] + (c^2/6)\,[3J(\omega_N) + 4J(0)] + Rex \qquad (2)$$

T_2 relaxation is dominated by $J(\omega_N)$ and $J(0)$
sensitive to motions with frequency 10^8-10^{12} s^{-1}
but also to exchange processes with frequency 10^4- 10^6 s^{-1}.

Heteronuclear NOE is the intensity change observed for ^{15}N when its attached ^{1}H is saturated, measuring the relative contribution of cross relaxation with ^{1}H to the relaxation of spin ^{15}N.

NOE = 1 + (γ_H/γ_N) (σ/ρ); where σ is the cross-relaxation term and ρ is $1/T_1$.
In case of the hetonuclear NOE for ^{15}N nuclei its value is less than 1.0 because the gyromagnetic ratios of the ^{15}N and ^{1}H nuclei have different signs.



$$NOE = 1 + T_1(\gamma_H/\gamma_N)(d^2/4)\,[\,6J(\omega_H+\omega_N) - J(\omega_H-\omega_N)] \qquad (3)$$

The NOE is dominated by the difference of $J(\omega_H+\omega_N)$ and $J(\omega_H-\omega_N)$ and is therefore sensitive to a range of the highest frequency motions.

In equations 1-3

$$d = (\mu_o h/8\pi^2)\,\gamma_H\,\gamma_N\,\langle r_{NH}^{-3}\rangle$$

$$c = \omega_N\,(\sigma_{\parallel} - \sigma_{\perp})/\sqrt{3}$$

where

ω_i , Larmor freq. of nucleus * 2π, eg. 500 MHz ^{1}H & 50 MHz ^{15}N

γ_i , gyromagnetic ratios

$\langle r_{NH}\rangle$, 1.02 Å

$\sigma_{\parallel} - \sigma_{\perp}$, -160 p.p.m. (ax. symm. chem. shift tensor) for backbone ^{15}N-^{1}H

h , Plank's constant

μ_o , permeability of free space

Rex , is an line-broadening contribution to $1/T_2$ which arises from conformational averaging on a time scale of 30 ns to 10 ms.

The three measurable experimental parameters are generally insufficient to determine the spectral density function $J(\omega_i)$ at the five frequencies explicitly. There are several approaches to the determination of the spectral density function. Firstly, the values of $J(\omega_H+\omega_N)$, $J(w_H)$ and $J(\omega_H-\omega_N)$ can be assumed to be approximately equal, in the case of ^{15}N relaxation, given that $\omega_H \sim 10^* \omega_N$ (Peng & Wagner, 1992). If Rex makes negligible contributions to $1/T_2$ (i.e. $1/T_2 >> Rex$), the expressions (1-3) above can be reduced to equations, which can be solved to give the spectral density function at three frequencies (0, ω_N, ω_H) in terms of T_1, T_2 and NOE (see Chapter 8)

$$J(\omega_H) = 0.2^* \, 1/T_1 \, (NOE - 1) \, (4/d^2) \, (\gamma_N/\gamma_H) \qquad (4)$$

$$J(\omega_N) = 4/(T_1(3d^2 + 4c^2)) \, [1- (7/5)(NOE - 1)(\gamma_N/\gamma_H)] \qquad (5)$$

$$J(0) = 1/(3d^2 + 4c^2) \, [(6/T_2) - (3/T_1)- (18/5) \, (1/T_1) \, (NOE - 1)(\gamma_N/\gamma_H)] \qquad (6)$$

Secondly, the spectral density function may be determined explicitly at the five frequencies by the spectral density mapping approach of Peng & Wanger (1992). However, this requires additional and alternative measurements of experimental relaxation parameters and there are some uncertainties as to whether all the parameters can be measured with sufficient accuracy.

A third approach is to model the motion of the protein and to derive an expression of the spectral density function which contains parameters determined by the experimental data. This latter approach was already presented above as the simplest case of the spectral density function derived for a rigid sphere tumbling stochastically in solution.

$$J(\omega) = 2\tau_M/(1 + \omega^2\tau^2_R). \qquad (7)$$

The term τ_R, the molecular rotational correlation time, is the time constant of the exponentially decaying correlation function,

$$C(t) = \exp(-t/\tau_R); \qquad (8)$$

which is the inverse Fourier transform of the spectral density function.
τ_R is a physically realistic parameter which may be calculated from the equations describing the hydrodynamic behaviour of particles and can be measured by other experimental techniques (e.g. fluorescence anisotropic decay). A value of τ_R may be obtained from NMR on the basis of the relaxation parameters, because the ratio of T_1/T_2 depends principally on τ_R for sites of the protein which are undergoing no significant internal motions. In this case the dependence of the relaxation parameters on the correlation time τ_R is as shown in Figure 3 (adapted from Kay et al., 1989).

Figure 3

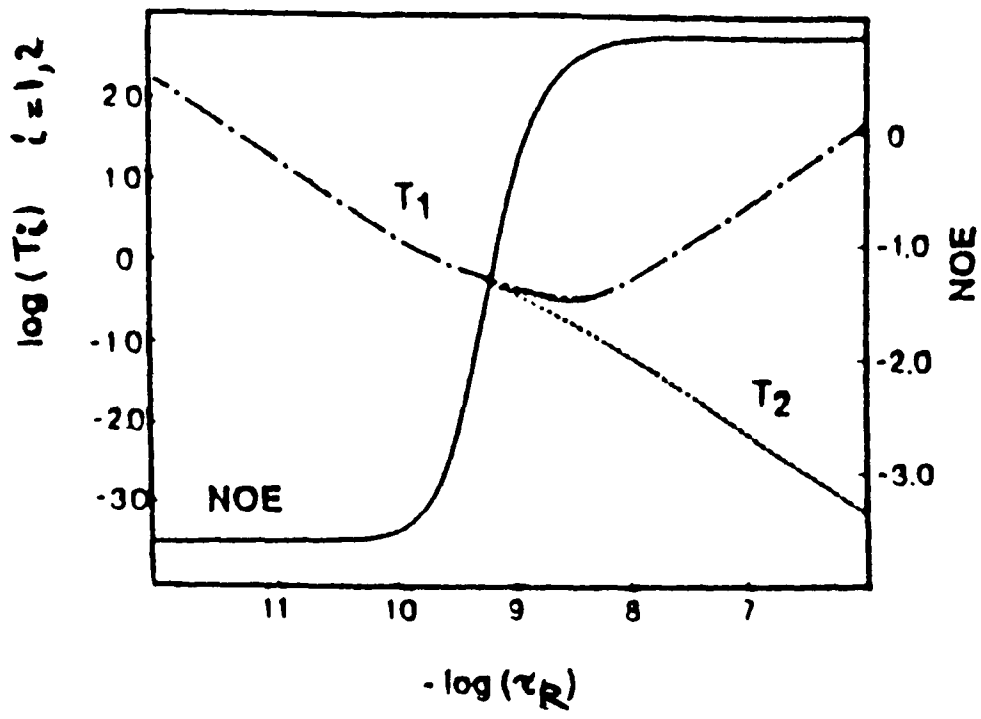
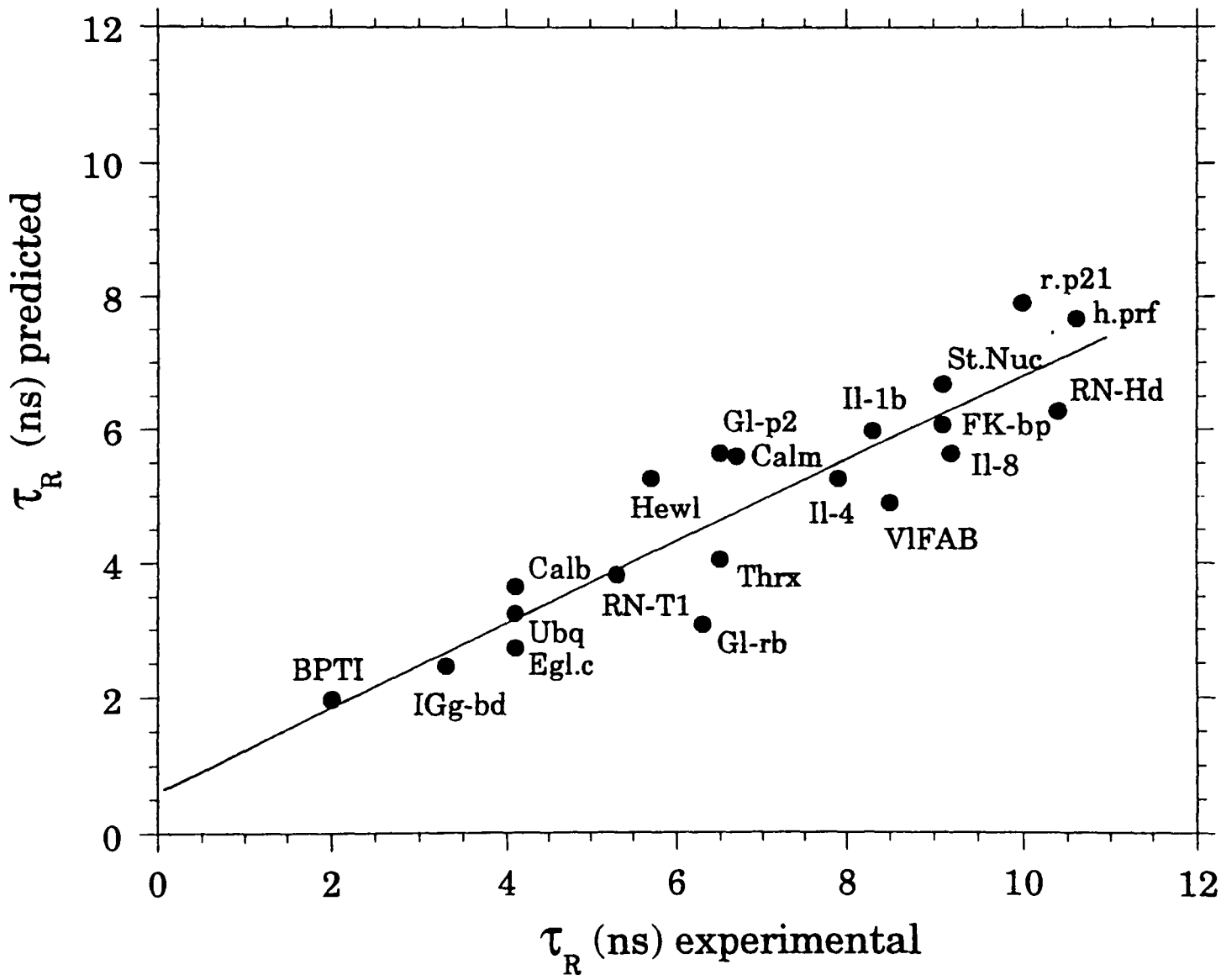


Figure 4



A comparison between the τ_R values determined for a number of proteins from ^{15}N relaxation studies and their values predicted on the basis of hydrodynamics calculation is shown in Figure 4. Values for τ_R obtained for a number of proteins from ^{15}N relaxation studies. These are **BPTI** (Szyperski et al., 1993), **IgG-b**inding **d**omain (Barchi et al., 1994), **Eg**lin **c** (Peng & Wagner, 1992), **Ubi**quitin (Schneider et al., 1992), **Calb**indin (Kördel et al., 1992), **RN**ase-**T1** (Fushman et al., 1994), **Hewl** - this study, **G**lucocorticoid receptor DNA **b**inding protein (Berglund et al., 1992), **Thio**redoxin (Stone et al., 1993), **FK**506-**bp** (Cheng et al., 1993), **Vl** of **FAB** (Freud et al., 1994), **IL-8** (Grasberger et al., 1993), **IL-4** (Redfield et al., 1992), **IL-1**β (Clore et al., 1990), **Calm**odulin -average of both domains (Barbato et al., 1992), **G**lucose **perm.II** (Stone et al., 1992), **Staph.Nuc**lease (Kay et al., 1989), **ras p21**.GDP (Kraulis et al., 1994), **RiboNuclease-H d**omain of human HIV reverse transcriptase (Powers, et al., 1992), **human profilin** (Constantine et al., 1993). The rotational correlation time can also be estimated on the basis of well known hydrodynamic equation. τ_R pred. = $V\eta/k_BT$ (Debye equation); where k_B is Boltzman's constant; T is Temp. in K and V is hydrated volume (in m^3) and η is viscosity (in $kgm^{-1}s^{-1}$). However, the effect of the different protein volumes, concentrations and temperatures used in these studies must be incorporated as $V = M_r/N_A (\upsilon_p + \omega/\rho_o)$ (Cantor & Schimmel, 1980) where M_r in g/mol-1, N_A Avogadro's No., υ_p partial specific volume of proteins, typically 0.72 cm^3/g; ρ_o - density of water taken as 1.0 g/cm^3. ω - mean hydration of proteins, taken as 0.4 g H_20/g protein. $\eta = \eta_o(1 + [\eta]C)$. where η_o is the viscosity of water at a given T (eg. 1.0, 0.87 and 0.69 cP at 20, 26, 37°C respectively - 1 cP = 10^{-3} $kgm^{-1}s^{-1}$); C is the protein concentration in g/cm^3 and $[\eta]$ the intrinsic viscosity of protein calculated as $[\eta] = \nu(\upsilon H_20 + \upsilon_p\omega)$ (Cantor & Schimmel, 1980) where ν is the protein shape, assumed here 2.5 for sphere; υ is partial specific volume of H_20 taken as 1.0 cm^3/g. υ_p and ω are as above. The linear correlation between experimental and predicted τ_R is excellent (Rc = 0.91). Using the above equations, however experimental τ_R may be estimated as τ_R estm. = 1.63 (τ_R pred. - 0.67); the RMS difference is 0.69ns. The value of τ_R for hen lysozyme is found to be similar to, albeit slightly lower than for other proteins of similar molecular weight. It is apparent that other factors, for example the presence of long, flexible loops, differences in protein hydration, shape and intramolecular association may result in values of τ_R which are increased over those obtained from theoretical estimates.

Description of internal motions in proteins by the modelling of spectral density functions.

