PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Insights into Protein Folding Using Physical Techniques: Studies of Lysozyme and <latex>$\alpha $</latex>-lactalbumin

Sheena E. Radford and Christopher M. Dobson

Phil. Trans. R. Soc. Lond. B 1995 **348**, 17-25
doi: 10.1098/rstb.1995.0041

Insights into protein folding using physical techniques: studies of lysozyme and α-lactalbumin

SHEENA E. RADFORD AND CHRISTOPHER M. DOBSON

Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, U.K.

SUMMARY

Understanding the process of protein folding, during which a disordered polypeptide chain is converted into a compact well-defined structure, is one of the major challenges of modern structural biology. In this article we discuss how a combination of physical techniques can provide a structural description of the events which occur during the folding of a protein. First, we discuss how the rapid kinetic events which take place during *in vitro* folding can be monitored and deciphered in structural terms. Then we consider how more detailed structural descriptions of intermediates may be obtained from NMR studies of stable, partly folded states. Finally, we discuss how these experimental strategies may be extended to relate the findings of *in vitro* studies to the events occurring during folding *in vivo*. The approaches will be illustrated using results primarily from our own studies of the c-type lysozymes and the homologous α-lactalbumins. The conclusions from these studies are also related to those from other systems to highlight their unifying features. On the basis of these results we identify some of the determinants of the events in folding and we speculate on the importance of these in driving folding molecules to their native states.

1. INTRODUCTION

Proteins are synthesized within the cell on ribosomes. Although folding in the cell is a highly complex process involving a cascade of helper proteins called the molecular chaperones, it is clear that the major events in the folding process occur after departure from the ribosome (Ellis 1994; Frydman *et al.* 1994) and perhaps even after release from the chaperones (Weissman *et al.* 1994). For extracellular proteins the process is even more complex because folding must be tightly coupled to translocation (Bychkova & Ptitsyn 1994) and in this case a large part of folding may occur following secretion from the cell. Despite these complexities, many proteins can fold efficiently and correctly in isolation, provided that suitable conditions are found.

Given the improbability that folding could occur in a finite time on a random search basis (Levinthal 1968; Sali *et al.* 1994) and that folding in the cell is not thought to be sterically driven by chaperones, but merely controlled, it is unlikely that the principles behind the folding process *in vivo* and *in vitro* will be fundamentally different (although the two situations may differ substantially in detail given the far from physiological conditions of most experiments *in vitro*). Studies of folding *in vitro*, where physical techniques capable of providing detailed structural information can be used most readily, can provide specific information about the folding process at the molecular level. It is then of great interest to investigate how the various auxiliary factors might moderate or control the process in the light of our understanding of the refolding of proteins in isolation. Efforts to extend the use of physical techniques to enable the achievement of these objectives are increasingly being made (Gray & Fersht 1993; Jackson *et al.* 1993; Hartl *et al.* 1994; Landry & Gierasch 1994; Robinson *et al.* 1994).

In this article, strategies to provide a structural description of the events in protein folding are discussed with particular reference to one family of proteins, the c-type lysozymes; these are small monomeric proteins with *ca.* 130 residues. The structure of the archetypal family member of this family – the hen protein – is shown in figure 1. There are four α-helices; two located towards the C-terminus and two towards the N-terminus of the polypeptide chain, which together with a short 3_{10} helix make up one 'domain' of the molecule (the α-domain). The other domain, the β-domain, is

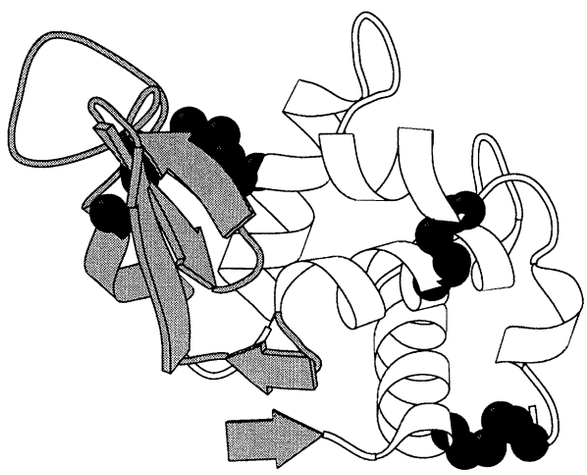

Figure 1. Schematic view of the structure of hen lysozyme (Blake *et al.* 1965). The α- and β-domains are shaded white and grey, respectively, and the four disulphide bonds are shown in black. The diagram was drawn using the program Molscript (Kraulis 1991).

formed from a triple-stranded antiparallel β-sheet, another short 3_{10} helix, and a long loop. A short antiparallel double-stranded β-sheet links these two domains, as does one of the four disulphide bridges. The disulphide bonds remain intact during the majority of the folding studies discussed here.

2. KINETIC REFOLDING *IN VITRO*

Refolding of most small proteins following dilution, for example, from high concentrations of urea or guanidinium chloride, is often completed within a second or so. Given that NMR spectroscopy is an inherently insensitive and slow technique, the simple view that structural changes taking place during refolding could be monitored by recording a series of spectra as a function of time during refolding, although appealing in principle is, for the majority of cases, impossible in practice. Only in a few cases, for example where proteins fold very slowly because of a need for cis/trans proline isomerism, has direct observation of folding by one-dimensional NMR methods been possible (Koide *et al.* 1993).

Despite these problems, methods have been developed to allow the power of NMR spectroscopy to be applied to the study of the structures of species sampled during folding by adopting an indirect approach. The key to this is that the rate of exchange of amide hydrogens with solvent is dependent on their environment in a protein structure. In the pulse labelling approach (Roder *et al.* 1988; Udgaonkar & Baldwin 1988), this property is utilized to trap deuterons at specific amide sites within regions of structure as they are formed during folding. These sites can be detected and assigned to individual residues by two-dimensional NMR methods by recording spectra of the protein once refolding is complete and making use of residue-specific assignments of the native protein. By this method, it was found that the two structural domains of lysozyme (despite undergoing cooperative unfolding and refolding at equilibrium; Imoto *et al.* 1972) form structure protected against hydrogen exchange at very different rates. In the majority of molecules, the α-domain forms faster than the β-domain (Radford *et al.* 1992). Although within both domains the protection kinetics appear similar (indicating that stabilization of each domain to a protected state occurs cooperatively), protection does not follow a simple exponential path. Further experiments have shown that this is because there are several alternate folding routes (Radford *et al.* 1992; Miranker *et al.* 1993).