It is apparent from the variation of relaxation parameters along the polypeptide chain, typically observed for proteins, that these molecules are not rigid spheres, but experience often considerable internal dynamics. This internal motion can in principle be modelled explicitly, analogously to the derivation of a correlation function for the isotropic tumbling of a rigid sphere. However, a more general and mathematically rigorous approach has been presented by Lipari & Szabo (1982) which does not make any assumptions about the precise physical processes and the geometry which describe the internal motions, but instead models the correlation function as a series of independent contributions,

$$C(t) = C_O(t)\, C_i(t) \qquad (9)$$

where $C_O(t)$ is the contribution of the overall rotational motion, defined above (equation 8) and

$$C_i(t) = S^2 + (1 - S^2)*\exp(-t/\tau_e) \qquad (10)$$

describes a faster motion with correlation time, τ_e, and an order parameter (S^2) which gives the relative contributions of the overall and the internal motions to the correlation function (Figure 5.1 below). It has values between 0 and 1 and is taken as a general measure of the geometrical restriction of the NH bond vector, relative to the remainder of the molecule. By contrast to explicit models for the dynamic processes, the parameters have no direct physical meaning themselves (see below). The formalism is therefore said to be model free.
The spectral density function corresponding to equation 9 is

$$J(\omega_i) = S^2\tau_R/(1 + \omega_i^2\tau_R^2) + (1 - S^2)\tau/(1 + \omega_i^2\tau^2) \qquad (11)$$

overall internal motion

where $\tau^{-1} = \tau_R^{-1} + \tau_e^{-1}$

τ_R - correlation time for global molecular motion
τ_e - correlation time effective internal motion
S^2 - order parameter

In the case that the internal motions are fast (τ_e < 20 ps) this expression simplifies to

$$J(\omega_i) = S^2\tau_R/(1 + \omega_i^2\tau_R^2) \qquad (12)$$

Clore et al., (1990) introduced an extension of the Lipari & Szabo formalism in which the correlation function is subfactored into internal motions occurring on two time-scales.

$$C_i(t) = S_S^2S_f^2 + S_f^2(1 - S_S^2)\exp(-t/\tau_S) + (1 - S_f^2)\exp(-t/\tau_f) \qquad (13)$$

The corresponding spectral density is now expressed as a function of three motions; a global motion, a slow and a fast internal motion.

$$J(\omega_i) = S_S^2S_f^2\tau_R/(1 + \omega_i^2\tau_R^2) + S_f^2(1 - S_S^2)\tau_S'/(1+\omega_i^2\tau_S'^2) + (1-S_f^2)\tau_f'/(1+\omega_i^2\tau_f'^2)$$

(14)

where $\tau_i'^{-1} = \tau_i^{-1} + \tau_R^{-1}$; i= s or f

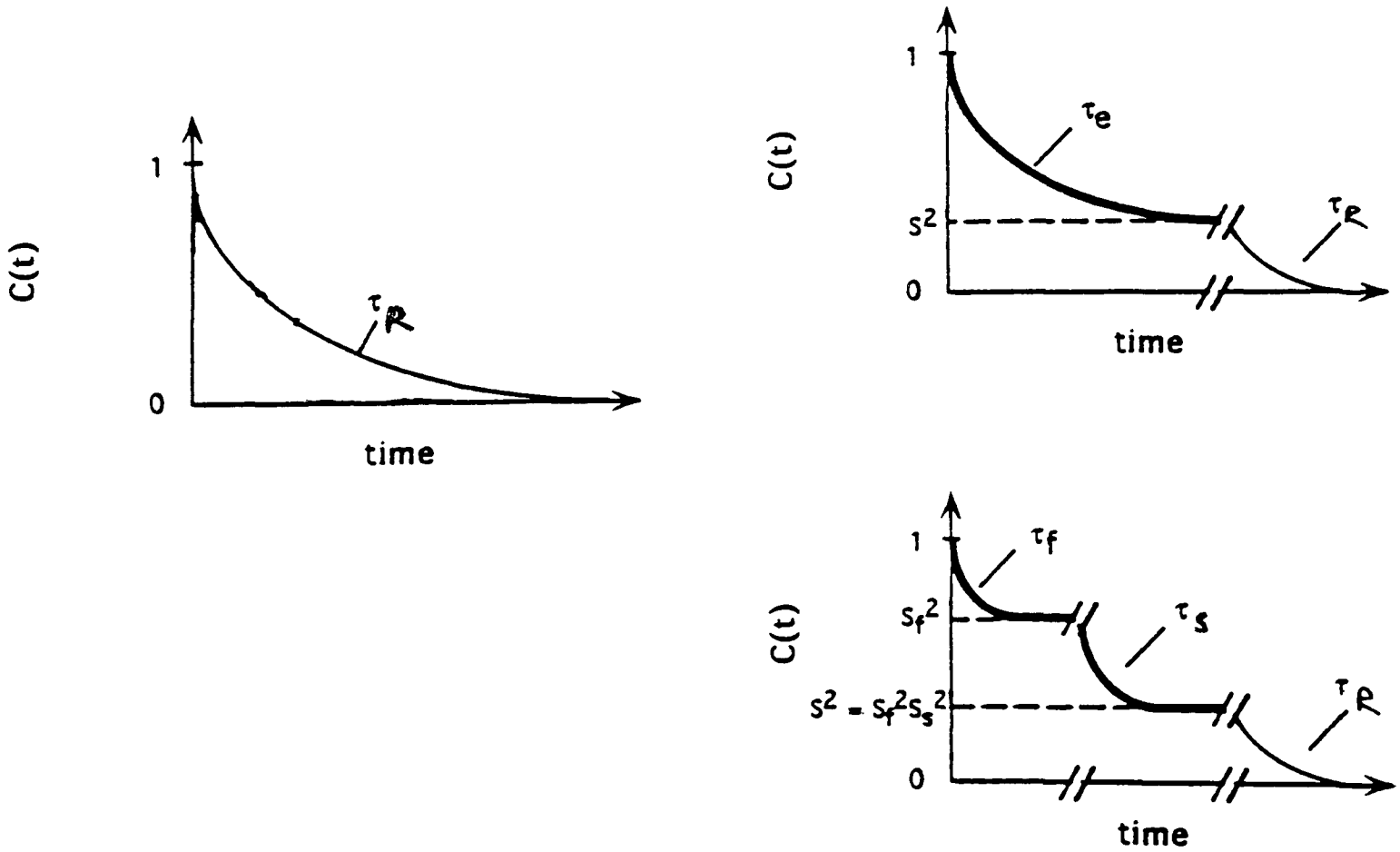
under conditions where $\tau_f < 20$ ps equation 14 reduces to

$$J(\omega_i) = S_f^2S_S^2\tau_R/(1 + \omega_i^2\tau_R^2) + S_f^2(1-S_S^2)\tau/(1 + \omega_i^2\tau^2) \qquad (15)$$

where $\tau^{-1} = \tau_R^{-1} + \tau_S^{-1}$

$$S^2 = S_f^2S_S^2 \qquad (16)$$

The correlation functions described by the models above are pictured together with the order parameters (S^2) and time constants (τ) in Figure 5 adapted from Alexandrescu & Shortle (1994). The breaks indicate the change in scale along the time axis: $\tau_R > \tau_S > \tau_e >> \tau_f$.



Given the experimental values T_1, T_2 & NOE and their relationship to the spectral density function $J(\omega_i)$ (equations 1-3), parameters in the model of the spectral density function (τ_e, S^2 etc.) may be optimised to best reproduce the experimental data by minimisation of the rms difference (eg. Palmer et al., 1991)

$$\Delta = [(T_1^e - T_1^c / \sigma T_1^e)^2 + (T_1^e - T_2^c / \sigma T_2^e)^2 + (NOE^e - NOE^c / \sigma NOE^e)^2]^{1/2} \quad (17)$$

where superscripts e and c denote the experimental data and calculated values, respectively and σ is the experimental error in the measurement.

The selection of the appropriate model of the spectral density function to describe the dynamics of a ^{15}N nucleus depends on the relaxation data. The value of the ratio of T_1/T_2, relative to that obtained for sites which undergo no significant internal motions and the magnitude of the NOE are commonly used as selection criteria (Clore et al., 1990). This is described in chapter 8.

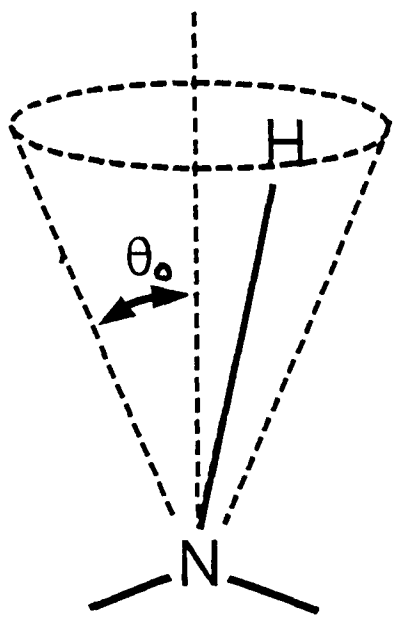
Physical Interpretation of S^2 (S_s^2, S_f^2), τ_e (τ_s) and Rex

The approach of Lipari & Szabo (1982) can be considered as modelfree because it does not involve geometric constraints as would be required for a physically explicit model. It follows that the meaning of the derived parameters is limited in a physical sense.

S^2 is a measure of the degree of spatial restriction of the motion.
τ_e, the effective correlation time, is a measure of the rate of the motion. However, both the geometrical restriction, the actual rate depends on the model to be used and their relationship to the derived parameters (S^2, τ_e etc.). In general any physical model that reproduce these parameters could be valid and the NMR relaxation data cannot discriminate between the alternative models.

Several physical models have been proposed for the motions of mainchain N-H vectors, including the following.

"Diffusion in cone" model (e.g. see Lipari & Szabo, 1982; Kördel et al., 1992)



$S_{cone} = 0.5 (\cos \theta_o)(1+ \cos\theta_o)$

can always find semiangle θ_o

$\theta_o = \cos^{-1} [0.5((1 + 8S_{cone})^{1/2} - 1)]$
where $0^o < \theta_o < 90^o$

$\tau_D = 1/D_w$ microscopic Wobbling diffusion constant

$(1-S^2)\tau_e = 6\tau_D$ (complicated function of[cos θ_o])
(Kördel et al., 1992)

for example for several residues of the native state the following values are obtained

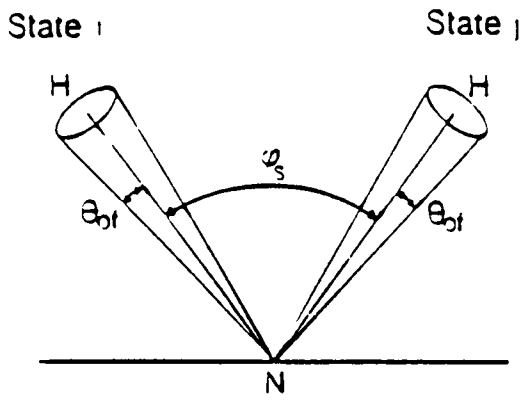
Gln 57 NH $S^2 = 0.94 \Rightarrow \theta_o = 12^o$
Ala 11 NH $S^2 = 0.89 \Rightarrow \theta_o = 16^o$
Ser 85 NH $S^2 = 0.54 \Rightarrow \theta_o = 36^o$

Alternatively, the motion may be considered as a vibrational lengthening of the NH bond. (Clore et al. 1989)

$S^2 \sim (R/r_{eff})^6$ where $R = \langle r_{NH} \rangle$ taken as 1.02 Å

This would result in an increased bond length of 1% and 11% for the NH of Gln57 and Ser85 respectively.

A two-site jump superimposed on the fast diffusion in a cone, has been used as a model to provide a physical depiction of the two internal motions described by the extended model of Lipari & Szabo (equation 14), assuming that S_S^2 and S_f^2 are totally independent.



$S_S^2 = (1+3\cos^2\varphi)/4$
where φ is the angle between the two orientations of the NH bond, which occur in equal populations (figure adapted from Clore et al., 1990)

for example residue Asn103 in the native state of hen lysozyme for

$S_S^2 = 0.63 \quad => \varphi = 45^o$

$S_f^2 = 0.83 \quad => \theta_o = 20^o$

so the effective reorientation $(0.5\psi + \theta_o)$ is 43°

The actual motions underlying the relaxation of the ^{15}N nucleus are likely to be in many instances a combination of several types of conformational transitions. Molecular dynamics simulations increasingly allow the study of these transitions in unprecedented atomic and temporal resolution. Several, trajectories have been used to derive correlation functions and order parameters, and to compare the results with the experimental data (Chandrasekhar et al., 1992; Palmer & Case; 1992; Erickson et al., 1993). While, there is broad agreement in some cases, the detailed comparison is as yet limited by the short duration of the simulations and the poor sampling of infrequent events.

References

Abragam, A. (1961) "Principles of Nuclear Magnetism" , Oxford University Press.

Barbato, G., Ikura, M., Kay, L.E., Pastor, R.W. & Bax, A. (1992) Biochem. 31, 5269-5278.

Barchi, J.J., Grasberger, B., Gronenborn, A.M. & Clore, G.M. (1994) Protein Science 3, 15-21.

Berlund, H., Kovacs, H., Dahlman-Wright, K., Gustavson, J.-A. & Hard, T. B. (1992) Biochem. 31, 12001-12011.

Cantor, C.L. & Schimmel, P. (1980) "Biophysical Chemistry Vol II" Freeman & CO, San Francisco.

Chandrasekhar, I., Clore, M.G., Gronenborn, A.M. & Brooks, R. (1992) J.Mol.Biol. 226, 239-250.

Cheng, J.W., Lepre, C.A., Chambers, S.P., Fulghum, J.R., Thomson, J.A. & Moore, J.M. (1993) Biochem 32, 9000-9010.

Clore, G.M., Driscoll, P.C., Wingfield, P.T. & Gronenborn, A.M. (1990) Biochem. 29, 7387-7401.

Constantine, K.L., Friedrichs, M.S., Goldfarb, V., Jeffrey, P.D., Sherriff, S. & Mueller, L. (1993) Proteins: Struct.Funct. & Genetics 15, 290-311.

Derome, A.E. (1988) "Modern NMR Techniques for Chemistry Research", Pergamon Press, Oxford.

Freund, C., Ross, A., Plückthun, A. & Holak, T.A. (1994) Biochem. 33, 3296-3303.

Fushman, D., Weiseman, R., Thuring, H. & Rüterjans, H. (1994) J.Biomol.NMR 4, 61-78.

Grasberger, B.L., Gronenborn, A.M. & Clore, G.M. (1993) J.Mol.Biol. 230, 364-372.

Kay, L.E., Torchia, D.A. & Bax, A. (1989) Biochem. 28, 8972-8979.

Kördel,J., Skelton, N.J., Akke, M., Palmer, A.G. & Chazin, W.J. (1992) Biochem. 31, 4856-4866.

Kraulis, P.J., Domaille, P.J., Campbell-Burk, S.L., van Aken, T. & Laue, E.D. (1994) Biochem. 33, 3515-3531.

Lipari, G. & Szabo,A. (1982) J.Am.Chem.Soc. 104, 4559-4570.

London, R.E. (1989) Meth.Enzym. 176, 358-375.

Palmer, A.G. & Case,D.A. (1992) J.Am.Chem.Soc. 114, 9059-9067.

Palmer, A.G., Rance, M. & Wright, P.E. (1991) J.AmChem.Soc. 113, 4370-4380.

Peng, J.W. & Wagner, G. (1992) Biochem. 31, 8571-8586.

Powers, R., Clore, G.M., Sahl, S.J., Wingfield, P.T. & Gronenborn, A. (1992) Biochem. 31, 9150-9157.

Redfield, C., Boyd, J., Smith, L.J., Smith, R.A.G & Dobson, C.M. (1992) Biochem. 31, 10431-10437.

Schneider, D.M., Dellwo, M.J. & Wand, J.A (1992) Biochem. 31, 3645-3652.

Stone, M.J., Chandrasekhar, K., Holmgren, A., Wright, P.E. & Dyson, H. J. (1993) Biochem. 32, 426-435.

Stone, M.J., Fairbrother, W.J., Palmer, A.G., Reizer, J., Saier, M.H. & Wright, P.E. (1992) Biochem. 31, 4394-4406.

Szyperski, T., Luginbühl, P., Otting, G., Günert, P., & Wüthrich, K. (1993) J.Biomol.NMR 3, 151-164.

Wüthrich, K. (1986) "NMR of Proteins and Nucleic Acids,", Wiley, N.Y.

Appendix II

Preliminary Structure Calculation of the TFE Denatured State of Hen Lysozyme.

Objective

This appendix presents the approach used to compute an ensemble of structures for the TFE denatured state from the NMR restraints and summarises the results of these calculations. As potential long range NOEs could not be assigned unambiguously, structures were calculated using only medium and short range constraints, except for the disulphide cross-bridges, which were treated as covalent bonds. The aims of the calculations and the analysis of the resulting structures were several:

1) Calculation of structures and assessment of their agreement with input restraints.
2) Analysis of the structural propensities of the ensemble and comparison with other data on the TFE state, obtained principally from ^{15}N relaxation.
3) Assessment of the extent to which disulphides and the secondary structures define the overall tertiary topology of the denatured state in the absence of long range restraints.

Input parameters

A total of 1402 NOEs were used as input restraints in ranges 1.8-3.2 Å (strong), 1.8-3.8 Å(medium strong), 1.8-4.3 Å (medium) and 1.8-5.5 Å (weak), defined on the basis of their intensity relative to the autocorrelation peak. Pseudoatoms with corrections (Wüthrich et al., 1983) were used for aromatic rings and for other sets of equivalent and non-stereospecifically assigned protons. In addition, a further 0.5 Å was added to the upper limits in the case of NOEs involving methyl groups (Wüthrich, 1986). The NOEs were distributed as follows: 1159 NOEs involving mainchain NH , 105 NOEs involving NH_2 groups of asparagine and glutamine and 138 NOEs involving tryptophan ring protons (see further classification in Table 1). A plot of the number of NOEs versus the sequence is shown in Figure 1a,b. In addition to the above, the NOE list also included distance restraints for 48 CO_i-NH_{i+4} hydrogen bonds, set to 1.8±0.2 Å for NH-O and 2.8±0.3 Å for N-O distances. These were assigned for mainchain sites which have both significant protection against hydrogen exchange and are located in regions identified to be part of helical structures by the other NMR parameters. Similarly, 33 $^3J_{N\alpha}$ measurements were used as 27

α-helical ϕ angle restraints (-60±30°) and as ϕ angle restraints outside the β-region (60±150°) for 6 sites, which have $^3J_{N\alpha}$ values < 5.0 Hz but are not located in helix structures. An additional 31 α-helical ϕ restraints were used supplemented the data for sites, which are located in helices, particularly in the B-, C- and D-helices defined by the other NMR parameters, and for which no coupling constant could be measured.

Protocol for Structure Calculation

NMR structures were computed using the programme X-PLOR (Brünger, 1988) using a two-step simulated annealing protocol designed to ensure good global sampling of conformational space (Nilges et al., 1991). The first annealing run was initiated by generation of starting structures with random ϕ, ψ angles and a short energy minimisation, with high force constants, to ensure correct covalent stereochemistry. Initial force constants were set for NOE as 25 kcalmol^{-1}Å^{-2}; dihedral angle as 50 kcalmol^{-1}rad^{-2}. Bond lengths were fixed by the SHAKE algorithm (Rychaert et al., 1977) used in conjunction with reduced planarity and chirality force constants during the high temperature dynamics and cooling stages of the first annealing run. The system was subjected to dynamics for 15 ps at 1000K in presence of the experimental restraints. A soft potential well function was used as well as a low value for the van der Waals (vdW) repulsion, allowing atoms to pass through each other. Disulphide crosslinks were treated as NOEs throughout this first annealing run. The vdW and asymptotic force constants were increased stepwise from 0.1 to 1.0 kcalmol^{-1}Å^{-2} and from 0.002 to 0.1 kcalmol^{-1}Å^{-4} respectively during a further 1ps of dynamics. Cooling of the system from 1000 to 300K proceeded in steps of 25K over the next 2.8 ps with the NOE set as the normal square-well potential, and a vdW repulsion term of 4 kcalmol^{-1}Å^{-4}. This was followed by restrained Powell minimisation, with angle and improper force constants set to their nominal high values.

Disulphide cross-links of the polypeptide were incorporated as bonds at the start of a second annealing run. Resulting strain was removed by restrained Powell minimisation. In the course of a 3.75 ps dynamics simulation at 1000K, force constants for experimental restraints were increased stepwise from 0.5 to 50 kcalmol^{-1}Å^{-2} for NOE and from 0.3 to 200 kcalrad^{-2} for dihedral angles; together with the vdW force constant which was increased from 0.001 to 0.25 kcalmol^{-1}Å^{-4}. The system was then cooled as described above and energy minimised in the presence of the experimental restraints.

Agreement of structures with the NMR parameters

27 structures of the TFE state, resulting from the calculations above, were used for further analysis. Their rms deviation from NOE and covalent restraints is poor in comparison with parameters obtained during a similar calculation of the native state structure (Table 2). This reduced level of quality of structures for a denatured state may be anticipated for a number of reasons: The input data are likely to be highly incomplete in the present case, as no long range NOEs could be defined from ^{15}N edited spectra. Furthermore, NOEs are biased towards the closest distances protons approach each other and these are averaged over the conformations which coexist in solution. These different conformations, could in principle give rise to NOEs which range form sets which are mutually exclusive to highly complementary. The structure calculation programme tries to satisfy all these constraints at once within the same one structure. This may not possible in many instances and more refined procedures have to be used in order to calculate realistic structures, by either time- or ensemble averaging of the restraints (eg. Torda et al., 1990; Mierke et al., 1994). Such approaches have been successfully applied to structure determinations of small peptide and protein systems (eg. Smith et al., 1994).

Nevertheless, remarkably the average number of NOEs 69±7 which is violated in any of the structures by more than 0.5 Å is small, comprising less than 5% of the input data set. The sum total of the violating 1867 NOEs of the 27 structures, occur for 273 of the 1506 input NOEs, suggesting that the great majority of NOEs is satisfied by all the structures of the ensemble. Approximately half of these 273 NOEs are violated in less than any three of the structures; around a third of NOEs are violated in most of the structures. Noticeably, the majority of these NOEs involve tryptophan sidechains and it is possible that spin diffusion and the aforementioned conformational averaging may be responsible for the violations. These NOE violations account for approximately half of the average 69 violations seen in the structures. Amongst the other violations of NOEs, 6 on average occur for restraints which represent hydrogen bonds. These are found with variable frequency particularly at the N-termini of helices again suggesting additional conformations may be present. Improvements in the representation of the TFE state based on these NMR restraints could be made by using alternative protocols which take the dynamic nature of this state into account. Moreover the present structures have not been refined in the presence of forcefields appropriate for full molecular dynamics. While this has become routine for native proteins, causing significant improvements in the quality of the derived structures, it appeared not to be appropriate for refinement of the structure of

the TFE denatured state. This may be at least in part because of the effects of TFE on the structural energetics are neglected. However, the overall number of violations is remarkably low for a partially structured state given the limitations of the present procedure. This may in part be due to the fact that structure is well defined by the constraints at the local level, including α-helical secondary structure. There is good correspondence of the calculated ensemble of local secondary structures with the NMR data and this is further reinforced by comparison with a range of additional structural parameters.

Definition of secondary structure

The regional variation of structures is shown in Figure 2a-c, characterised by the variance of dihedral angles over the ensemble (MacArthur & Thornton, 1993), local rmsd (here over a sliding window of 3 residues) and the ensemble average surface accessibility of the mainchain. Significantly, the regions with lower than average values are highly coincident for each of the parameters considered, suggesting that they are all good indicators for local structure. Furthermore, there is good correspondence between regions of the polypeptide chain with low peptide bond accessibility and their dynamics, as described by the order parameter S^2 in Figure 2d. This relationship between structure and dynamics was apparent for the native state of hen lysozyme (Chapter 6) and is also likely to hold here.

The nature of these local structures can be identified on the basis of the values of ϕ and ψ along the polypeptide chain (Figure 3) and by direct superposition (Figure 4). Figure 3, shows that the majority of the ordered regions, correspond to stretches of the polypeptide in helical conformations ($\phi \sim -65^o$, $\psi \sim 50^o$). A superposition of the mainchain in the regions which form the principal helical conformation in the TFE state are shown Figure 4. It is noticeable that the extent of definition of α-helical structure (rmsd of superposition) is not identical for the different helices in the TFE denatured state and is poorer than that of α-helices in the native state (Table 3). Helix C and its N-terminal extension (res. 85-89) are not as clearly defined as helix A for example. These differences, are likely to be a direct result of the incompleteness of the NMR data, but also of the dynamics in these regions. The latter contributes substantial resonance broadening, particularly in the second half of the protein.

Overall topology of the TFE state

It is possible that disulphides and rigid α-helical secondary structure both constrain the overall topology of the polypeptide chain, even in the absence of long range NOE restraints. Higher order structures involving secondary

structures result, by definition, in well defined helix-helix angles. This is illustrated in the case of the ensemble of 15 structures, calculated by Smith et al., (1993) for the native state of hen lysozyme in Figure 5a. By contrast, helix-helix angles are not as well defined over the ensemble of structures for the TFE denatured state (Figure 5b). Significantly, the distribution of angles is not random, however, and there appears to be only a small number of helix-helix angles which are populated (rather akin to sidechain rotamers). For example the helices A and B are found frequently with a relative orientation of around 100° to each other. Intriguingly, each helix-helix angle distribution contains a significant area of overlap with the range of angles found in the native state. However, none of the 27 molecules has helix-helix angles identical to the native state and therefore higher order correlations are weak for many of the helices. This is shown in Figure 6.

Disulphide bond connectivities have a profound effect on the structure of the polypeptide chain and its dynamics (see main chapter). Significantly, the lengths of loops bounded by disulphides are unequal, thus constraining the polypeptide chain into two substructures in the denatured state. These regions residues 64-94 and residues 6-30-linked to residues 115-127 are shown schematically and as the superposition of the ensemble of structures of the TFE denatured state in Figure 7. Superposition of these local clusters with the native structure in these regions results in a large rms deviation of around 10 Å, and it is apparent that the native conformation is considerably more complex than the simple and disulphide dominated topologies in the TFE state. Surface accessibilities were computed in order to assess the effect of disulphides or secondary structure on the topology of the TFE state. This was done by a comparison of polypeptide chain accessibility between calculations, which take into account only local residues (±4 residue neighbours in sequence), and those which take into account the whole molecule and therefore possible longer range interactions (Figure 8). In case of such calculations on the native conformation it is apparent that a significant proportion of the accessible area of sidechains is buried as a result of tertiary contacts (e.g. the region corresponding to residues 24-36). By contrast the difference in average surface area buried locally, within helix structures, and that buried by additional long range contacts, consequence of the overall topology of the protein, is small over the ensemble of structures of the TFE denatured state. The greatest differences are in close proximity to the disulphide pair **64-80** and **76-94**, regions of the protein with little regular secondary structure. Therefore, helix formation and disulphides, have little effect on the apparent compactness of the derived structures as judged by a

probe with radius of 1.4 Å and significant area of non-polar surface remains accessible to solvent in this model of the TFE state.

Limitations of the present model

In several respects, the present structural model of the TFE state is not satisfactory. For example, it is not clear whether the helix-helix angles are a mere consequence of the absence of long-range interactions (except disulphides). The NMR data suggests, extensive interactions between non-sequential sidechains exist near disulphides. When these can be defined unambiguously new restraints would need to be balanced to achieve similar helix orientations. Indeed, structures calculated using a list of 50, in part rather speculative, NOEs are found to have significantly different, and an increased number of non-native helix-helix angles. While, substructures may or may not be native-like in the present model of the TFE denatured state, assembly of these structures into the native topology is not straightforward, suggesting that considerable rearrangements have to occur before a native-like, though "molten globule" like topology is attained.

The ensemble of structures derived for the TFE state here in absence of long range NOEs is greatly expanded with respect to the native structure, with approximate dimensions being 24x35x40 Å for the native and 30x30x60 Å for the TFE state. However, by contrast to the present model, low angle x-ray scattering suggests that the radius of gyration of the TFE state is, similarly to other denatured states, expanded by only 25% relative to that of the native state (Goto, Y. personal communication). Furthermore, tryptophan sidechains are not accessible to surface probes in Photo-CIDNP spectroscopy (Broadhurst, R.L., personal communication) of the TFE state. By contrast, sidechain accessibiliy is largely unaffected by long range structural effects in the present model of the TFE state and tryptophan sidechains remain largely exposed, together with a large proportion of the non-polar surface area. While, TFE is known to weaken hydrophobic interactions in favour of hydrogen bonding, it is unlikely to abolish them completely. Further studies on the structure of the TFE denatured state of a doubly ^{15}N, ^{13}C labelled hen lysozyme are in progress by heteronuclear NMR to resolve some of these issues and to characterise potential long range contacts. If this is successful the resulting additional restraints should be used to calculate a second generation of structures.

Possible implications for protein folding

Possible implications of the differential persistency and dynamics of structures in the TFE denatured state for the folding process are discussed in the chapter

7 and 8. The finding that it is the native-like helices which are the most persistent and dynamically rigid was of particular significance and suggested how productive folding may be initiated from such preformed secondary structure (Chapter 4). Given the present model calculations on the structure of the TFE state, the result, that the distribution of helix-helix angles are non-random in the ensemble of the derived structures and that a significant number of them have, at least in part, native-like helix-helix orientations, is intriguing in the light of diffusion collision models (Weaver & Karplus, 1994). Predefinition and preorientation of helices in a partially folded state are both likely to reduce conformational space significantly.

References

Brünger, A.T. (1988) X-PLOR manual, Yale University Press, C.T.

Karplus, M. & Weaver, D.L. (1994) Protein Science 3, 650-668.

MacArthur, M.W., & Thornton, J.M. (1993) Proteins: Struct.Funct. & Genetics 17, 232-251.

Mierke, D.F., Kurz, M. & Kessler, H. (1994) J.Am.Chem.Soc. 116, 1042-1049.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G. & Thornton, J.M. (1992) Proteins: Struct.Funct. & Genetics 12, 345-364.

Nilges, M., Kuszewski, J. & Brünger, A.T. (1991) in "Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy" (Hoch, J.C., Poulsen, F.M. & Redfield, C., eds.) p. 451-455, Plenum Press, N.Y.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H.J.C. (1977) J.Comput.Phys. 23, 327-341.

Smith, L.J., Sutcliffe, M.J., Redfield, C. & Dobson, C.M. (1993) J.Mol.Biol. 229, 930-44.

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

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, N.Y.

Wüthrich, K., Billeter, M. & Braun, W. (1983) J.Mol.Biol. 169, 949-61.