To complement these studies we have introduced the idea of monitoring the cooperativity of folding and the populations of folding intermediates by detection of hydrogen exchange labelling by electrospray ionization mass spectrometry (ESI-MS) (Miranker *et al.* 1993). The basis of this method is that the incorporation of deuterons instead of protons within a protein molecule increases its mass, and that ESI-MS enables this to be measured, for a protein the size of lysozyme to a mass resolution better than 1 Da. This approach has enabled us to show that the α-domain can fold to a protected state independently of the β-domain, although the converse is not true, and there is no evidence to suggest that persistent structuring of the β-domain can occur in the absence of a stably structured α-domain. In addition, the multi-exponential protection kinetics arise from the existence of populations of molecules with distinct folding kinetics; some 25% of the population appears able to fold to a native-like state within 10 ms, whereas in the remainder of the molecules this process takes in excess of 300 ms.

These results provide considerable insight into the nature of species formed during the folding process. They rely, however, on hydrogen exchange protection which, because of the manner in which the experiments are performed, can only detect highly stabilized folding intermediates. To gather information about less structured states and to interpret further the structural details of protected states, these methods must be related to other indicators of structure gained through complementary techniques (Dobson *et al.* 1994; Evans & Radford 1994). We have therefore carried out a variety of stopped-flow optical measurements under conditions as similar as possible to the hydrogen exchange measurements (Radford *et al.* 1992; Itzhaki *et al.* 1994). A summary of these experiments and their results is given in table 1. Of particular importance are circular dichroism (CD) experiments in the far ultraviolet (UV) region, which monitors formation of secondary structure, and in the near UV which detects the immobilization of aromatic residues within a close packed structure. In addition, intrinsic fluorescence, fluorescence quenching and binding of the dye 1-anilino naphthalene sulphonic acid (ANS), which are thought to reflect hydrophobic collapse, have been measured. Finally, formation of the functional native protein can be detected through binding of a fluorescently labelled inhibitor that binds to the active site cleft which lies between the two folding domains (Itzhaki *et al.* 1994).

3. SCHEMATIC PATHWAYS OF FOLDING

The acquisition of this array of structural information enables us to propose a schematic folding pathway for lysozyme that is consistent with the currently available experimental evidence (see figure 2). This indicates that within a few ms a compact state of the protein is formed which has a native-like content of secondary structure (Chaffotte *et al.* 1992; Radford *et al.* 1992). Primary evidence for this comes from the very rapid kinetics of formation of native-like ellipticity monitored by far UV CD (Chaffotte *et al.* 1992; Radford *et al.* 1992), from the observation that ANS fluorescence is enhanced most strongly at the earliest measurable time following initiation of folding, and from the resilience of this state to fluorescence quenching by iodide ions (Itzhaki *et al.* 1994).

This state is shown in figure 2 to be heterogeneous; this provides a mechanism for kinetically distinct populations of folding molecules. The ESI-MS data require that such heterogeneity occurs early in folding before the onset of protection (Miranker *et al.* 1993), but refolding experiments from denatured states generated by different means indicates that residual structure in