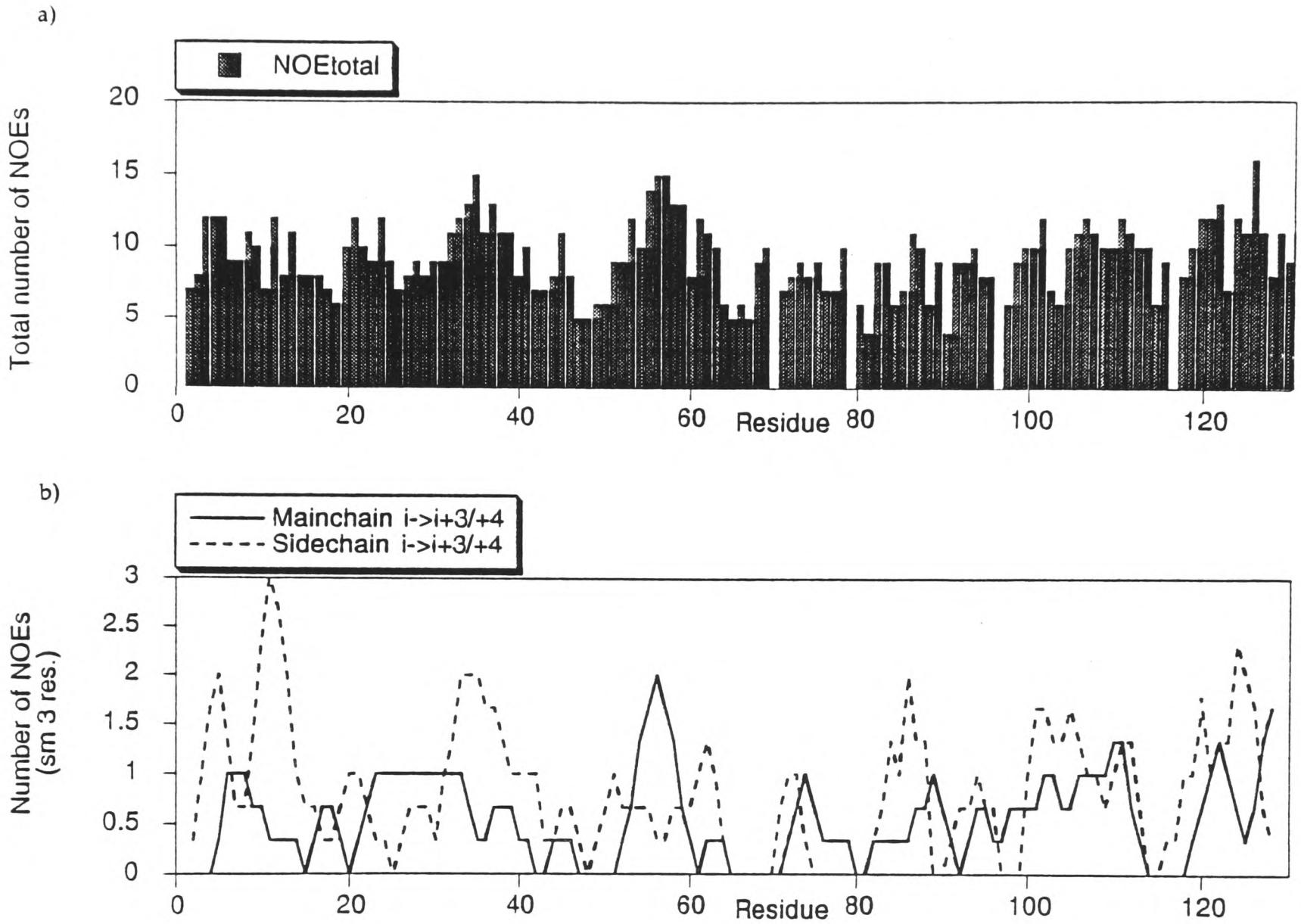


Figure 1
Number of short and medium range NOEs versus the protein sequence
a) all short and medium range NOEs b) medium range NOEs: mainchain Hi to mainchain Ni+3 and Ni+4 (continuous line) and sidechain Hi to mainchain Ni+3 and Ni+4 NOEs (dashed line) are plotted as smoothed over a sliding window of 3 residues.

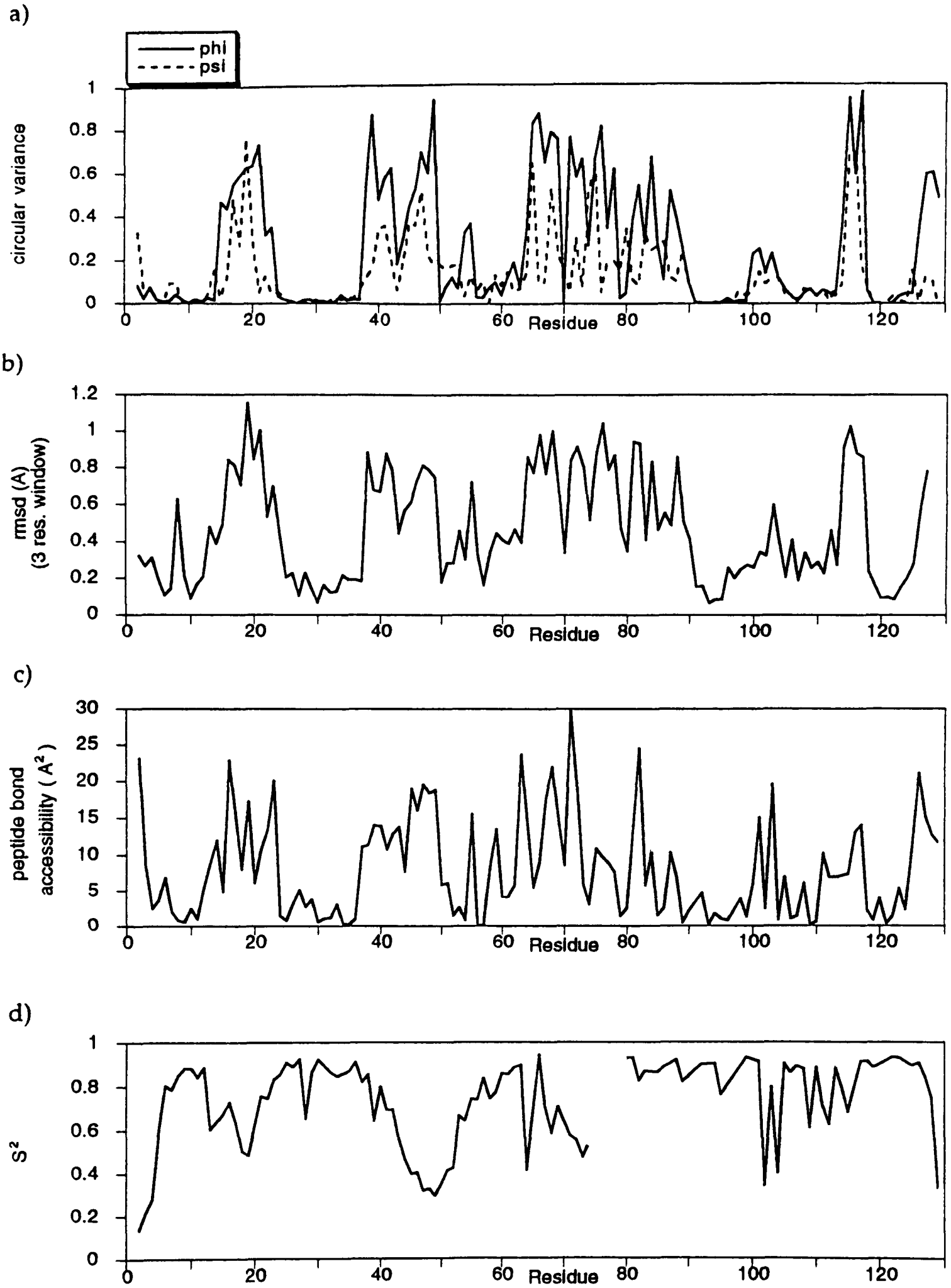


Figure 2

Regional variation of structure defined by a) circular covariance in ϕ and ψ over the ensemble of structures (MacArthur & Thornton, 1993) b) local rmsd by pairwise superimposition of members of the ensemble of structures by their mainchain atoms over a sliding window of 3 residues c) mean accessible surface area of the peptide bond in the ensemble (see Cahpter 6) d) Order parameters, S^2, derived for mainchain dynamics in the TFE denatured state (Chapter 8).

Figure 3

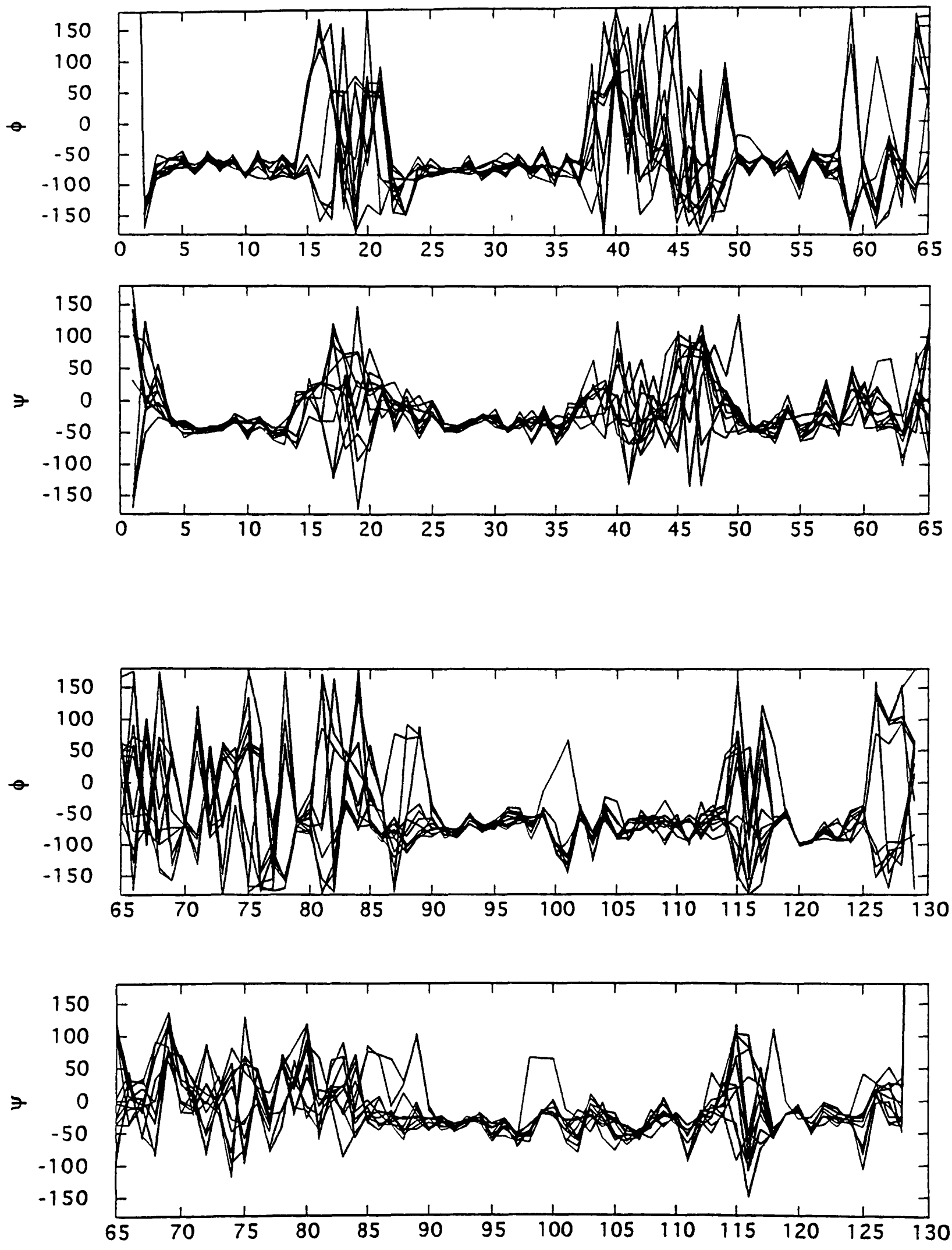


Figure 3
Plot of dihedral angles ϕ and ψ for members of the ensemble along the polypeptide chain. The angles were calculated using PROCHECK (Morris et al., 1992).

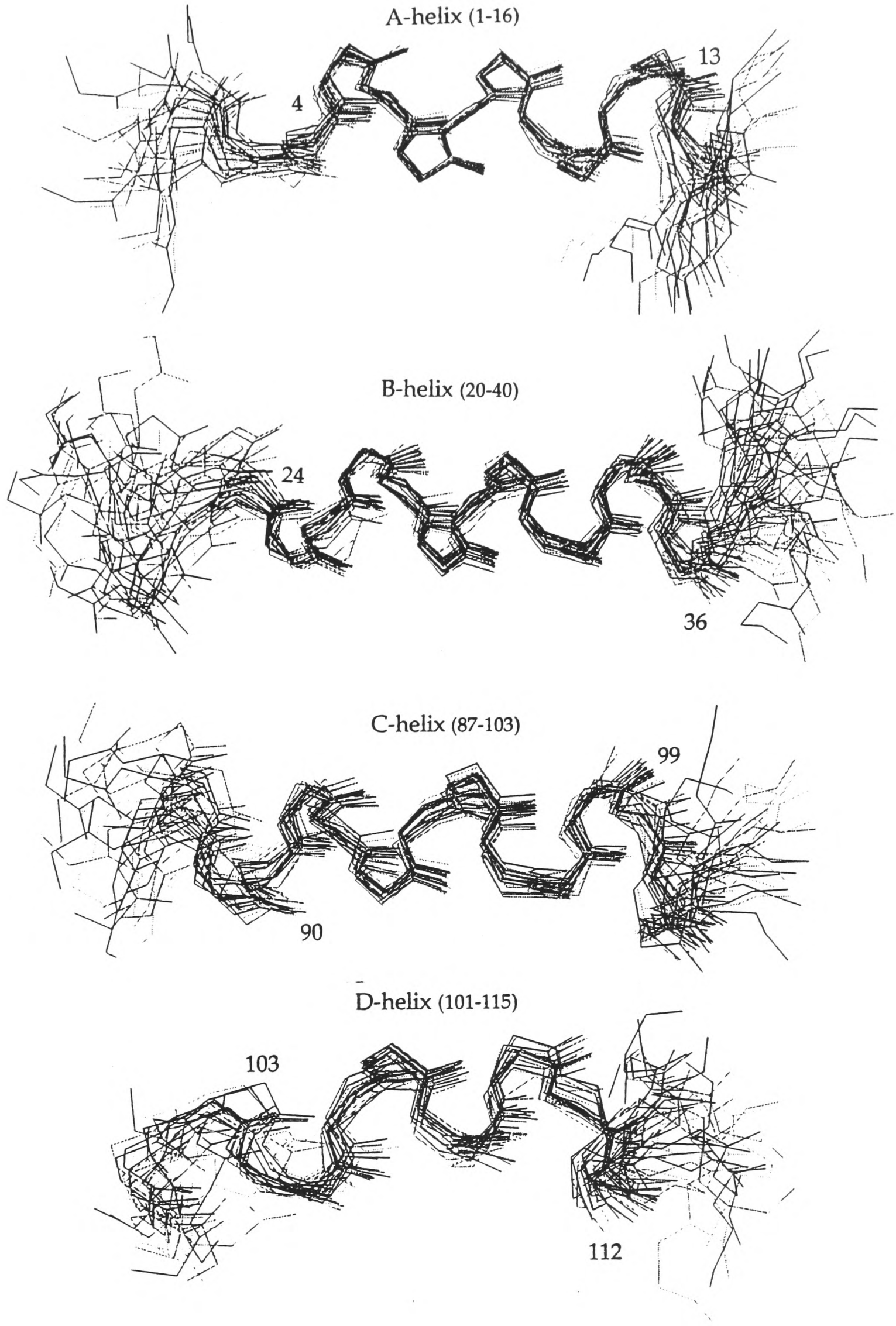


Figure 4
Superposition of mainchain atoms in regions with well defined dihedral angles, corresponding to the principal helices in the TFE state. The ensembles of the 20 best structures for each of the regions are shown and the rms deviations are listed in Table 3.