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

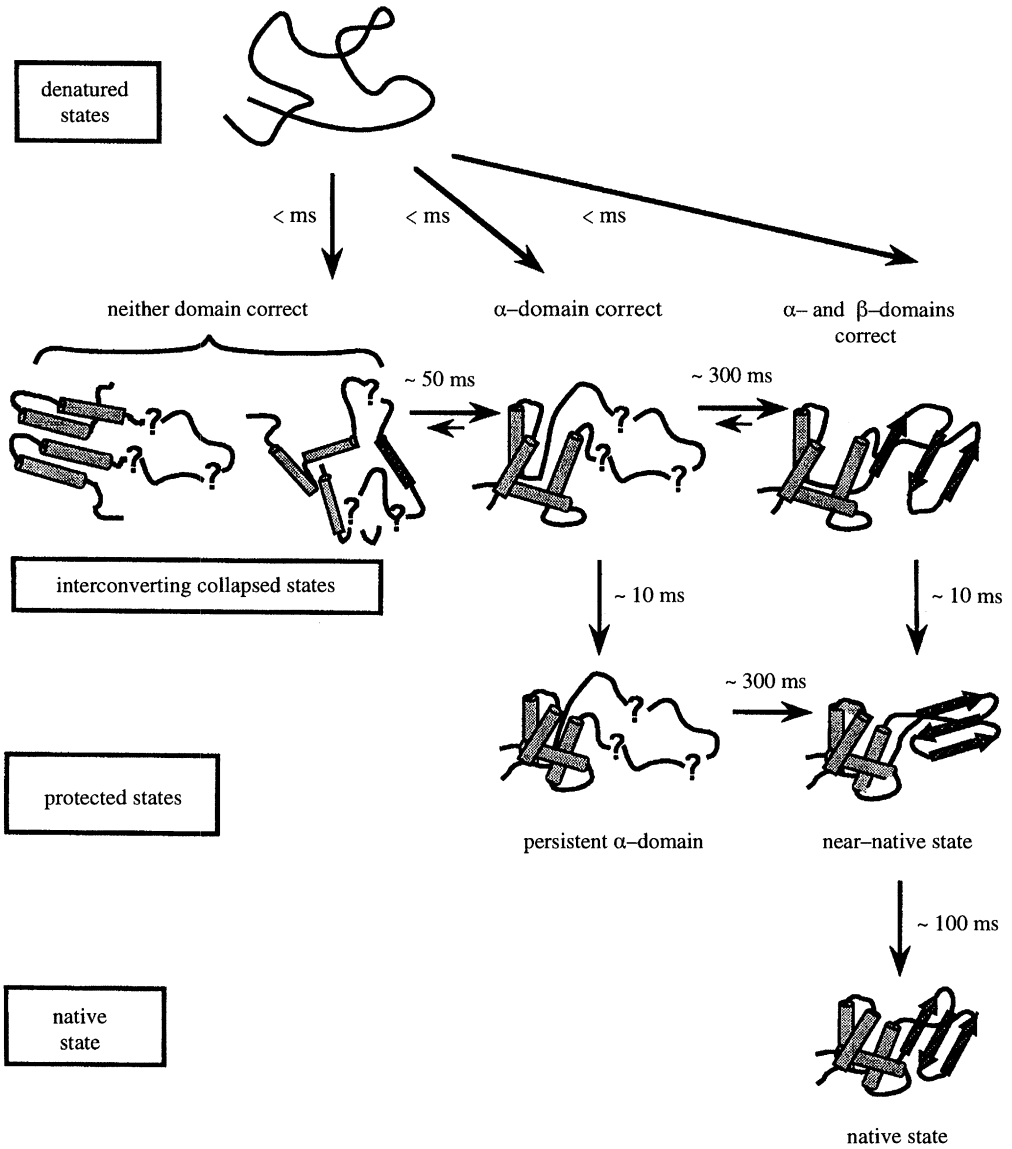

Figure 2. Schematic representation of a possible folding pathway of hen lysozyme deduced from the eight complementary methods listed in table 1. The four native disulphide bonds that remain intact throughout the folding process are not shown.

guanidinium chloride is unlikely to be the origin of this phenomenon (Kotik *et al.* 1995). Rather, we believe it arises from the existence of both correctly folded and misfolded secondary structure or the existence of distinct topological forms resulting from the rapid collapse of the polypeptide chain; recent theoretical predictions provide support for this general idea (Thirumalai & Guo 1995). These forms may be particularly significant in the folding of lysozyme because the polypeptide chain is constrained by the four native disulphide bonds. Studies of the folding of a three disulphide derivative of lysozyme lacking the disulphide bond (Cys 6-Cys 127) which links the N- and C-terminal regions shows, however, that although the presence of this disulphide bond stabilizes the α-domain intermediate, its loss does not appear to change the folding pathway fundamentally (Denton *et al.* 1994; Eyles *et al.* 1994).

At the early stage (< ms) of folding the structure formed is not persistent, indicating that these states are highly dynamic and rapidly fluctuating. Estimates of protection factors for the amide hydrogens in these states indicate that these are less than *ca.* 50. The formation of persistent structure occurs in the next steps along the pathway, and is indicated by protection factors in excess of 500 (Radford *et al.* 1992). Most molecules fold to form persistent structure in the α-domain on a *ca.* 50 ms timescale, whereas the formation of a stably structured β-domain takes significantly longer and is not complete in most molecules until at least 300 ms. On the pathway shown in figure 2, however, a fast folding population of molecules is seen to form persistent structure in both domains on a *ca.* 10 ms timescale. The fact that no such phase is seen in the fluorescent inhibitor binding experiments shows that the active site cleft has not yet formed in these molecules and suggests that another step is needed to bring the domains together to form the fully native state. The timescale of this process is not resolved from that of formation of the native enzyme in the population of slow folding molecules, suggesting that if this step is common to both pathways, it must occur on

Table 1. *Different techniques used to determine the kinetic folding pathway of hen lysozyme*

(All experiments were performed at pH 5.2, 20 °C. The initial denatured state was formed in 6 M guanidinium chloride and refolding was initiated by 11-fold dilution into the refolding buffer. Relevant references to this work are given in the text. HX, hydrogen exchange.)

method	information	result
HX labelling and NMR		
amides	formation of persistent secondary structure	parallel pathways observed: the major folding intermediate is protected in the α-domain
Trp indoles	formation of persistent tertiary structure	hydrophobic core formation precedes α-domain stabilization
HX labelling and ESI-MS	detection of intermediates and folding populations	parallel paths confirmed: independent folding of α-domain, but not the β-domain, demonstrated
far UV CD	formation secondary structure (and possibly some tertiary interactions)	(i) near native secondary structure formed within 2 ms (ii) ellipticity greater than that of the native structure seen after *ca.* 50 ms (iii) slowest phase monitors formation of the ellipticity of the native state
near UV CD	immobilization of aromatic residues	single exponential similar to (iii) above
intrinsic fluorescence	environment of Trp and Tyr residues	(i) change in environment of Trp residues occurs within 2 ms (ii) excess fluorescence quenching in intermediate at *ca.* 50 ms (iii) slow phase monitors recovery of fluorescence to value of native state
fluorescence quenching	solvent accessibility of fluorophore	(i) substantial protection in 2 ms indicates burial of Trp residues (ii) native-like resistance to quenching achieved within *ca.* 100 ms
ANS binding	exposure of hydrophobic surface	2 ms intermediate shows maximum fluorescence enhancement with ANS
inhibitor binding	formation of native state	native state forms in the slowest observable phase; only a single kinetic step is observed at 20 °C

a timescale of about 100 ms. Preliminary results investigating the temperature dependence of folding support this result (A. Matagne, S.E. Radford & C.M. Dobson, unpublished data).

The existence of a schematic pathway is of considerable value in the design of further experimental strategies to test it. Particular areas of interest concern the nature of the collapsed states, the apparent inability of the β-domain to fold to stable structure independently of the α-domain, and the origin of the cooperativity involved in the formation of the structural domains. One strategy we have adopted to explore the former is to probe the development of exchange protection of side chain hydrogens of tryptophan residues (see table 1). Lysozyme has six such residues, four of which are located within the hydrophobic core in the α-domain of the native protein (Blake *et al.* 1965). The kinetics of protection of these hydrogens have been studied in detail, and preliminary data suggest that for several of these protection occurs, at least to a limited extent, before the development of the stable α-domain (Radford *et al.* 1992; C. Morgan, A. Miranker & C.M. Dobson, unpublished data). This, like the ANS binding and fluorescence quenching experiments, suggests that a rudimentary hydrophobic core, at the very least, is formed early during the folding process, from, or within which, the development of persistent secondary structure takes place.

A further approach is to examine the folding of lysozymes from different species. The most important finding from these studies so far is that the refolding of the human variant (which is about 40% identical in sequence to hen lysozyme), while resembling that of the hen protein in outline, differs from it in detail (Hooke *et al.* 1994). Specifically, although refolding studies of the human protein show that the α-domain forms before the β-domain in most molecules, formation of the former domain is not fully cooperative and amides located in the A-, B- and C-terminal 3_{10} helices become protected before those in the C- and D-helices. Thus it seems that this region of structure forms a distinct stable subdomain in the folding process. Although such a subdomain is not observed in the hen protein, its absence suggests that if it is formed it is not stable enough to resist hydrogen-exchange labelling. Although specific site-directed mutants will be needed to probe the molecular origin of such a difference, the observation itself provides evidence that the co-operative but local assembly of regions of secondary structure is a key aspect of folding.

4. THE STRUCTURES OF FOLDING INTERMEDIATES

The combination of the different kinetic techniques described above enables the structural characteristics

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

of the various intermediates to be inferred from these experiments. To test the conclusions from such studies, however, much more detailed structural studies are required. One approach to achieving this is to produce and study stable analogues of the transient intermediates which allow the full potential of NMR in the elucidation of the structure and dynamics of these states to be realized (Baum *et al.* 1989; Hughson *et al.* 1990; Darby *et al.* 1991; Alexandrescu *et al.* 1993). If characteristics such as their optical and hydrogen exchange properties can be correlated with those of their transient counterparts, then the conclusions from detailed studies of the stable analogues could provide great insight into the nature of the species formed in the kinetic folding process.

One approach to this is to study peptide fragments of the protein that are unable to fold to a native structure in the absence of the remainder of the polypeptide chain. The characterization of peptides has been widely used to search for conformational preferences in different regions of a sequence to provide clues as to the origin of early events in folding (Dyson *et al.* 1992; Kemmink & Creighton 1993; Waltho *et al.* 1993; Wu *et al.* 1993). A complementary strategy is to study fragments designed to probe interactions between regions of the protein in intermediate states (Oas & Kim 1988; Staley & Kim 1990). In the case of lysozyme we have synthesized the entire polypeptide chain in four segments with this purpose in mind. One result of particular interest is the discovery that the conformation of a peptide corresponding to the major β-sheet of the protein is able to adopt a variety of conformational states, depending on the solution conditions and its oligomeric state (Yang *et al.* 1994). This might provide clues as to the origin of misfolding events if indeed these are involved in the slow folding of the β-domain in the major pathway of folding. Detailed structural studies of the remaining three peptides by NMR are currently underway.

Another approach to the identification of species whose structures are potentially related to those of folding intermediates is to expose intact proteins to mild denaturing conditions. This strategy is of particular importance to our studies of the folding of lysozyme because α-lactalbumins, which are structural homologues of lysozymes, form partly folded states during equilibrium unfolding, for example at low pH (Kuwajima 1989). In studies using optical techniques such states (known as molten globules) have been correlated with early intermediates in the folding of both α-lactalbumins and lysozymes (Kuwajima 1989). Furthermore, NMR and ESI-MS studies have shown that these states are only very weakly protected from hydrogen exchange (the protection factors of the majority of amides are less than than ten (Baum *et al.* 1989; Chyan *et al.* 1993; Buck *et al.* 1994) and that the protected amides are located predominately in the α-domain (Baum *et al.* 1989; Chyan *et al.* 1993). In these respects they resemble most closely the molecules populated in the first few milliseconds of the lysozyme refolding experiments, suggesting that characterization of these stable states could provide considerable insight into the nature of very early events in the folding of lysozyme. In this regard an important recent development is that the entire α-domain of human α-lactalbumin has been generated by protein engineering techniques and shown to form a stable state at neutral pH in the absence of denaturants with characteristics of the molten globule state of the intact protein (Peng & Kim 1994).

Structural studies using NMR of the molten globule state of α-lactalbumin have shown that despite clear evidence for secondary structure and some regions of local tertiary structure, the side-chains of at least the large majority of residues are substantially disordered (Baum *et al.* 1989; Alexandrescu *et al.* 1993). One interesting possibility is that this disorder, reflected in the broadening of resonances in the NMR spectrum (Baum *et al.* 1989; Alexandrescu *et al.* 1993), includes the slow interconversion of species as diverse as those thought to be responsible for the heterogeneity of the kinetic folding of lysozyme discussed above. Finally, recent work has indicated that states related to the α-lactalbumin molten globules may exist for some lysozymes. One of these, from the equine protein, is intriguing because this lysozyme is unusual in that it binds calcium; in this respect it has characteristics of both the lysozymes and α-lactalbumins (Van Dael *et al.* 1993). A second example has recently been found for human lysozyme (Haezebrouck *et al.* 1995) and it is interesting to speculate that this difference from the hen protein, whose equilibrium folding is highly cooperative under all the conditions explored so far, is related to the differences in the kinetic refolding behaviour of the two proteins.

Given that the molten globule states of α-lactalbumin appear to resemble the collapsed states formed during the kinetic refolding of lysozyme, the question of finding models for other states on the folding pathway arises. Detailed NMR studies of both the native and highly denatured states of lysozyme have been reported (Evans *et al.* 1991; Smith *et al.* 1993). The former correlates very well with the structure of the protein in the crystalline state (Blake *et al.* 1965; Smith *et al.* 1993). The structure of lysozyme in guanidinium chloride or urea appears to differ significantly from that of an archetypal random coil; of particular interest is the significant involvement of hydrophobic residues in residual structure (Broadhurst *et al.* 1991; Evans *et al.* 1991). This structure, however, does not appear to be persistent (Buck *et al.* 1994), nor to affect the events occurring during refolding (Kotik *et al.* 1995). Models for the more persistently structured states have proved harder to generate, although a partly folded state of lysozyme that has stable secondary structure in the absence of well-defined tertiary structure can be formed in TFE solution (Buck *et al.* 1993). In this state the secondary structure is most persistent in regions of the protein that are helical in the native state. This suggests that an environment which stabilizes structure in a polypeptide, such as might occur in the early stages of collapse, could generate native-like elements of secondary structure through a highly non-specific mechanism.