Figure 4

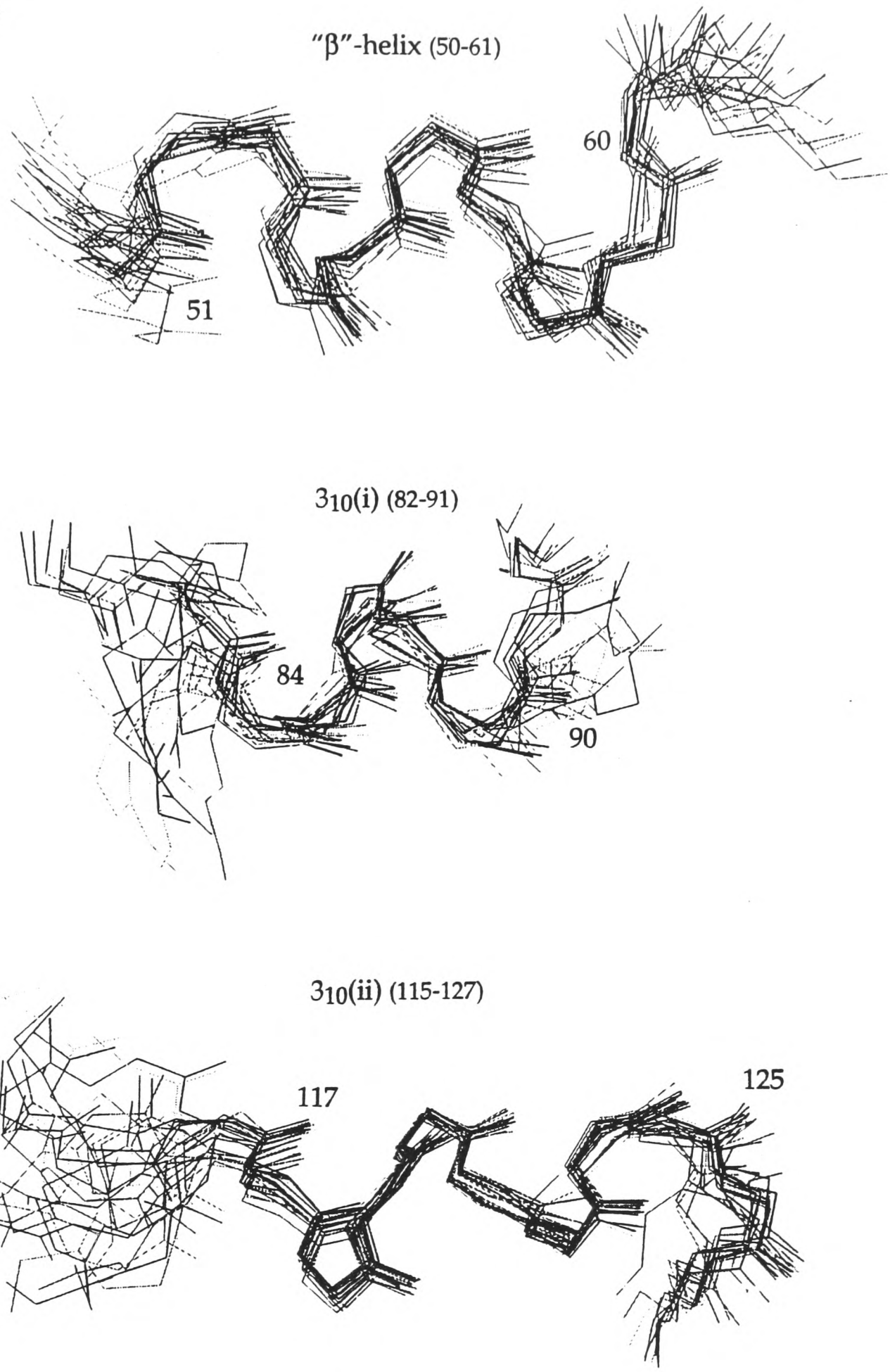


Figure 5 a)

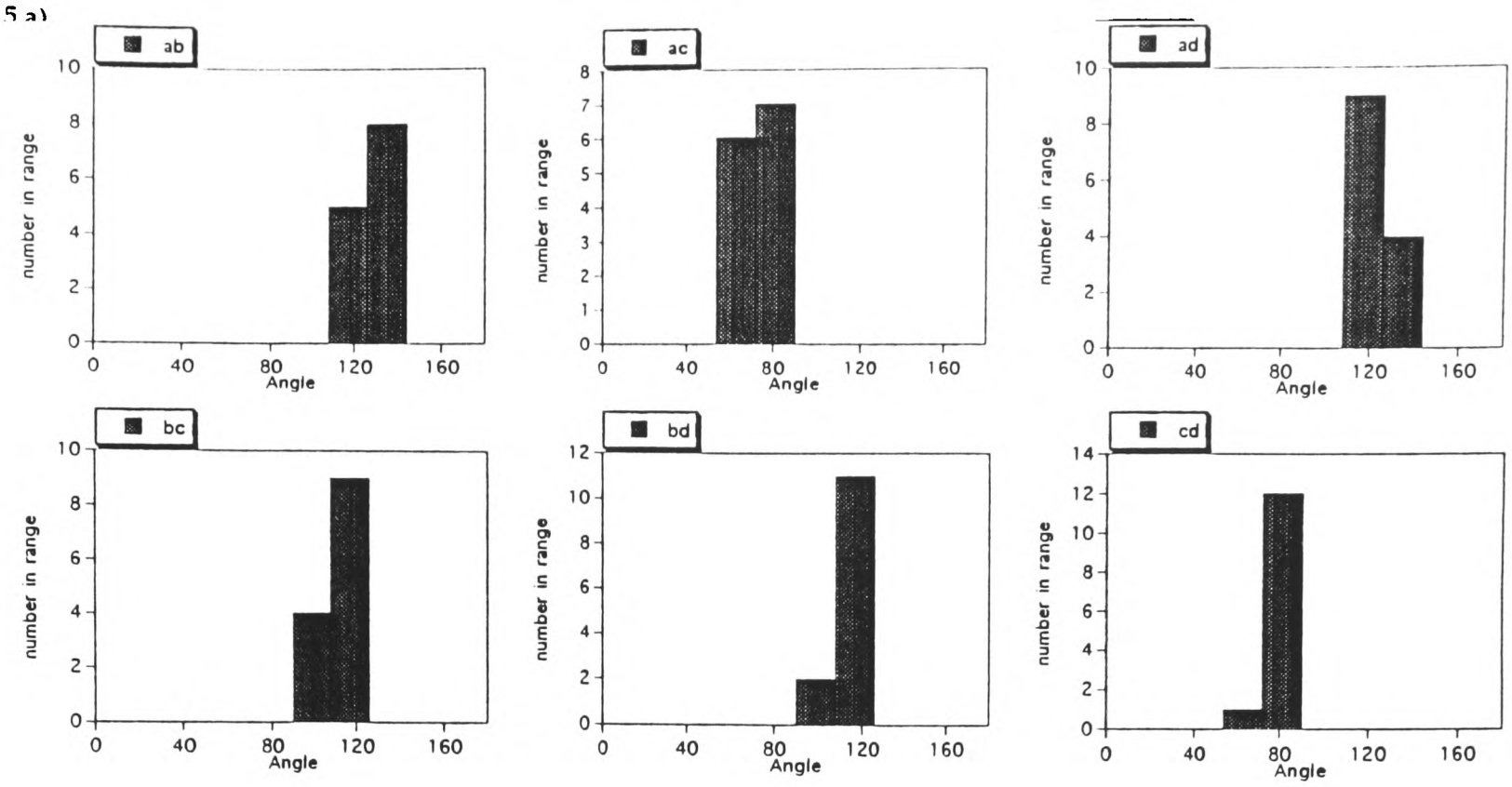


Figure 5 b)

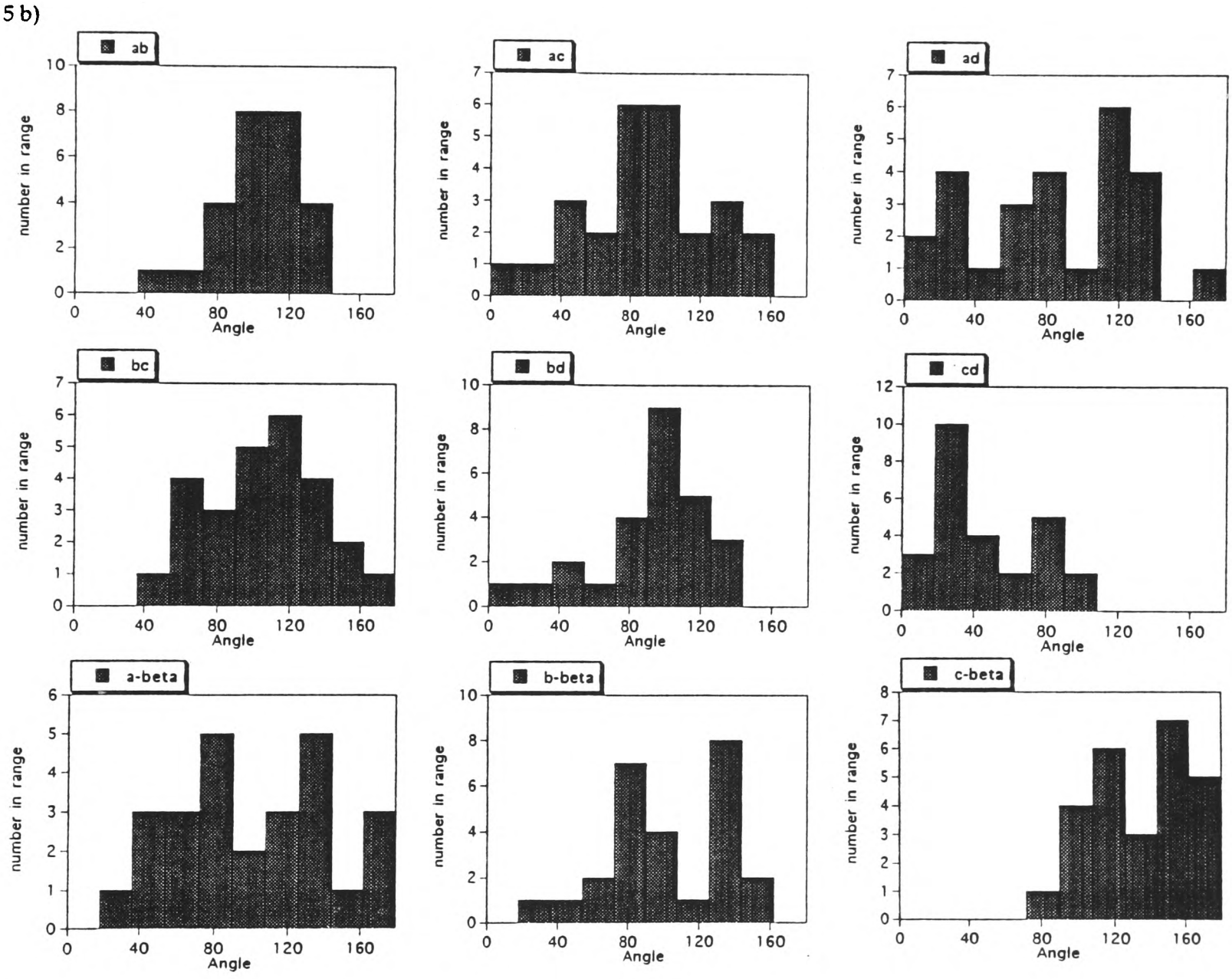


Figure 5

Plot of helix-helix angles in members of a) the native state ensemble (Smith et al., 1993) b) the TFE state. In case of b) also the angles between the non-native helix, stabilised by TFE in the region of the polypeptide chain which corresponds to part of the β-sheet in the native state, and the other principal helices is shown.

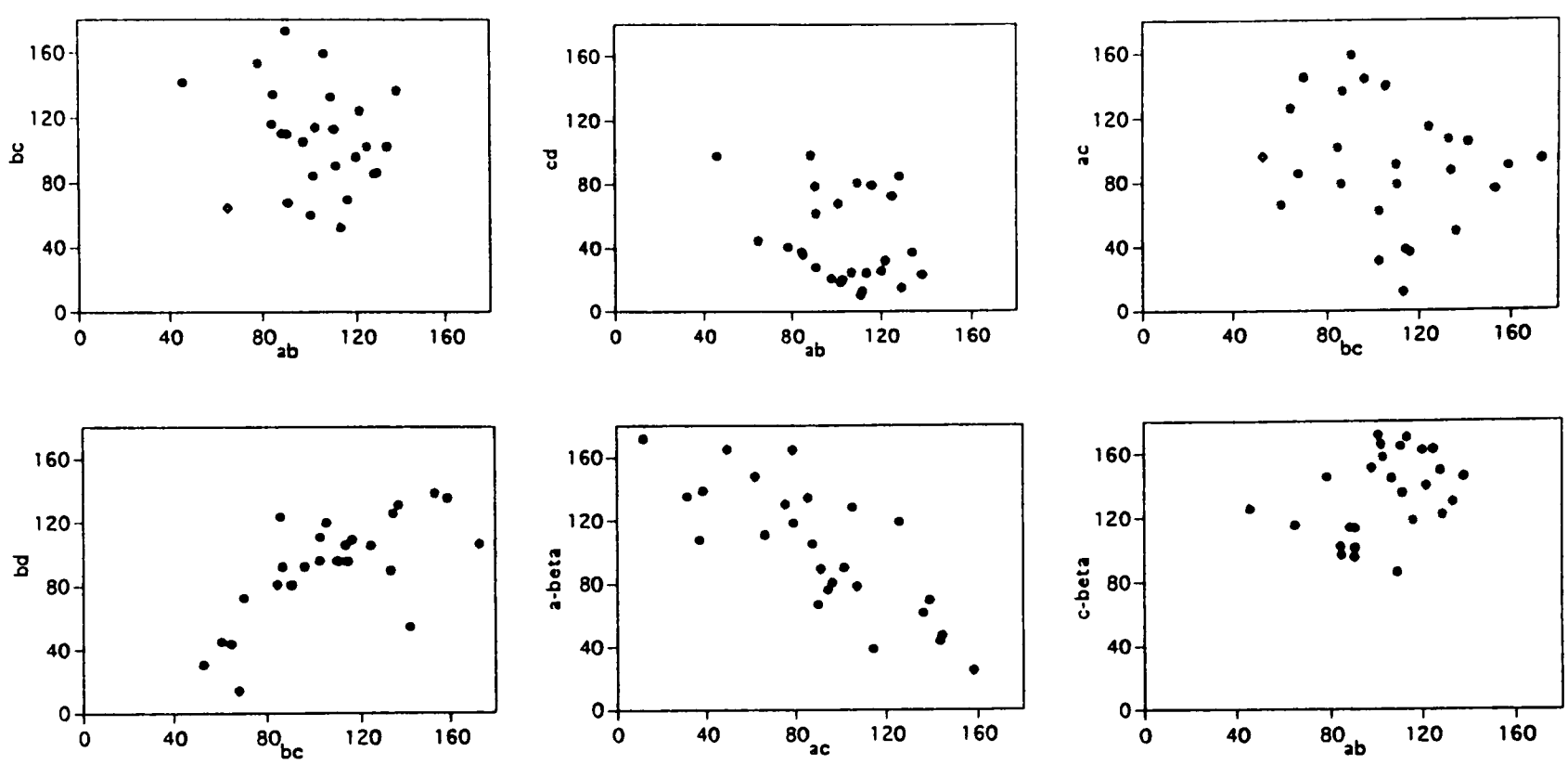


Figure 6
Plot of selected helix-helix angles in order to illustrate possible higher order correlations. Helices C and D, for example, are almost always collinear and therefore their orientation towards helix B is similar. Equally, the plot comparing the angles between helix A and the non-native helix (in the β-sheet region of the native protein) with the angles between the A and C-helices, points to the mostly antiparallel relationship between helix C and the part of the β-sheet which has adopted helix structure in the TFE state.

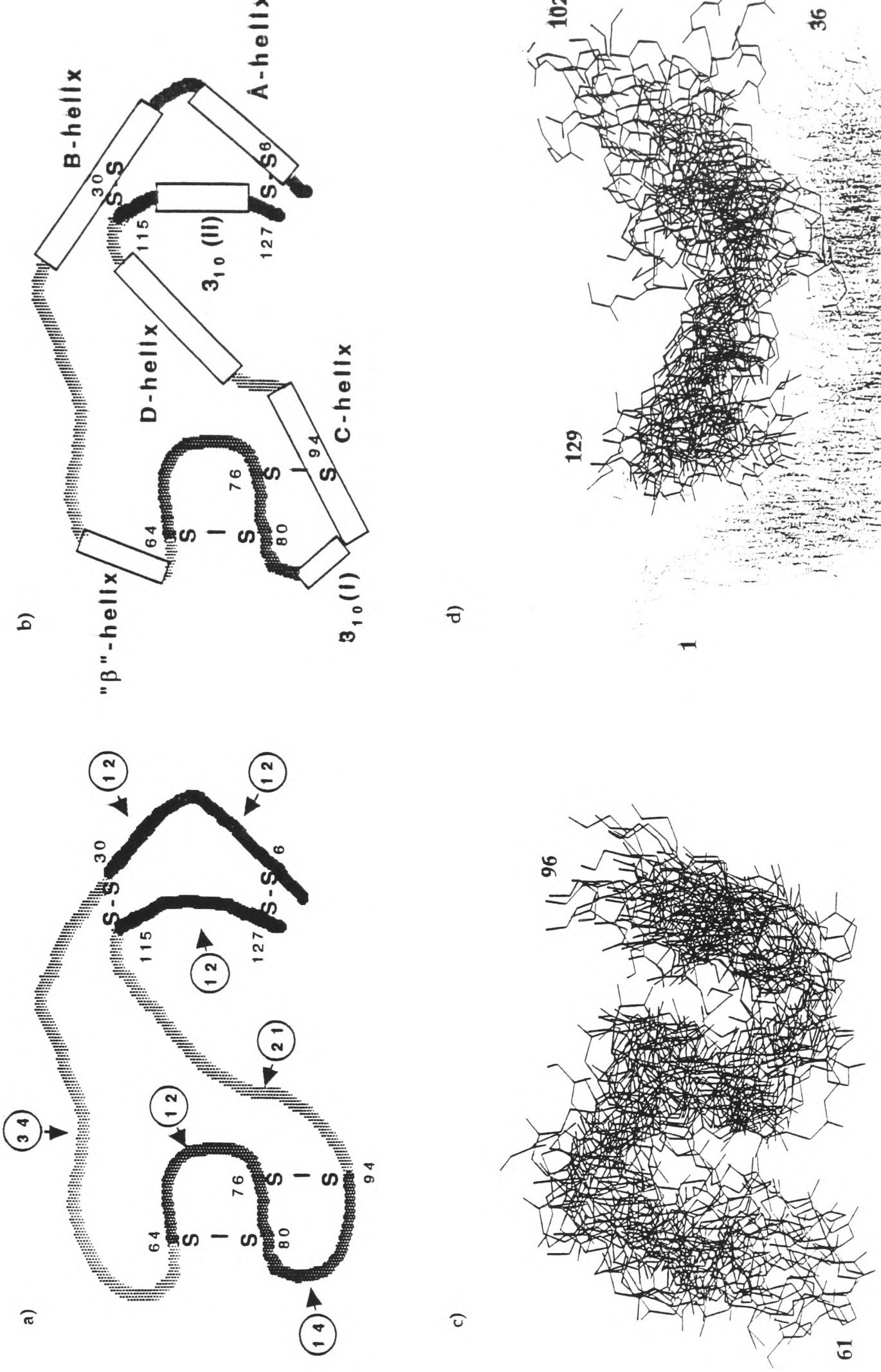


Figure 7

a) Schematic representation of the polypeptide chain of hen lysozyme as it is cross-linked by disulphides. The lengths of the loops between disulphides are shown, with exception to the loop comprising parts of helices A and B, which is subdivided into two parts. b) as above, but showing the approximate position of helices in the TFE denatured state.

c) and d) show a superposition of the ensemble of TFE structures for the two regions of the structures: residues 64-94 and 6-30-linked to 115-127; the pairwise rmsd between the structures of the ensemble are 4.72±1.18 Å and 6.20±1.27 Å respectively.

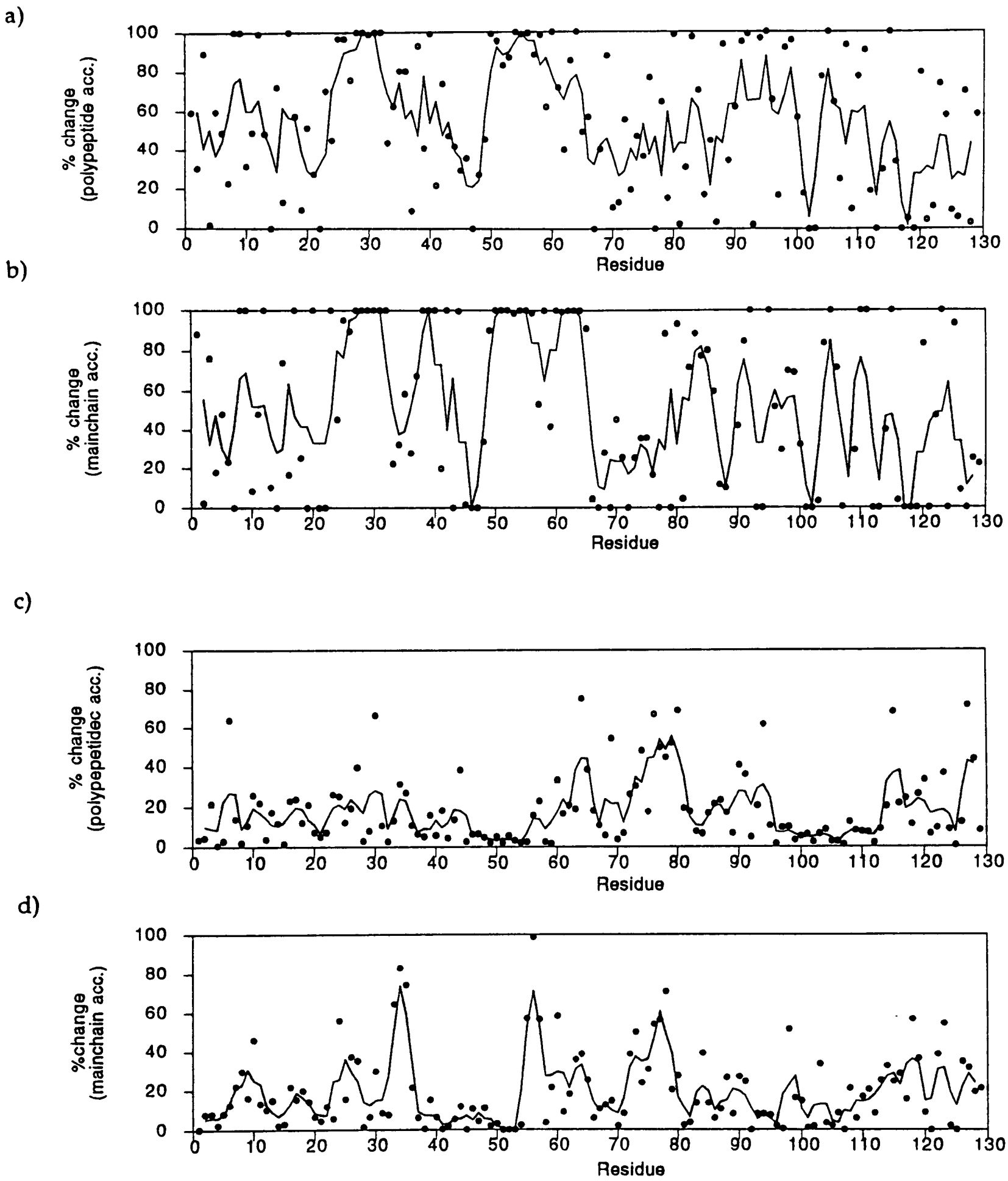


Figure 8

Accessible surface areas were calculated for each residue as the centre of a sliding window of 9 residues and in the presence of the entire polypeptide chain. The % change in accessible areas (relative to the area calulated considering only local contacts) are shown for the mainchain and sidechains in the native state in plot a), b) and the TFE denatured state in c) and d) respectively.

Supplementary Appendix

SI: Rates of amide exchange in min^{-1}.

(a) Native State, pH 7.5, 30°C.
(b) TFE denatured State, pH 2.0, 27°C, 50% TFE (v/v).
(c) Urea denatured State, pH 2.0, 20°C, 8 M urea.

SII: ^{15}N Relaxation data and derived order parameters for the native state of hen lysozyme.

(a) ^{15}N relaxation data for mainchain sites.
(b) Derived order parameters for the mainchain.
(c) ^{15}N relaxation data for sidechains.

SIII: ^{15}N relaxation data for the TFE state of hen lysozyme.

SI: Rates of amide exchange in min^{-1}.

Residue	Native (a)	TFE (b)	Urea (c)
3.0000	7.69e-01	1.53e-01	5.04e-02
8.0000	3.23e-03	1.07e-02	3.80e-02
9.0000	4.55e-04	3.99e-03	2.92e-01
10.000	1.59e-04	6.49e-03	5.26e-01
11.000	2.94e-03	2.34e-03	2.86e-01
12.000	5.00e-06	2.49e-03	6.68e-02
13.000	1.35e-03	5.57e-03	6.86e-02
14.000		2.33e-02	9.11e-02
15.000	8.33e-02	1.02e-01	
17.000	1.72e-01	3.82e-02	3.53e-01
23.000	1.85e-01	3.82e-02	4.15e-01
27.000	7.69e-03	1.38e-01	1.90e-01
28.000	2.63e-05	1.12e-02	1.58e-01
29.000	5.00e-06	8.42e-04	2.73e-02
30.000	5.00e-05	1.01e-02	4.54e-02
31.000	2.44e-05	5.37e-03	7.12e-02
32.000	8.33e-05	3.32e-03	4.31e-01
33.000	8.33e-05	2.04e-02	7.84e-02
34.000	1.72e-03	7.51e-03	1.08e-01
35.000	8.33e-02	1.05e-02	8.12e-02
36.000	1.75e-01	1.24e-02	1.98e-01
37.000	1.12e-01	4.37e-02	2.07e-01
38.000	1.89e-02	1.58e-02	8.45e-02
39.000	3.03e-03	3.18e-02	1.92e-01
40.000	1.23e-01	5.00e-02	1.24e-01
41.000	7.69e-01	2.64e-02	3.95e-01
42.000	1.54e-02	4.73e-02	1.49e-01
44.000	4.00e-02	9.15e-02	1.47e-01
46.000	3.33e-01	8.52e-02	2.94e-01
50.000	4.00e-01	8.61e-02	
51.000	5.88e-01	7.25e-02	1.45e-01
52.000	1.49e-04	1.19e-01	2.01e-01
53.000	7.69e-05	2.17e-02	9.34e-02
54.000	6.25e-04		
55.000	2.44e-01		
56.000	2.38e-02	8.86e-03	7.58e-02
57.000	1.08e-04	1.35e-02	2.53e-01
58.000	4.55e-05	2.21e-03	4.04e-02
59.000	7.69e-02	8.09e-03	1.35e-01
60.000	1.25e-04	2.42e-02	2.14e-01
61.000	4.17e-03	2.55e-02	5.00e-02
63.000	6.67e-03	5.00e-03	6.48e-02
64.000	1.05e-04	1.71e-02	5.95e-02
65.000	7.69e-03	3.32e-02	5.30e-02
75.000	3.57e-02	2.85e-02	9.12e-02
76.000	1.02e-02	2.37e-02	1.44e-01
78.000	2.17e-02	3.87e-03	2.78e-02
80.000	3.70e-02	2.04e-02	
82.000	9.09e-01		1.01e-01
83.000	5.56e-04	1.52e-02	5.92e-02

Residue	Native (a)	TFE (b)	Urea (c)
84.000	2.63e-01	1.17e-02	1.23e-01
85.000		1.08e-02	
92.000	5.26e-05	9.49e-04	2.48e-02
93.000	1.82e-01	3.15e-03	7.45e-02
94.000	2.00e-01	4.05e-03	4.44e-02
95.000	3.45e-06	6.50e-04	7.52e-02
96.000	5.00e-06	6.90e-04	1.96e-01
97.000	3.70e-03	2.21e-03	9.52e-02
98.000	4.17e-05	1.26e-03	8.00e-02
99.000	1.64e-04	1.52e-03	5.30e-02
100.00	2.86e-01	1.32e-02	
108.00	6.67e-02	3.00e-03	2.91e-01
111.00	2.44e-02	1.11e-03	1.34e-01
112.00	1.27e-01	1.17e-03	5.46e-02
114.00	7.69e-01	1.34e-03	
115.00	1.49e-01	1.91e-02	8.60e-02
123.00	7.14e-02	5.79e-03	1.74e-01
124.00	3.03e-02	1.58e-03	2.48e-02
125.00	2.70e-01	7.40e-03	2.27e-01

(a) Native State, pH 7.5, 30°C.

(b) TFE denatured State, pH 2.0, 27°C, 50% TFE (v/v).

(c) Urea denatured State, pH 2.0, 20°C, 8 M urea.

SII: ^{15}N Relaxation data and derived order parameters for the native state of hen lysozyme.

(a) ^{15}N relaxation data for mainchain sites.

Residue	R1	+/-	R2	+/-	NOE	+/-
2	2.16	0.03	7.10	0.10	0.84	0.04
3	2.18	0.03	7.72	0.11	0.78	0.04
4	2.26	0.03	6.85	0.11	0.77	0.04
5	2.37	0.02	6.97	0.07	0.69	0.03
6	2.24	0.02	7.41	0.07	0.77	0.03
7	2.32	0.02	7.38	0 06	0.73	0.02
8	2.42	0.02	7.54	0.07	0.72	0.02
9	2.41	0.02	8.05	0.07	0.75	0.02
10	2.36	0.02	7.50	0.06	0.75	0.02
11	2.39	0.02	7.48	0.06	0.71	0.02
12	2.35	0.03	7.76	0.09	0.77	0.03
13	2.42	0.02	7.82	0.08	0.74	0.02
14	2.15	0.03	8.60	0.11	0.71	0.03
15	2.22	0.03	7.05	0.09	0.77	0.05
17	2.38	0.03	7.52	0.08	0.75	0.04
18	2.27	0.02	7.29	0.06	0.76	0.02
19	2.30	0.04	6.95	0.13	0.75	0.04
20	2.25	0.02	7.16	0.08	0.74	0.03
21	2.38	0.03	7.38	0.11	0.76	0.03
22	2.58	0.04	10.80	0.17	0.78	0.05
23	2.29	0.03	7.81	0.12	0.66	0.04
24	2.30	0.03	7.61	0.13	0.70	0.03
25	2.27	0.03	8.03	0.13	0.74	0.03
26	2.33	0.03	7.95	0.13	0.75	0.03
27	2.44	0.03	8.14	0.10	0.77	0.03
28	2.28	0.02	7.95	0.08	0.77	0.04
29	2.36	0.03	8.22	0.11	0.76	0.04
31	2.39	0.02	8.04	0.07	0.74	0.02
32	2.41	0.02	8.24	0.07	0.74	0.03
33	2.36	0.02	7.74	0.07	0.77	0.02
34	2.39	0.03	7.92	0.09	0.78	0.04
35	2.30	0.03	8.07	0.13	0.76	0.03
36	2.24	0.05	7.77	0.20	0.77	0.06
37	2.51	0.04	8.16	0.14	0.82	0.04
38	2.33	0.02	7.67	0.08	0.76	0.04
39	2.32	0.03	7.64	0.09	0.80	0.04
40	2.35	0.03	7.86	0.13	0.74	0.03
41	2.24	0.04	7.99	0.17	0.80	0.05
42	2.28	0.02	8.04	0.08	0.71	0.03
43	2.13	0.04	7.15	0.13	0.76	0.04
44	2.19	0.03	6.98	0.10	0.75	0.04
45	2.03	0.02	6.65	0.07	0.75	0.03
46	2.19	0.03	7.04	0.10	0.77	0.04
47	2.01	0.02	6.75	0.08	0.71	0.02
48	2.01	0.02	7.41	0.10	0.80	0.04
49	2.13	0.03	7.02	0.10	0.72	0.03
51	2.33	0.07	7.51	0.26	0.81	0.04
52	2.34	0.04	7.56	0.12	0.73	0.03
53	2.27	0.05	7.37	0.18	0.75	0.04
54	2.34	0.06	7.82	0.21	0.71	0.05
55	2.42	0.09	8.26	0.37	0.71	0.05
56	2.38	0.06	7.97	0.26	0.73	0.05
57	2.44	0.05	8.04	0.18	0.71	0.04
58	2.34	0.05	7.78	0.19	0.71	0.06
59	2.38	0.07	8.46	0.29	0.77	0.06
60	2.42	0.07	8.48	0.29	0.76	0.05
61	2.41	0.04	8.30	0.16	0.88	0.04
62	2.17	0.04	7.40	0.15	0.72	0.07
63	2.35	0.05	7.65	0.17	0.74	0.07
64	2.38	0.06	7.75	0.22	0.74	0.06
65	2.23	0.05	7.97	0.19	0.75	0.05
66	2.31	0.08	7.56	0.28	0.84	0.05
67	2.26	0.03	7.09	0.11	0.76	0.03

Residue	R1	+/-	R2	+/-	NOE	+/-
68	2.04	0.04	7 67	0.14	0.77	0.04
69	2.03	0.03	6.32	0 12	0.74	0.04
71	1.92	0.02	6.07	0.06	0 71	0.03
72	1.98	0.01	6.60	0.04	0.67	0.02
73	2.26	0.06	7 62	0.22	0.70	0.07
74	2.26	0.05	7 50	0.16	0.79	0.05
75	2.47	0.05	8.05	0.18	0.71	0.04
76	2.37	0.07	8.01	0.27	0.73	0.05
77	2.32	0.02	7.79	0.08	0.80	0.03
78	2.39	0.03	7.69	0.10	0.76	0.03
80	2.39	0.03	7.79	0.08	0.82	0.03
81	2.24	0.02	8.39	0.09	0.77	0.02
82	2.31	0.03	8.56	0.10	0.78	0.04
83	2.17	0.04	8.42	0.14	0.73	0.04
84	2.18	0.04	7.93	0.15	0.78	0.07
85	1.67	0.02	4 88	0.07	0.27	0.06
86	2.10	0.03	7.33	0.11	0.71	0.03
87	2.09	0.03	7.61	0.12	0.77	0.03
88	2.08	0.03	7.76	0.14	0.80	0.04
89	2.40	0.04	7.84	0.14	0.73	0.02
90	2.35	0.03	7.92	0.11	0.75	0.02
91	2.26	0.03	7.14	0.09	0.78	0.03
92	2.43	0.03	7.97	0.11	0.75	0.03
93	2.43	0.03	7.87	0.08	0.71	0.02
94	2.40	0.03	7.87	0.11	0.71	0.03
95	2.48	0.02	7.68	0.08	0.70	0.02
96	2.48	0.03	7.72	0.13	0.74	0.03
97	2.46	0.03	7.93	0.09	0.73	0.04
98	2.42	0.04	8.06	0.16	0.67	0.04
100	2.41	0.04	7.55	0.14	0.68	0.04
101	2.31	0.03	7.29	0.09	0.64	0.03
102	2.01	0.03	6.04	0.08	0.64	0.04
103	2.08	0.02	5.28	0.06	0.49	0.03
104	2.25	0.03	6.87	0.10	0.62	0.03
105	2.29	0.03	7.97	0.14	0.74	0.06
106	2.49	0.02	8.32	0.08	0.71	0.03
107	2.43	0.02	7.86	0.09	0.64	0.04
108	2.19	0.03	7.60	0.13	0.75	0.04
109	2.26	0.02	7.91	0.10	0.69	0.03
111	2.20	0.03	8.24	0.10	0.77	0.04
112	2.32	0.03	8.26	0.09	0.74	0.03
113	2.29	0.03	7.82	0.10	0.74	0.03
114	2.21	0.03	7.61	0.12	0.74	0.05
115	2.07	0.03	7.58	0.12	0.81	0.06
116	2.20	0.02	7.96	0.08	0.75	0.04
117	2.13	0.03	6.82	0.10	0.72	0.03
118	1.89	0.03	7.17	0.12	0.81	0.04
119	2.12	0.02	6.73	0.08	0.79	0.03
120	2.07	0.02	6.92	0.06	0.73	0.03
121	2.36	0.02	7.72	0.06	0.72	0.02
122	2.37	0.02	7.94	0.05	0.77	0.02
123	2.35	0.02	7.57	0.07	0.70	0.03
124	2.41	0.04	7.58	0.12	0.82	0.06
125	2.28	0.02	7.46	0.06	0.72	0.03
126	2.13	0.03	7.06	0.10	0.73	0.03
127	2.01	0.01	6.58	0.05	0.74	0.03
128	1.99	0.02	6.92	0.07	0.76	0.03
129	1.62	0.02	5.96	0.06	0.62	0.03

SII: ^{15}N Relaxation data and derived order parameters for the native state of hen lysozyme.

(b) Derived order parameters.

res	Ss2	+/-	(Sf2) S2	+/-	(tau,s) tau,e	+/-	Rex	+/-
2	1.00	0.00	0.83	0.01	0.00	0.00	0.00	0.00
3	1.00	0.00	0.83	0.01	0.00	0.00	0.63	0.14
4	1.00	0.00	0.83	0.01	20.08	16.74	0 00	0.00
5	1.00	0.00	0.85	0.01	71 82	17.26	0.00	0.00
6	1.00	0.00	0.86	0.01	0.00	0.00	0.00	0.00
7	1.00	0.00	0.88	0.01	0.00	0.00	0.00	0.00
8	1.00	0.00	0.89	0.01	65.66	22.71	0.00	0.00
9	1.00	0.00	0.93	0.01	0.00	0.00	0.00	0.00
10	1.00	0.00	0.89	0.01	28.55	17.98	0.00	0.00
11	1.00	0.00	0.89	0.01	74.28	19.74	0.00	0.00
12	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
13	1.00	0.00	0.92	0.01	0.00	0.00	0.00	0.00
14	1.00	0.00	0.82	0.01	0.00	0.00	1.57	0.14
15	1.00	0.00	0.84	0.01	19.16	18.97	0.00	0.00
17	1.00	0.00	0.89	0.01	46.53	32.69	0.00	0.00
18	1.00	0.00	0.86	0.00	0.00	0.00	0.00	0.00
19	1.00	0.00	0.84	0.01	28.37	19.45	0.00	0.00
20	1.00	0.00	0.85	0.01	24.13	14 43	0.00	0.00
21	1.00	0.00	0.89	0.01	27.43	21.67	0.00	0.00
22	1.00	0.00	0.99	0.01	0.00	0.00	2.38	0.20
23	1.00	0.00	0.88	0.01	68.67	28.75	0.00	0.00
24	1.00	0.00	0.89	0.01	0.00	0.00	0.00	0.00
25	1.00	0.00	0.87	0.01	0.00	0.00	0.63	0.16
26	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
27	1.00	0.00	0.94	0.01	0.00	0.00	0.00	0.00
28	1.00	0.00	0.87	0.01	0.00	0.00	0.50	0.11
29	1.00	0.00	0.90	0.01	0.00	0.00	0.51	0.14
31	1.00	0.00	0.93	0.01	0.00	0.00	0.00	0.00
32	1.00	0.00	0.94	0.01	0.00	0.00	0.00	0.00
33	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
34	1.00	0.00	0.92	0.01	0.00	0.00	0.00	0.00
35	1.00	0.00	0.88	0.01	0.00	0.00	0.55	0.16
36	1.00	0.00	0.86	0.02	0.00	0.00	0.45	0.26
37	1.00	0.00	0.96	0.01	0.00	0.00	0.00	0.00
38	1.00	0.00	0.90	0.01	0.00	0.00	0.00	0.00
39	1.00	0.00	0.89	0.01	0.00	0.00	0.00	0.00
40	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
41	1.00	0.00	0.86	0.02	0.00	0.00	0.67	0.21
42	1.00	0.00	0.87	0.01	0.00	0.00	0.60	0.10
43	1.00	0.00	0.83	0.01	0.00	0.00	0.00	0.00
44	1.00	0.00	0.83	0.01	0.00	0.00	0.00	0.00
45	1.00	0.00	0.78	0.01	0.00	0.00	0.00	0.00
46	1.00	0.00	0.83	0.01	0.00	0.00	0.00	0.00
47	1.00	0.00	0.78	0.01	0.00	0.00	0.00	0.00
48	1.00	0.00	0.77	0.01	0.00	0.00	0.84	0.13
49	1.00	0.00	0.82	0.01	0.00	0.00	0.00	0.00
51	1.00	0.00	0.89	0.02	0.00	0.00	0.00	0.00
52	1.00	0.00	0.89	0.01	0.00	0.00	0.00	0.00
53	1.00	0.00	0.87	0.01	0.00	0.00	0.00	0.00
54	1.00	0.00	0.91	0.02	0.00	0.00	0.00	0.00
55	1.00	0.00	0.94	0.03	0.00	0.00	0.00	0.00
56	1.00	0.00	0.92	0.02	0.00	0.00	0.00	0.00
57	1.00	0.00	0.94	0.01	0.00	0.00	0.00	0.00
58	1.00	0.00	0.90	0.01	0.00	0.00	0.00	0.00
59	1.00	0.00	0.91	0.03	0.00	0.00	0.67	0.36
60	1.00	0.00	0.93	0.03	0.00	0.00	0.58	0.35
61	1.00	0.00	0.95	0.01	0.00	0.00	0.00	0.00
62	1.00	0.00	0.85	0.01	0.00	0.00	0.00	0.00
63	1.00	0.00	0.90	0.01	0.00	0.00	0.00	0.00
64	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
65	1.00	0.00	0.86	0.02	0.00	0.00	0.69	0.26
66	1.00	0.00	0.89	0.02	0.00	0.00	0.00	0.00
67	1.00	0.00	0.85	0.01	18.75	16.55	0 00	0.00

res	Ss2	+/-	(Sf2) S2	+/-	(tau,s) tau,e	+/-	Rex	+/-
68	1.00	0.00	0.78	0.01	0.00	0.00	0.98	0.18
69	1.00	0.00	0.76	0.01	15.37	10.87	0.00	0.00
71	1.00	0.00	0.72	0.01	19.21	6.70	0.00	0 00
72	1.00	0.00	0.76	0.00	19.69	7.15	0.00	0.00
73	1.00	0.00	0.88	0.02	0.00	0.00	0.00	0.00
74	1.00	0.00	0.87	0.01	0.00	0.00	0.00	0.00
75	1.00	0.00	0.94	0.02	0.00	0.00	0.00	0.00
76	1.00	0.00	0.92	0.02	0.00	0.00	0.00	0.00
77	1.00	0.00	0.90	0.01	0.00	0.00	0.00	0.00
78	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
80	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
81	1.00	0.00	0.86	0.01	0.00	0.00	1.07	0.12
82	1.00	0.00	0.88	0.01	0.00	0.00	1.02	0.13
83	1.00	0.00	0.83	0.01	0.00	0.00	1.33	0.18
84	1.00	0.00	0.83	0.01	0.00	0.00	0.82	0.18
85	1.00	0.00	0.54	0.01	86.06	6.26	0.00	0.00
86	1.00	0.00	0.80	0.01	0.00	0.00	0.47	0.15
87	1.00	0.00	0.80	0.01	0.00	0.00	0.79	0.15
88	1.00	0.00	0.80	0.01	0.00	0.00	0.97	0.17
89	1.00	0.00	0.92	0.01	0.00	0.00	0.00	0.00
90	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
91	1.00	0.00	0.85	0.01	8.42	11.67	0.00	0.00
92	1.00	0.00	0.93	0.01	0.00	0.00	0.00	0.00
93	1.00	0.00	0.93	0.01	0.00	0.00	0.00	0.00
94	1.00	0.00	0.92	0.01	0.00	0.00	0.00	0.00
95	1.00	0.00	0.92	0.01	112.17	32.46	0.00	0.00
96	1.00	0.00	0.92	0.01	62.97	62.28	0.00	0.00
97	1.00	0.00	0.94	0.01	0.00	0.00	0.00	0.00
98	1.00	0.00	0.92	0.01	140.14	83.30	0.00	0.00
100	1.00	0.00	0.89	0.01	90.68	63.44	0.00	0.00
101	1.00	0.00	0.84	0.01	102.33	22.78	0.00	0.00
102	1.00	0.00	0.67	0.01	51.26	1.51	0.00	0.00
103	0.63	0.01	0.83	0.02	1707.88	185.88	0.00	0.00
104	1.00	0.00	0.77	0.10	95.46	16.30	0.00	0.00
105	1.00	0.00	0.88	0.01	0.00	0.00	0.47	0.19
106	1.00	0.00	0.96	0.01	0.00	0.00	0.00	0.00
107	1.00	0.00	0.91	0.01	165.00	70.06	0.00	0.00
108	1.00	0.00	0.84	0.01	0.00	0.00	0.45	0.17
109	1.00	0.00	0.85	0.01	49.11	15.30	0.63	0.13
111	1.00	0.00	0.84	0.01	0.00	0.00	1.05	0.14
112	1.00	0.00	0.89	0.01	0.00	0.00	0.69	0.12
113	1.00	0.00	0.89	0.01	0.00	0.00	0.00	0.00
114	1.00	0.00	0.87	0.01	0.00	0.00	0.00	0.00
115	1.00	0.00	0.79	0.01	0.00	0.00	0.84	0.15
116	1.00	0.00	0.84	0.01	0.00	0.00	0.80	0.10
117	1.00	0.00	0.81	0.01	0.00	0.00	0.00	0.00
118	1.00	0.00	0.72	0.01	0.00	0.00	1.01	0.17
119	1.00	0.00	0.80	0.01	2.65	4.99	0.00	0.00
120	1.00	0.00	0.80	0.00	0.00	0.00	0.00	0.00
121	1.00	0.00	0.91	0.01	0.00	0.00	0.00	0.00
122	1.00	0.00	0.92	0.00	0.00	0.00	0.00	0.00
123	1.00	0.00	0.90	0.01	0.00	0.00	0.00	0.00
124	1.00	0.00	0.90	0.01	17.46	29.03	0.00	0.00
125	1.00	0.00	0.87	0.00	0.00	0.00	0.00	0.00
126	1.00	0.00	0.82	0.01	0.00	0.00	0.00	0.00
127	1.00	0.00	0.77	0.00	0.00	0.00	0.00	0.00
128	1.00	0.00	0.76	0.01	0.00	0.00	0.44	0.09
129	1.00	0.00	0.60	0.01	25.22	5.75	0.79	0.08

SII: ^{15}N Relaxation data and derived order parameters for the native state of hen lysozyme.

(c) ^{15}N relaxation data for sidechains.

Relaxation parameters for (i) tryptophan, (ii) arginine and (iii) asparagine and glutamine sidechains.

(i) Trp

Residue	R1e	+/-	R1s	+/-	R2e	+/-	R2s	+/-	NOEe	+/-	NOEs	+/-
28	2.09	0 02	2 08	0 02	6.64	0 08	6.70	0 05	0 68	0 01	0 68	0 01
62	1.57	0 01	1 57	0.01	3 79	0 02	3.79	0 02	0 25	0 00	0 25	0.01
63	2 04	0.03	2 04	0 03	7.39	0.14	7.40	0.13	0 67	0 02	0 67	0.02
108	2.04	0 03	2 04	0 02	6.51	0.10	6 53	0.07	0 64	0.01	0 64	0 01
111	2 02	0 03	2 02	0 03	7.41	0.12	7.41	0 12	0.71	0.01	0.74	0 00
123	1.98	0.02	1.95	0.01	6 23	0 05	6 35	0 04	0.71	0 01	0.74	0 00

(ii) Arg

Residue	R1e	+/-	R1s	+/-	R2e	+/-	R2s	+/-	NOEe	+/-	NOEs	+/-
61	1.03	0.08	1.02	0.07	2.35	0 62	2 68	0.19	-0.45	0 04	-0.45	0 04
73	0 78	0 07	0 78	0 07	1.58	0 62	1 51	0 14	-1.76	0 12	-1.76	0 12
112	1.07	0 13	1 07	0.12	4 36	1 56	4.51	1 37	-0.30	0 05	-0.30	0.05
114	0 84	0.07	0 83	0.07	2 84	0.67	2.93	0 53	0 04	0 01	0.04	0 01
125	0.70	0 08	0.70	0 07	3.40	0.98	3.35	0 96	-2.77	0 19	-2.77	0 19

(iii) Asn/Gln

Residue

Res	R1e	+/-	R1s	+/-	R2e	+/-	R2s	+/-	N600e	+/-	N600s	+/-	N750e	+/-	N750s	+/-	N500e	+/-	N500s	+/-
19	1.83	0.05	1.74	0.02	6.77	0.08	6.82	0.07	0.38	0.02	0.38	0.01	0.30	0.02	0.29	0.01	0.39	0.04	0 44	0 01
27	2.65	0.27	2.66	0.27	13.96	0.74	14.99	0.73	0.59	0.21	0.78	0.00	0.87	0.08	0.80	0.00	0.82	0.23	0 75	0.00
37	1.89	0.06	1.99	0.04	8.32	0.22	7.98	0.16	0.51	0.04	0.53	0.02	0.50	0.04	0.48	0.03	0.54	0.08	0.56	0 02
39	2.75	0.06	2.75	0.07	12.84	0.22	12.83	0.22	0.82	0.06	0.78	0.00	0.80	0.05	0.80	0.00	0.80	0.13	0.75	0.00
41	1.27	0.01	1.28	0.01	3.61	0.05	3.57	0.04	-0.91	0.04	-0.94	0.02	-0.94	0.03	-0.94	0.02	-0.94	0.05	-0 86	0.03
44	2.10	0.03	2.14	0.02	8.67	0.09	8.54	0.07	0.50	0.04	0.51	0.02	0.45	0.03	0.45	0.02	0.46	0.07	0 54	0 02
46	2.30	0.09	2.30	0.09	11.50	0.31	11.51	0.29	0.76	0.17	0.78	0.00	0.74	0.10	0.80	0.00	0.64	0.18	0 75	0 00
57	3.25	0.55	3.06	0.14	12.77	0.61	12.82	0.57	0.89	0.10	0.78	0.00	0.83	0.22	0.80	0.00	0.87	0.32	0.75	0 00
59	2.93	0.49	2.91	0.47	17.02	1.12	16.89	1.15	0.77	0.20	0.78	0.00	0.92	0.39	0.80	0.00	0.84	0.24	0.75	0.00
65	2.15	0.04	2.18	0.03	9.09	0.15	8.97	0.12	0.70	0.07	0.66	0.03	0.65	0.04	0.65	0.04	0.55	0.04	0.66	0 02
74	2.79	0.11	2.78	0.11	13.42	0.35	13.41	0.35	0.88	0.14	0.78	0.00	0.79	0.08	0.80	0.00	0.76	0.19	0 75	0 00
77	1.66	0.03	1.64	0.03	3.37	0.34	4.50	0.18	-0.61	0.03	-0.64	0.02	-0.55	0.03	-0.49	0.04	-0.66	0 08	-0 70	0.05
93	2.31	0.04	2.31	0.05	9.94	0.16	9.94	0.15	0.49	0.05	0.55	0.03	0.55	0.05	0.50	0.04	0.52	0 10	0 57	0 02
103	1.90	0.07	1.87	0.07	4.87	0.25	5.10	0.20	-0.20	0.04	-0.12	0.03	0.10	0.03	0.09	0.03	-0.13	0.07	-0 27	0 04
106	2.26	0.08	2.24	0.07	9.83	0.25	9.89	0.22	0.51	0.10	0.58	0.04	0.59	0.06	0.55	0.05	0.50	0.11	0 59	0.03
113	2.30	0.06	2.32	0.04	7.99	0.16	7.95	0.12	0.02	0.01	0.03	0.01	0.11	0.02	0.10	0.01	0 08	0.04	0 01	0 02
121	1.72	0.03	1.70	0.01	5.92	0.05	5.93	0.04	-0.16	0.03	-0.22	0.01	-0.43	0.01	-0.43	0.01	-0.16	0.06	-0 06	0.01

SII: ^{15}N Relaxation data and derived order parameters for the native state of hen lysozyme.

(c) ^{15}N relaxation data for sidechains.

The experimental (e) as well as the simulated (s) data are listed. Both components of the ^{15}N-$^{1}H_2$ correlation pair of asparagine and glutamine sidechains, depending on overlap, were treated separately for the extraction of relaxation data and calculation of order parameters. The differences in S^2 values derived from the relaxation parameters extracted for the two components of the ^{15}N-$^{1}H_2$ correlations were found to be small, and in the majority of cases well within the uncertainties estimated from spectral noise; they are therefore shown here as the average. Certain sets of data were noisier than those used for the mainchain, reflecting the considerable linebroadening evident for some of the sidechain resonances. Assuming the same value of τ_R as for the mainchain, the corresponding T_1/T_2 ratio at 600 MHz is 4.17 ± 0.18; The criteria of Clore et al., (1990) were used initially to select the model of the spectral density function to be fitted (see Appendix 1). Furhermore, an attempt was made to distinguish between the simple and extended model for those asparagine and glutamine sidechains which had reduced values for the NOE (< 0.7). Whenever τ_e (simple model) > 25 ps, the predicted frequency dependence of the NOE can becomes fairly large and a change of about 20% is predicted over the frequency range 500-750 MHz. However, the data did not show clearly such a dependence within the experimental errors for several of them. Overall, the extended model appeared to produce the better agreement with experimental data including a smaller frequency dependence to the NOE and this model was chosen for 4 of the sidechains.

SIII: ^{15}N relaxation data for the TFE state of hen lysozyme.

Residue	R1 600	+/-	R2 600	+/-	NOE600	+/-	NOE750	+/-
2	1.28	0 01	3.17	0.03	0.03	0 01	-0.12	0.01
3	1.42	0 01	4 62	0.02	0 10	0.01		
4	1.48	0.01	5.60	0.05	0.17	0.02	0.37	0.01
5	1.39	0 02	9.93	0.15	0.52	0.06	0.58	0.01
6	1.38	0.03	12.08	0.31	0.56	0.07	0.68	0.02
7	1.34	0.03	12.15	0.17	0.51	0.07	0.68	0.03
8	1.31	0.02	11.68	0.35			0.68	0.02
9	1.28	0.01	13.43	0.31	0.57	0.06		
10	1.30	0.02	13.09	0.30	0.60	0.07	0.69	0.02
11	1.33	0.02	12.52	0.27	0.62	0.07	0.67	0.02
12	1.34	0.03	13.58	0.44	0.60	0.09	0.72	0.03
13	1.34	0.03	9.91	0.27			0.69	0.03
15	1.35	0.03	10.92	0.30	0.50	0.06	0.63	0 02
16	1.54	0.03	13.02	0.37	0.70	0.08	0.49	0.02
17	1.37	0.02	10.31	0.18	0.47	0.05	0.55	0.01
18	1.38	0.02	8.59	0.12	0.33	0.04	0.49	0.01
19	1.39	0.02	8.28	0.12	0.38	0.04	0.44	0.01
21	1.38	0.03	11.93	0.34	0.36	0.05	0.51	0.02
22	1.33	0.03	11.11	0.38	0.33	0.06	0.49	0.03
23	1.31	0.03	12.43	0.39	0.46	0.07	0.60	0.02
24	1.27	0.03	12.83	0.46	0.47	0.08	0.55	0.03
25	1.05	0.05	20.74	1.92	0.50	0.15	0.64	0.06
26	1.04	0.07	29.67	4.44	0.38	0.11	0.57	0.06
27	1.11	0.05	14.50	1.14	0.68	0.17	0.68	0.05
28	0.88	0.09	16.72	2.98	0.59	0.21		
29	0.90	0.11	18.05	5.04	0.75	0.61	0.72	0.12
30	1.01	0.12	22.47	3.67	0.54	0.16	0.79	0.09
32	0.91	0.11	20.75	7.05	0.52	0.23	0.80	0.14
33	0.89	0.11	20.44	5.13	1.05	0.72	0.75	0.10
35	0.94	0.11	18.57	4.70	0.59	0.24	0.64	0.09
36	1.03	0.06	18.64	2.13	0.89	0.42	0.63	0.07
37	1.36	0.04	19.50	2.76	0.55	0.18	0.57	0.05
38	1.16	0.05	14.30	1.00	0.34	0.10	0.54	0.05
39	1.21	0.04	10.27	0.78	0.42	0.08	0.53	0.03
40	1.26	0.04	11.74	0.61	0.34	0.07	0.43	0.03
41	1.35	0.03	10.54	0.30	0.27	0.05	0.43	0.02
42	1.40	0.03	10.96	0.31	0.34	0.05	0.42	0.02
43	1.38	0.03	9.24	0.25	0.28	0.05	0.43	0.02
44	1.36	0.02	8.00	0.15	0.22	0.03	0.45	0.02
45	1.48	0.01	7.21	0.08	0.30	0.03	0.38	0.01
46	1.51	0.02	7.35	0.11	0.27	0.03	0.42	0.01
47	1.46	0.02	6.07	0.09	0.19	0.02	0.34	0.01
48	1.57	0.02	6.39	0.08	0.23	0.03	0.43	0.01
49	1.46	0.01	5.77	0.07	0.19	0.03	0.37	0.01
50	1.45	0.01	6.54	0.07	0.23	0.02	0.44	0.01
51	1.45	0.02	7.37	0.12	0.27	0.03	0.44	0.01
52	1.49	0.02	7.64	0.11	0.28	0.03	0.47	0.01
53	1.43	0.02	10.28	0.23	0.36	0.04	0.53	0.02
54	1.47	0.03	9.80	0.37	0.31	0.05	0.66	0.04
55	1.41	0.03	11.53	0.39	0.40	0.07	0.57	0.02
56	1.43	0.03	11.73	0.41	0.41	0.06	0.57	0.03
57	1.33	0.03	11.31	0.68	0.52	0.08	0.62	0.03
58	1.34	0.03	10.90	0.38	0.67	0.13	0.55	0.03
59	1.34	0.04	11.15	0.44	0.44	0.07	0.62	0.04
60	1.33	0.05	13.09	1.25	0.54	0.10	0.63	0.04
61	1.34	0.06	13.53	1.08	0.50	0.11	0.64	0.04
62	1.26	0.10	12.41	1.98	0.61	0.16	0.66	0.05
63	1.20	0.22	14.69	4.85	0.62	0.25	0.68	0.10
64	2.55	0.96	24.70	13.50	0.38	0.24	0.59	0.15
66	1 23	0.14	22.96	6.79	0.54	0.25	0 72	0.11
67	1.44	0.05	10.70	0.63	0.50	0.12	0.57	0.04
68	1.36	0.03	9.67	0.31	0.39	0.07	0.60	0 03
69	1.37	0.05	11.05	0.46	0 29	0.05	0.44	0 03

Residue	R1 600	+/-	R2 600	+/-	NOE600	+/-	NOE750	+/-
71	1.34	0.04	9.52	0.37	0.34	0.07	0.46	0.04
72	1.36	0.03	9.28	0.28	0.38	0.05	0.57	0.03
73	1.43	0.06	8.31	0.56	0.48	0.13	0.57	0.04
80	1.17	0.07	18.55	2.86	0.81	0.31	0.67	0.11
81	1.14	0.12	17.70	4.40	0.80	0.21	0.62	0.09
82	1.23	0.11	11.90	2.90	0.37	0.12	0.52	0.11
83	1.24	0.06	12.48	1.02	0.42	0.10	0.61	0.04
85	1.21	0.07	15.89	2.41	0.37	0.08	0.63	0.07
86	1.16	0.12	16.97	2.71	0.51	0.22	0.56	0.08
88	1.16	0.15	16.41	4.44	0.59	0.27	0.65	0.10
89	1.14	0.07	19.30	2.73	0.79	0.31	0.28	0.05
92	0.97	0.10			0.90	0.44	0.59	0.08
94	0.97	0.13	22.49	3.79	1.56	1.20	0.65	0.06
95	0.97	0.10	34.68	9.08	0.90	0.44	0.74	0.08
99	1.03	0.11	22.80	3.80	1.18	0.53	0.68	0.08
101	1.18	0.05	16.33	1.51	0.41	0.11	0.61	0.04
102	1.26	0.04	6.19	0.21	0.28	0.07	0.47	0.03
103	1.24	0.03	11.51	0.38	0.38	0.07	0.47	0.04
104	1.24	0.03	6.92	0.21	0.90	0.09	0.40	0.03
105	1.08	0.06	17.05	1.39	0.66	0.17	0.56	0.04
106	1.06	0.04	16.01	1.11	0.34	0.08	0.49	0.04
107	1.04	0.06	15.25	1.88	0.48	0.14	0.54	0.06
108	0.94	0.08	20.72	3.59	0.39	0.15	0.69	0.09
109	1.00	0.07	9.42	0.74	0.62	0.30	0.60	0.10
110	0.92	0.09	27.32	6.29	0.36	0.11	0.63	0.09
111	0.90	0.11	21.78	7.22			0.83	0.14
112	1.76	0.30	26.41	8.70	0.57	0.27	0.79	0.16
113	1.05	0.05	13.40	1.03	0.56	0.16	0.72	0.09
114	1.01	0.06	21.24	3.55	0.52	0.21	0.66	0.08
115	0.87	0.11	27.72	7.01	1.04	0.78	0.76	0.10
117	1.07	0.06	18.87	3.21	0.63	0.18	0.60	0.07
118	1.09	0.05	19.17	1.95	0.58	0.15	0.62	0.05
119	1.13	0.06	14.41	1.03	0.39	0.09	0.62	0.04
122	1.17	0.03	12.72	0.48			0.68	0.04
123	1.15	0.04	15.77	1.29	0.52	0.10	0.67	0.04
125	1.20	0.04	16.57	0.88	0.42	0.06	0.64	0.04
126	1.25	0.04	18.48	1.24	0.52	0.10	0.65	0.04
127	1.21	0.03	13.18	0.41	0.48	0.08	0.68	0.03
128	1.27	0.01	11.74	0.17	0.47	0.04	0.55	0.02
129	1.37	0.01	6.15	0.05	0.12	0.02	0.27	0.01