Although the structure of an analogue of the highly structured molten globules formed late in folding has

not yet been characterized for lysozyme, structures of other proteins that are partly unfolded (at least locally) are now beginning to emerge which may share characteristics with such states (Feng *et al.* 1994; Redfield *et al.* 1994). At low pH, one such example is the four helix bundle protein, interleukin-4. Under these conditions, the protein has both near and far UV CD spectra that differ significantly from those of the native state, and an enhanced fluorescence in the presence of ANS (Redfield *et al.* 1994). It is, however, sufficiently native-like to allow its structure to be defined. This shows few significant differences from the fully native state, except the loss of several turns of one helix. Some 30% of the polypeptide chain, however, is significantly disordered, as indicated most dramatically by nuclear relaxation measurements. Whether or not this has any direct relation to folding intermediates remains to be established, but it stresses the point that regions of secondary structure can be highly persistent, able to protect strongly against hydrogen exchange, whereas other regions of the native structure are still substantially unstructured.

5. RELATING *IN VIVO* AND *IN VITRO* FOLDING EVENTS

The folding process of a protein *in vitro* (see, for example, that of hen lysozyme described in figure 2) is likely to differ in many details from that in the cell. Apart from the obvious differences in the refolding conditions (temperature, pH, ionic strength and protein concentration), the presence of disulphide bonds at the onset of folding is artificial and is likely to affect significantly the stability of the various intermediate states, as it does the native state (Doig & Williams 1991; Cooper *et al.* 1992). A large part of this stabilization is likely to result from the reduction in the entropy of unfolded states relative to more highly folded ones, rather than from direct stabilization of specific structural features of particular states. Comparison of folding in the absence and presence of disulphide bonds, therefore, promises to throw considerable light on factors stabilizing intermediates and on the relation of the folding process *in vitro* to the situation occurring *in vivo*. A further complication in this comparison arises, however, as folding of proteins in the cell is assisted by proteins acting as catalysts such as protein disulphide isomerase and prolyl isomerase (Jaenicke 1993; Freedman *et al.* 1994). In addition, proteins such as the molecular chaperones, which are not thought to catalyse folding reactions but to increase the yield of the native protein by preventing aggregation of partly folded intermediates, are known to be important in folding in the cell (Ellis 1994; Hartl *et al.* 1994).

On the chaperone-assisted folding pathway the chaperonins (named GroEL in *Escherichia coli*) are thought to facilitate folding of their substrate proteins by providing a central cavity within which folding might take place in a protected environment (Langer *et al.* 1992; Chen *et al.* 1994). However, recent evidence using a mutant form of GroEL which is able to bind the substrate protein but not release its product, has suggested that even in the chaperone-assisted pathway a substantial portion of the folding reaction might take place in the cytosol (Weissman *et al.* 1994). One of the major challenges, therefore, is to determine the relation between folding pathways determined *in vitro* with the situation occurring *in vivo*. Deciphering the molecular details of folding pathways *in vivo* is, however, no trivial task; not only because of the complexity of the GroEL tetradecamer (Braig *et al.* 1994) and the limited structural information about its interaction with the co-chaperone GroES (Landry & Gierasch 1994), but also in that elucidation of the structure of the bound substrate protein in the presence of the chaperonin is also required. From studies of ANS binding, intrinsic tryptophan fluorescence and proteinase sensitivity there is some evidence that the bound state has molten-globular characteristics (Martin & al 1991). Further characteristics of the nature of such a state, however, requires more detailed information such as that described above for intermediates formed *in vitro*.

How this can be achieved, however, is not obvious, as the size of the GroEL oligomer (> 800 kDa) is orders of magnitude larger than is amenable to direct NMR studies, precluding direct measurement of hydrogen exchange protection in the substrate protein by this usually powerful method.

One approach to overcoming this problem is to permit hydrogen exchange to take place within the intact complex, then to dissociate the components before recording the NMR spectrum of the released protein ligand. Such a procedure has been used to establish that cyclophilin bound to GroEL is substantially less stable than the native protein, and is in a state unable to protect amide hydrogens sufficiently to allow their detection (Zahn *et al.* 1994). However, this experiment could only put an upper limit of *ca.* 1000 on the protection factors in the bound state; such a value would not exclude any of the states on the lysozyme folding pathway except the native one. The success of the ESI-MS experiments described above to monitor hydrogen exchange has, therefore, prompted us to exploit this technique to characterize the structural details of a state of α-lactalbumin, which by reduction of a single disulphide bond and rearrangement of the remaining three disulphides has been shown to form a stable complex with GroEL (Hayer-Hartl *et al.* 1994; Robinson *et al.* 1994).

The idea behind this experiment is shown in figure 3. Hydrogen exchange is allowed to take place within the complex, but in this case the rate of hydrogen exchange is measured directly and in real time by introducing the intact complex into the mass spectrometer. Once this has occurred, the rapid loss of water from the protein in the gas phase prevents further exchange from taking place. By adjusting the characteristics of the mass spectrometer, conditions can be found under which the complex will dissociate in the mass spectrometer without perturbing the pattern of labelling in the protein ligand, allowing the mass of the α-lactalbumin molecules to be determined and hence the exchange process to be monitored. This experimental strategy has recently been put into practice (Robinson *et al.* 1994). The degree of protection measured for the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

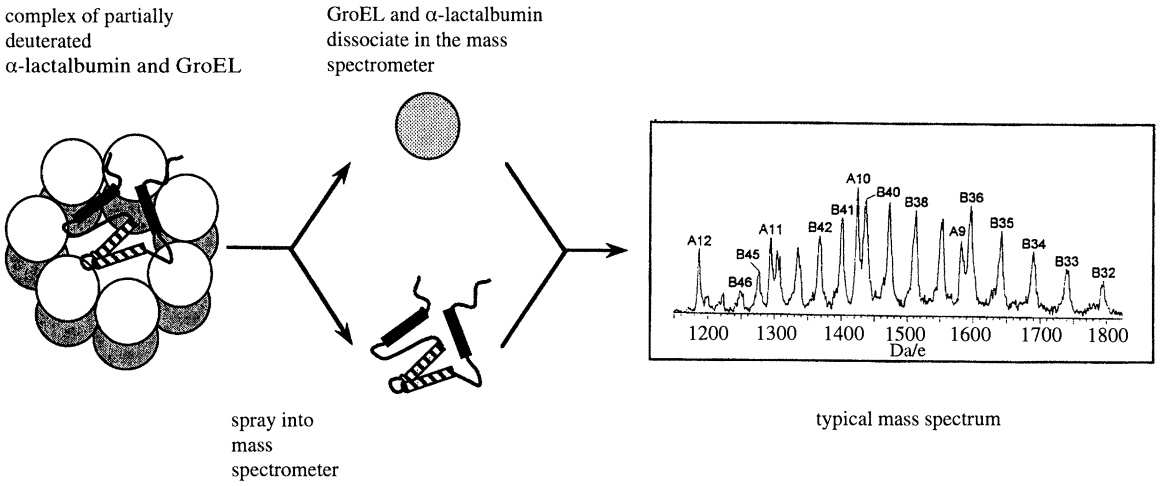

Figure 3. Schematic diagram of the experiment designed to monitor hydrogen exchange in α-lactalbumin bound to GroEL. The complex of GroEL with α-lactalbumin (which had previously been deuterated at all exchangeable sites) is diluted into H_2O and the rate of hydrogen exchange is measured directly by ESI-MS. In the diagram helices shaded with stripes denote regions protected from exchange, helices coloured black represent exchange labile regions. A typical mass spectrum of the complex is also shown. The charge state series labelled A arises from α-lactalbumin, that labelled B arises from the GroEL monomer (Robinson *et al.* 1994).

α-lactalbumin derivative bound to GroEL is much lower than that for the native state, in accord with the results for GroEL-bound cyclophilin (Zahn *et al.* 1994). What is even more striking, however, is that the bound protein is more highly protected than expected for a highly unfolded state. The extent of protection is very similar to that observed for a three-disulphide derivative of α-lactalbumin that exists in a molten globule state under the conditions of these experiments (Ewbank & Creighton 1993; Robinson *et al.* 1994). This suggests that GroEL interacts with folding intermediates having structures analogous to the disordered molten globules such as those formed in the early stages of the lysozyme folding pathway. In accord with this, recent experiments using intrinsic fluorescence measurements suggest that GroEL binds to the early collapsed states formed in the lysozyme folding pathway (M.P. Botwood & S.E. Radford, unpublished data).

Mass spectrometry has emerged, therefore, as a powerful method to complement NMR in monitoring hydrogen exchange. It provides novel information about the cooperativity of exchange processes, and about the populations of individual species having different numbers of protected amides. It can be applied to proteins of greater mass than can readily be studied by NMR, and requires only very small quantities of material. Information at the level of individual residues or regions of the sequence can be obtained by relating the results to those from complimentary NMR studies. In addition, we are exploring the possibility that specific hydrogen exchange information can also be obtained, at least in favourable cases, by using procedures that fragment the protein within the mass spectrometer. If such a procedure were to prove viable for protein molecules released from GroEL in the gas phase, then further details of the nature of the bound state are certain to emerge. Given that ESI-MS is in its infancy, the long term prospects for its widespread application in folding studies look excellent.

7. CONCLUSIONS

The studies of lysozyme and α-lactalbumin summarized here have revealed the power of using a battery of physical techniques to complement NMR in an attempt to describe the structural transitions occurring during the folding of a protein. The results have also enabled us to speculate on the nature of these states, and how this relates to the mechanism by which proteins are able to fold efficiently and accurately. Key ideas include the fact that a rapid collapse to a state with a rudimentary hydrophobic core could stabilize or even promote the formation of native-like secondary structure, and limit enormously the extent of conformational space that needs then to be explored to locate the native state itself. The observation of kinetically distinct populations of molecules emerging from the collapsed state supports the idea that only a proportion of molecules have attained a native-like topology in this process, and that the remainder fold more slowly because reorganizational events are needed before folding can proceed to the native state. The probability that misfolding occurs within a population of molecules is likely to be larger for longer polypeptide chains, and this might provide a considerable driving force towards the evolution of domains able to fold to a large extent independently of other regions of the structure. It would also be consistent with the fact that several small proteins, containing both β-sheets and α-helices, appear to fold extremely efficiently without the generation of molecules unable to fold rapidly from the collapsed state (Jackson & Fersht 1991; Briggs & Roder 1992; Kragelund *et al.* 1995).

As protein folding is usually fast, extremely accurate, and often rather insensitive to changes in conditions and even amino acid sequence, it seems likely that it must depend on rather simple universal characteristics of protein sequences and structure, such as their patterns of hydrophobic and hydrophilic residues (Dill

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

et al. 1993). The importance of these patterns in determining characteristic folds is evident from their successful utilization in methods designed to predict which of a set of known structures a given sequence is most likely to favour (Bowie *et al.* 1991; Jones & Thornton 1993). Furthermore, it would be remarkable if the features of proteins that have developed to enable a protein to find a unique fold *in vivo* from an apparently overwhelming number of alternatives could by chance allow the same solution to be achieved by a fundamentally different route *in vitro*. In the case of α-lactalbumin discussed here, it is evident that GroEL binds to rather disordered but compact species corresponding most closely to species formed early during the folding of lysozyme. Indeed, it is tempting to speculate if this is the case that it might act to destabilize misfolded species and to avoid the trapping of such states that might otherwise fail to undergo further folding or might simply aggregate.

We acknowledge support from the EPSRC, BBSRC and MRC through the Oxford Centre for Molecular Sciences. S.E.R. is a Royal Society 1983 University Research Fellow. The research of C.M.D. is supported in part by an International Research Scholars Award from the Howard Hughes Medical Foundation and by a Leverhulme Fellowship from the Royal Society. We thank our collaborators who have contributed to this work in many ways, especially P.A. Evans (University of Cambridge) for his invaluable contribution to the *in vitro* studies of lysozyme and α-lactalbumin.

REFERENCES

Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J. & Dobson, C.M. 1993 Structure and dynamics of the acid-denatured molten globule state of α-lactalbumin: a two-dimensional NMR study. *Biochemistry* **32**, 1707–1718.

Baum, J., Dobson, C.M., Evans, P.A. & Hanley, C. 1989 Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig α-lactalbumin. *Biochemistry* **28**, 7–13.

Blake, C.C.F., Koenig, D.F., Mair, G.A. & Sarma, R. 1965 Crystal structure of lysozyme by X-ray diffraction. *Nature, Lond.* **206**, 757–761.

Bowie, J.U., Lüthy, R. & Eisenberg, D. 1991 A method to identifying protein sequences that fold into a known three-dimensional structure. *Science, Wash* **253**, 164–170.

Braig, K., Otwinowski, Z., Hedge, R. *et al.* 1994 The crystal structure of the bacterial chaperonin GroEL at 2.8Å. *Nature, Lond.* **371**, 578–586.

Briggs, M.S. & Roder, H. 1992 Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2017–2021.

Broadhurst, R.W., Dobson, C.M., Hore, P.J., Radford, S.E. & Rees, M.L. 1991 A photochemically induced dynamic nuclear polarization study of denatured states of lysozyme. *Biochemistry* **30**, 405–412.

Buck, M., Radford, S.E. & Dobson, C.M. 1993 A partially folded state of hen egg white lysozyme in trifluoroethanol: structural characterization and implications for protein folding. *Biochemistry* **32**, 669–678.

Buck, M., Radford, S.E. & Dobson, C.M. 1994 Amide hydrogen exchange in a highly denatured state of hen egg white lysozyme in urea. *J. molec. Biol.* **237**, 7–254.

Bychkova, V.E. & Ptitsyn, O.B. 1994 The molten globule in vitro and in vivo. *Chemtracts Biochem. molec. Biol.* **4**, 133–163.

Chaffotte, A., Guillou, Y. & Goldberg, M.E. 1992 Kinetic resolution of peptide bond and side chain far-UV circular dichroism during the folding of hen egg white lysozyme. *Biochemistry* **31**, 9694–9702.

Chen, S., Roseman, A.M., Hunter, A.S. *et al.* 1994 Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. *Nature, Lond.* **371**, 261–264.

Chyan, C., Wormald, C., Dobson, C.M., Evans, P.A. & Baum, J. 1993 Structure and stability of the molten globule state of guinea pig α-lactalbumin: a hydrogen exchange study. *Biochemistry* **32**, 5681–5691.

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

Darby, N.J., van Mierlo, C.P.M. & Creighton, T.E. 1991 The 5–55 disulphide intermediate in folding of BPTI. *FEBS Lett.* **279**, 61–64.

Denton, M.E., Rothwarf, D.M. & Scheraga, H.A. 1994 Kinetics of folding of guanidine-denatured hen egg white lysozyme and carboxymethyl (Cys6, Cys127)-lysozyme: A stopped flow absorbance and fluorescence study. *Biochemistry* **33**, 11225–11236.

Dill, K.A., Fiebig, K.M. & Chan, H.S. 1993 Cooperativity in protein folding kinetics. *Proc. natn. Acad. Sci. U.S.A.* **90** 1942–1946.

Dobson, C.M., Evans, P.A. & Radford, S.E. 1994 Understanding how proteins fold: the lysozyme story so far. *TIBS* **19**, 31–37.

Doig, A.J. & Williams, D.H. 1991 Is the hydrophobic effect stabilizing or destabilizing in proteins? The contribution of disulphide bonds to protein stability. *J. molec. Biol.* **217**, 389–398.

Dyson, H.J., Sayre, J.R., Merutka, G., Shin, H.-C., Lerner, R.A. & Wright, P.E. 1992 Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II. Plastocyanin. *J. molec. Biol.* **226**, 819–835.

Ellis, R.J. 1994 Roles of molecular chaperones in protein folding. *Curr. Opin. struct. Biol.* **4**, 117–122.

Evans, P.A. & Radford, S.E. 1994 Probing the structure of folding intermediates. *Curr. Opin. struct. Biol.* **4**, 100–106.

Evans, P.A., Topping, K.D., Woolfson, D.N. & Dobson, C.M. 1991 Hydrophobic clustering in non-native states of a protein: interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins Struct. Funct. Genet.* **9**, 248–266.

Ewbank, J.J. & Creighton, T.E. 1993 Structural characterization of the disulphide folding intermediates of bovine α-lactalbumin *Biochemistry* **32**, 3694–3707.

Eyles, S.J., Robinson, C.V., Radford, S.E. & Dobson, C.M. 1994 Kinetic consequences of removal of a disulphide bond on the folding of hen lysozyme. *Biochemistry.* **33**, 1534–1538.

Feng, Y., Sligar, S.G. & Wand, A.J. 1994 Solution structure of apocytochrome b562. *Nature struct. Biol.* **1**, 31–35.

Freedman, R.B., Hirst, T.R. & Tuite, M.F. 1994 Protein disulphide isomerase: building bridges in protein folding. *TIBS* **19**, 331–336.

Frydman, J., Nimmesgern, E., Ohtsuka, K. & Hartl, F.U. 1994 Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature, Lond.* **370**, 111–117.

Gray, T.E. & Fersht, A.R. 1993 Refolding of barnase in the presence of GroE. *J. molec. Biol.* **232**, 1197–1207.

Haezebrouck, P., Joniau, M., Van Dael, H., Hooke, S.D., Woodruff, N.D. & Dobson, C.M. 1995 An equilibrium molten globule state of human lyszoyme at low pH. *J. molec. Biol.* **246**, 384–389.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Hartl, F.-U., Hlodan, R. & Langer, T. 1994 Molecular chaperones in protein folding: the art of avoiding sticky situations. *TIBS* **19**, 20–25.

Hayer-Hartl, M., Ewbank, J.J., Creighton, T.E. & Hartl, F.U. 1994 Conformational specificity of the chaperonin GroE for the compact folding intermediates of α-lactalbumin. *EMBO J.* **13**, 3192–3202.

Hooke, S.D., Radford, S.E. & Dobson, C.M. 1994 The refolding of human lysozyme: a comparison with the structurally homologous hen lysozyme. *Biochemistry* **33**, 5867–5876.

Hughson, F.M., Wright, P.E. & Baldwin, R.L. 1990 Structural characterisation of a partly folded apomyoglobin intermediate. *Science, Wash.* **249**, 1544–1548.

Imoto, T., Johnson, L.N., North, A.C.T., Philips, D.C. & Rupley, J.A. 1972 *The enzymes* (ed. P.D. Boyer), 3rd edn, vol. 7, pp. 666–867. Orlando, Florida: Academic Press.

Itzhaki, L.S., Evans, P.A., Dobson, C.M. & Radford, S.E. 1994 Tertiary interactions in the folding pathway of hen lysozyme: kinetic studies using fluorescent probes. *Biochemistry* **33**, 5212–5220.

Jackson, G.S., Staniforth, R.A., Halsall, D.J. *et al.* 1993 Binding and hydrolysis of nucleotides in the chaperone catalytic cycle: implications for the mechanism of assisted protein folding. *Biochemistry* **32**, 2554–2563.

Jackson, S.E. & Fersht, A.R. 1991 Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428–10433.

Jaenicke, R. 1993 Role of accessory proteins in protein folding. *Curr. Opin. struct. Biology*, **3**, 104–112.

Jones, J. & Thornton, J. 1993 Protein fold recognition. *J. comp. molec. Des.* **7**, 439–456.

Kemmink, J. & Creighton, T.E. 1993 Local conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor and their roles in folding. *J. molec. Biol.* **234**, 861–878.

Koide, S., Dyson, H.J. & Wright, P.E. 1993 Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry* **32**, 12299–12310.

Kotik, M., Radford, S.E. & Dobson, C.M. 1995 The folding pathway of lysozyme denatured in guanidine hydrochloride and in DMSO. *Biochemistry* **34**, 1714–1724.

Kragelund, B.B., Robinson, C.V., Knudsen, J., Dobson, C.M. & Poulsen, F.M. 1995 A simple protein structural motif folds efficiently and cooperatively. *Biochemistry* (In the press.)

Kraulis, P.J. 1991 Molscript: a program to produce both detailed and schematic plots of protein structures. *J. appl. Cryst.* **24**, 946–950.

Kuwajima, K. 1989 The molten globule as a clue for understanding the folding and cooperativity of globular protein structures. *Proteins Struct. Funct. Genet.* **6**, 87–103.

Landry, S.J. & Gierasch, L.M. 1994 Polypeptide interactions with molecular chaperones and their relationship to in vivo protein folding. *A. Rev. Biophys. Biomol. Stuct.* **23**, 645–669.

Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F.U. 1992 Chaperone-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO J.* **11**, 4757–4765.

Levinthal, C. 1968 Are there pathways for protein folding. *J. Chim. Phys.* **65**, 44–45.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L. & Hartl, F.-U. 1991 Chaperonin-mediated protein folding at the surface of GroEL through a 'molten globule'-like intermediate. *Nature, Lond.* **352**, 36–42.

Miranker, A., Krupa, G.H., Robinson, C.V., Aplin, R.T. & Dobson, C.M. 1995 Site specific analysis of proteins by mass spectrometry. (Submitted.)

Miranker, A., Robinson, C.V., Radford, S.E., Aplin, R.T. & Dobson, C.M. 1993 Detection of transient protein folding populations by mass spectrometry. *Science, Wash.* **262**, 896–899.

Oas, T.G. & Kim, P.S. 1988 A peptide model of a protein folding intermediate. *Nature, Lond.* **336**, 42–47.

Peng, Z.Y. & Kim, P.S. 1994 A protein dissection study of a molten globule. *Biochemistry* **33**, 2136–2141.

Radford, S.E., Dobson, C.M. & Evans, P.A. 1992 The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature, Lond.* **358**, 302–307.

Redfield, C., Smith, R.A.G. & Dobson, C.M. 1994 Structural characterisation of a highly-ordered 'molten globule' at low pH. *Nature struct. Biol.* **1**, 23–29.

Robinson, C.V., Groß, M., Eyles, S.J. *et al.* 1994 Hydrogen exchange protection in GroEL-bound α-lactalbumin probed by mass spectrometry. *Nature, Lond.* **372**, 646–651.

Roder, H., Elöve, G.A. & Englander, S.W. 1988 Structural characterization of folding intermediates in cytochrome c by hydrogen exchange labelling and proton NMR. *Nature, Lond.* **335**, 700–704.

Sali, A., Shakhnovich, E. & Karplus, M. 1994 How does a protein fold? *Nature, Lond.* **369**, 248–251.

Senko, M. W. & McLafferty, F. W. 1994 Mass spectrometry of macromolecules: has its time now come? *A. Rev. Biophys. Biomolec. Struc.* **23**, 763–785.

Smith, L., Sutcliffe, M.J., Redfield, C. & Dobson, C.M. 1993 Structure of hen lysozyme in solution. *J. molec. Biol.* **229**, 930–944.

Staley, J.P. & Kim, P.S. 1990 Role of a subdomain in the folding of BPTI. *Nature, Lond.* **344**, 685–688.

Thirumalai, D. & Guo, Z. 1995 Nucleation mechanism for protein folding and theoretical predictions for hydrogen-exchange labelling experiments *Biopolymers* **35**, 137–140.

Udgaonkar, J.B. & Baldwin, R.L. 1988 NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature, Lond.* **335**, 694–699.

Van Dael, H., Haezebrouck, P., Morozova, L., Arico-Muendel, C. & Dobson, C.M. 1993 Partially folded states of equine lysozyme: Structural characterization and significance for protein folding. *Biochemistry* **32**, 11886–11894.

Waltho, J.P., Feher, V.A., Merutka, G., Dyson, H.J. & Wright, P.E. 1993 Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *Biochemistry* **32**, 6337–6347.

Weissman, J.S., Kashi, Y., Fenton, W.A. & Horwich, A.L. 1994 GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78**, 693–702.

Wu, L.C., Laub, P.B., Elöve, G.A., Carey, J. & Roder, H. 1993 A non-covalent peptide complex as a model for an early folding intermediate of cytochrome c. *Biochemistry* **32**, 10271–10276.

Yang, J.J., Pitkeathly, M. & Radford, S.E. 1994 Far UV-circular dichroism reveals a conformational switch in a peptide fragment from the β-sheet of hen lysozyme. *Biochemistry* **33**, 7345–7353.

Zahn, R., Spitzfaden, C., Ottiger, M., Wüthrich, K. & Plückthun, A. 1994 *Nature, Lond.* **368**, 261–265.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES