2

Protein Folding and its Implications for the Production of Recombinant Proteins

ROMAN HLODAN*, STEWART CRAIG† AND ROGER H. PAIN*

* *Department of Biochemistry and Genetics, The University of Newcastle upon Tyne, UK and* † *British Biotechnology Ltd, Cowley, Oxford, UK*

Protein folding: the basic problem facing the biotechnologist

The ultimate biotechnological achievement would be to regenerate a fertile egg from its three minutes boiled counterpart. A 'living' entity would have been created from a patently dead assembly of macromolecules. Perutz introduced a meeting on Protein folding (Jaenicke, 1980) by drawing attention to the work of Anson and Mirsky that established the basic principles for such a process of renaturation. They showed that heat-precipitated egg albumin can be resolubilized and 'refolded' to native albumin. The principal change that is being reversed is not chemical, involving covalent change, but physico-chemical, involving non-covalent and conformational changes.

Proteins fold *in vitro* because the three-dimensional functional state is more stable than the fully unfolded state (Anfinsen, 1967). This thermodynamic statement underlies all the basic and biotechnological research aimed at renaturing proteins. It is, however, research on the kinetic process that has provided insights into how the protein folds up. While each protein poses its own problems in folding and has to be considered on its own merits—or more usually, *de*merits!—it is this research that has provided basic technical expertise and some underlying rationale on which the biotechnology industry has been able to build when faced with recombinant proteins in the form of an insoluble, inactive and apparently intractable gunge. The oft-voiced com-

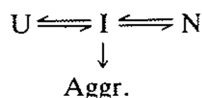
---

Abbreviations: ANS, anilino-naphthalene sulphonic acid; bGH, bovine growth hormone; BPTI, bovine pancreatic trypsin inhibitor; CD, circular dichroism; EDTA, ethylenediaminetetraacetic acid; *E. coli*, *Escherichia coli*; FMD, foot-and-mouth disease; FPLC, fast protein liquid chromatography; GuHCl, guanidine hydrochloride; hGH, human growth hormone; HPLC, high performance liquid chromatography; IFN, interferon; IGF, insulin-like growth factor; IL-, interleukin; LDH, lactate dehydrogenase; MBP, maltose binding protein; MES, 2-[*N*-morpholino]ethanesulphonic acid; NMR, nuclear magnetic resonance; PA, penicillin acylase; PDGF-β, platelet-derived growth factor β; PDI, protein disulphide isomerase; PGK, phosphoglycerate kinase; PPI, peptidyl prolyl *cis–trans* isomerase;RBP, ribose binding protein; RNase A, ribonuclease A; Rubisco, ribulose bisphosphate carboxylase; SDS, sodium dodecyl sulphate; TGF, transforming growth factor; TNF, tumour necrosis factor; tPA, tissue plasminogen activator; TSP, tailspike protein.

Biotechnology and Genetic Engineering Reviews — Vol. 9, December 1991
0264–8725/91/09/47–88 $20.00 + $0.00 © Intercept Ltd, P.O. Box 716, Andover, Hampshire SP10 1YG, UK

plaint that research on protein folding is not relevant to how the protein folds *in vivo* is itself irrelevant, for the biotechnological problem is that of how to refold a protein reproducibly and in high yield *in vitro*.

Proteins fold from the unfolded state stabilized by denaturants such as guanidinium chloride or urea, through a series of conformational states of generally increasing order. The basic problem underlying attempts to renature or refold proteins lies in the fact that these intermediate states, under conditions where the native state is stable, exhibit a tendency to aggregate—a cul-de-sac reaction that competes effectively with the folding pathway to the functional protein. The intermediate state that persists for the longest period of time, and is therefore most prone to form aggregates, is the state I immediately before the rate-determining step.



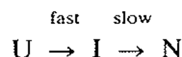
Recent work has shown that this state, or collection of states, is compact but 'sticky', due to an excess of solvent-exposed non-polar groups, and exhibits a tendency to aggregate. This tendency is particularly marked with folding subunits that will ultimately assemble into multisubunit proteins, and nature appears to have evolved a mechanism to combat this competing process. The folder's art is likewise to act as a chaperone who prevents improper liaisons between refolding polypeptides and allows them to realize their true potential.

This review attempts, therefore, to survey briefly the aspects of protein folding research that bear on the biotechnological problem and to follow this with a more detailed account of ways in which the problem has been tackled and frequently overcome. For those who wish to pursue the basic aspects in more detail, the recent review by Kim and Baldwin (1990) lists other recent reviews on various aspects of the field.

How proteins fold

THE EARLY STAGES OF PROTEIN FOLDING

It is becoming clear that the early stages of folding from the unfolded state, U, involve both rapid formation of secondary structure and a rapid collapse of the polypeptide chain into a compact intermediate state, I, from which the native protein, N, is then generated much more slowly.



What is not clear at present is whether secondary structure formation succeeds or precedes the collapse of the polypeptide chain or whether it is possible to generalize on this point for all proteins under all conditions of

folding. Until recently, it has proved difficult to characterize the early stages of protein folding. This is because of the speed with which they occur, within the time-range of milliseconds, making them difficult to detect.

Indications about how folding is initiated have come, first, from studies with peptides in solution. This has largely been possible due to recent developments in nuclear magnetic resonance (NMR) techniques (Wright, Dyson and Lerner, 1988). It has been found that quite small peptides are capable of taking up regular conformations. The structures so far detected do not adopt a stable conformation, but interconvert rapidly with non-structured forms.

Studies with peptides corresponding to the C-peptide of RNase A (Shoemaker *et al.*, 1987) and Pα, the terminal helix of bovine pancreatic trypsin inhibitor (BPTI) (Goodman and Kim, 1989) show that α-helical structures can be formed under favourable conditions with peptides of about 12 amino acids in length. Helix formation was found to be dependent on both pH and the amino-terminal charge. The conclusion drawn from both of these studies is that the introduction of charged groups can have a stabilizing effect. Stabilization of the helix dipole may serve to stabilize intermediates in protein folding. However, a more recent study has shown that helix formation in C-peptide is not stabilized solely by dipole interactions (Shoemaker *et al.*, 1990). An interaction between the imidazolium ring of the protonated form of His-12 and the side-chain ring of Phe-8, which is present in the native structure of RNase A, contributes to the stabilization of the peptide.

A reverse turn structure has been detected in a 12-residue peptide corresponding to residues 98–106 of the influenza virus haemagglutinin HA1 (Wright, Dyson and Lerner, 1988). It is also retained in a peptide as short as four or five amino acids. The isomeric form of the proline involved in the turn is important in maintaining the structure in aqueous solution, the *trans* isomer being preferred over the *cis*. The primary determinant of β-turn stability however, appears to be the amino acid at position 3 in the structure, where there is a strong preference for glycine.

Oas and Kim (1988) characterized by NMR a synthetic peptide from BPTI, in which the carboxy-terminal α-helical region is connected to a β-sheet structure by a native disulphide bond. The formation of this disulphide bond, between Cys-30 and Cys-51 is known to be of key importance in the refolding of BPTI. Most of the secondary structure present in the corresponding region of the native protein also occurs in the peptide, which demonstrates that folding can occur autonomously at localized regions along the polypeptide chain. The secondary structure was lost upon reduction of the disulphide, indicating the importance of disulphide bond formation in its maintenance.

These and similar studies show that there is a significant tendency for elements of a polypeptide chain to 'flicker in' to secondary structures, presumably for a long enough time to create a sufficient probability for mutual stabilization to give larger elements of structure (Anfinsen, 1972). They suggest also that the β-turns that occur frequently in folded proteins act positively to create a turn, rather than acting as a loose hinge. In this way they would speed the formation of antiparallel strand pairs and form structured

polar sites that would tend to locate on the surface of a compact intermediate state.

Recently, deuterium-exchange techniques have been applied to characterize the early folding intermediates in proteins (Roder, Elöve and Englander, 1988; Udgaonkar and Baldwin, 1988, 1990; Bycroft *et al.*, 1990). A similar approach was used in all of these cases. The native protein was unfolded in D_2O and all the amide hydrogen atoms exchanged for deuterium. Folding was initiated in D_2O before dilution, after varying time intervals, into H_2O-based solvent (*Figure 1*). Thus, all amide groups would be protonated except those where deuterium was trapped by the formation of a stable, hydrogen-bonded structure. Refolding was then allowed to proceed to completion in slow hydrogen-exchange conditions and two-dimensional NMR used to locate where D replaced amide H. In this way, the rates of formation of structures involving backbone amide groups participating in intramolecular hydrogen bonds, or the exclusion of amides from the solvent environment could be monitored. These results show rapid formation of some secondary structure: a hydrogen-bonded network corresponding to the amino- and carboxy-terminal helices of cytochrome *c* was formed within 20 ms of initiating refolding (Roder, Elöve and Englander, 1988). Other structures within the protein were formed at different times during folding (*Figure 1*). Rapid formation of secondary structure was shown with barnase (Bycroft *et al.*, 1990) but the amide deuterons of three probes in the protein not involved in regular secondary structure elements were protected in a time-course identical to that of the overall folding process. A β-pleated sheet was formed in ribonuclease in similarly short times (Udgaonkar and Baldwin, 1988).

THE FORMATION OF COMPACT INTERMEDIATE STATES

The stable, compact or 'molten globule' states of carbonic anhydrase (Semisotnov *et al.*, 1987) and β-lactamase (V. Uverski and R.H. Pain, unpublished data) have been shown to bind the fluorescent probe anilino-naphthalene sulphonic acid (ANS). In contrast, the fully unfolded proteins in denaturant bind no ANS and native protein binds relatively small amounts. This property has been used to detect the formation and accumulation during folding of *kinetic* compact intermediate states for several proteins of different structural type, ranging from high α to high β (Ptitsyn *et al.*, 1990). These states are characterized, like the stable states, by a high proportion of native-like secondary structure, but otherwise few stable tertiary interactions and no enzyme activity. The kinetic states are, however, more compactly folded than the stable intermediates, being only marginally less stable than the native state. They are formed rapidly within time-scales ranging from 10 to 100 ms. Further discussion of the compact intermediate and molten globule states will be found in reviews by Ptitsyn (1987), Kuwajima (1989), Kim and Baldwin (1990) and Christensen and Pain (1991).

The binding of ANS to the compact molten globule intermediates, both those which accumulate transiently under folding conditions and those which are stable in intermediate denaturant concentrations, highlights the fact that

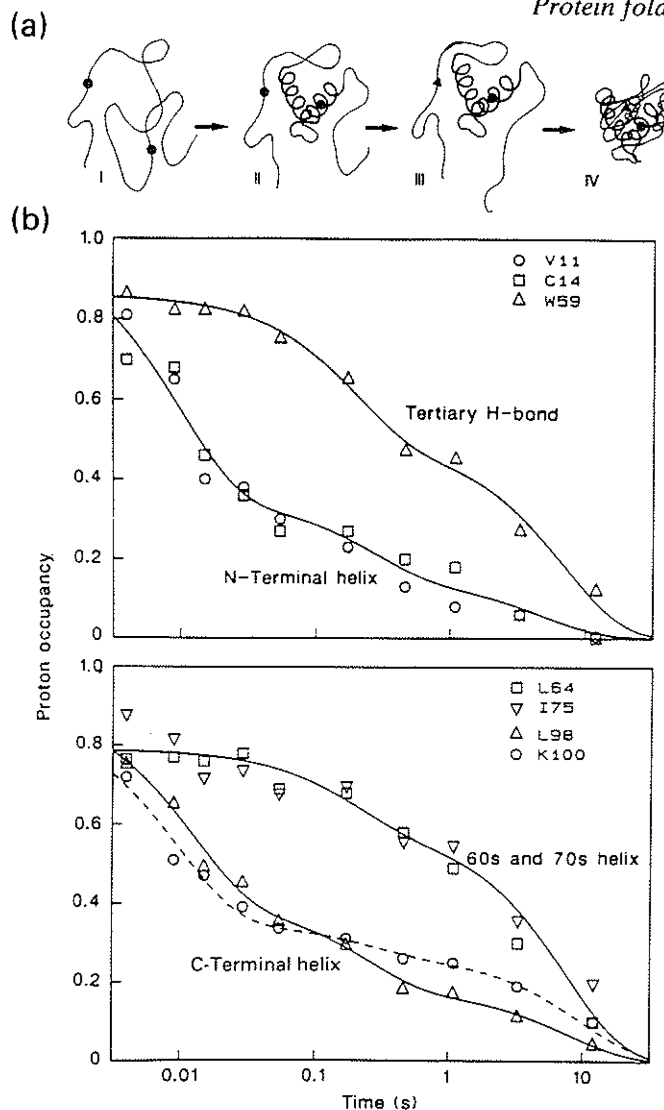


Figure 1. The identification of folding intermediates that can be characterized by NMR. (a) The protein is unfolded in guanidine hydrochloride (GuHCl) solution in D_2O, so that all exchangeable hydrogens become substituted by deuterium (I)—two such protons are represented as (●). Refolding is initiated by dilution into a D_2O-based solvent (II). At this stage of the experiment, intermediates may be formed in which exchangeable deuterons become protected from solvent. Dilution into excess H_2O, under conditions that favour rapid exchange, leads to protonation of all amides except for those protected by the formation of structure in the intermediate (III). This pulse is then terminated by refolding the protein under conditions where the exchange rate is slowest. In this way, amide sites involved at an early stage in structure formation would show little proton incorporation, whereas those involved at a later stage would show more (from Dobson and Evans, 1988). By increasing the time allowed for the protein to refold in D_2O-based solvent, the sequence of structure formation can be followed. In the case of cytochrome *c*, the extent of protection against hydrogen exchange for some of the backbone amide sites is shown in (b). Rapid protection from exchange (a low proton occupancy) is observed for NH sites in the N- and C-helices. Slower protection against exchange (a high proton occupancy) is observed at other sites along the polypeptide chain, including a central (60–80) helix. (From Roder, Elöve and Englander, 1988.)

they are 'sticky'. In general, therefore, they have a greater tendency than the native state to aggregate or to adhere to surfaces. Since the transient molten globule, or state I (*see* p. 50), comprises the most highly populated group of states on the folding pathway, it is at this stage that problems of aggregation and reduction of yield are most likely to occur. This point will be taken up again later in the review. The possible relevance of the molten globule state to protein translocation across membranes has been proposed (Bychkova, Pain and Ptitsyn, 1988). It is also a likely candidate for interaction with chaperone proteins (p. 62).

The question that the approaches mentioned so far have not been able to answer is, how do the collapsed folding intermediates compare with the native protein with respect to topology? Information about the 'nativeness' of the structures formed has been provided in an elegant study using monoclonal antibodies raised against the native protein as probes of folding (Blond and Goldberg, 1987). Rapid binding of antibody to the refolding β_2-subunit of tryptophan synthase demonstrated that native-like epitopes form very early. An additional slower phase to the reaction, however, indicated that the epitope changes slightly as further general conformational rearrangement to the native structure takes place.

THE RATE-DETERMINING STEPS IN PROTEIN FOLDING

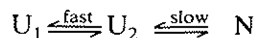
The rate-determining step of protein folding occurs after the rapid formation of the compact folding intermediate, which includes much secondary structure and from which the native state is generated. Forming a part of this slow stage of refolding are other processes, such as proline isomerization, and the formation and reshuffling of disulphide bonds.

Formation of the compact intermediate state can be viewed as being essentially an intra-domain event. Consistent with its proposed role in folding is the independent folding of individual domains in multidomain proteins. The domains of yeast phosphoglycerate kinase (PGK), for example, can refold independently (Adams, Burgess and Pain, 1985), as can those of an immunoglobulin light chain (Tsunenaga *et al.*, 1987) as well as the β_2-subunit of tryptophan synthase (Blond and Goldberg, 1987; Blond-Elguindi and Goldberg, 1990). The kinetics of *in vitro* renaturation of multidomain proteins, however, cannot be explained simply in terms of the independent refolding of domains. For example, an additional slow phase was detected in the refolding of an immunoglobulin light chain (Tsunenaga *et al.*, 1987) as well as the β_2-subunit of tryptophan synthase (Blond-Elguindi and Goldberg, 1990), suggesting that domains interact with each other during refolding *in vitro*, impeding the efficiency of folding. Further evidence for intramolecular interference being responsible for slowing the kinetics comes from studies on reduced and denatured serum albumin, which refolds more slowly than any of the three isolated domains (Johansen *et al.*, 1981).

Intermolecular association of intermediates during folding to the native state has also been documented, notably for bovine growth hormone (bGH) (Brems, 1988), but also for β-lactamase (V. Uverski and R.H. Pain,

unpublished results) and PGK (D. Dryden and R.H. Pain, unpublished results). While this might be expected to slow the rate of folding, no firm evidence yet exists that this is the case.

The proposal that *cis–trans* isomerization may play a determining role in protein folding arose as a result of the finding that several small monomeric proteins exhibit both fast and slow refolding kinetics to the native state. The refolding of thermally denatured bovine pancreatic RNase A, for example, was monitored by three criteria: the ability to bind inhibitor, enzyme activity and formation of tertiary structure (Garel and Baldwin, 1973). This work showed that there are two forms of the unfolded protein, spectroscopically indistinguishable, both lacking tertiary structure and the ability to bind inhibitor and yet which fold to the native state at very different rates. The model in its simplest form was proposed as follows:



$$U_1 \underset{}{\overset{\text{fast}}{\rightleftharpoons}} U_2 \underset{}{\overset{\text{slow}}{\rightleftharpoons}} N$$

This was confirmed by the diagnostic double-jump experiment, where the rapidly unfolded protein is refolded at varying times after unfolding. The amplitude of the slow refolding phase increases with time. Brandts, Halvorson and Brennan (1975) proposed that the slowly established equilibrium between U_1 and U_2 represents the kinetics of isomerization between the *cis* and *trans* forms of proline. This has been confirmed by studies on simple proline-containing peptides. The involvement of proline isomerization in protein folding has been shown in studies on the C fragment of an immunoglobulin light chain (Goto and Hamaguchi, 1982), pepsinogen (McPhie, 1980), *E. coli* thioredoxin (Kelley and Stellwagen, 1984), *iso*-cytochrome *c* (White, Berger and Wall, 1987) and RNase T1 (Kiefhaber *et al.*, 1990a,b). Supporting evidence for the slow folding phase being attributable to proline isomerization was provided in a study of the refolding of parvalbumin (Brandts, Brennan and Lin, 1977), which contains no proline residues. Though the kinetics were biphasic, the slower phase was much faster than that typical of proline isomerization, whereas in another parvalbumin containing a single proline residue, a slow refolding rate consistent with that observed for proline isomerization was observed (Lin and Brandts, 1978). The replacement of conserved prolines in thioredoxin (Kelley and Richards, 1987) and *iso*-2-cytochrome *c* (Wood *et al.*, 1988) also resulted in the loss of the slow phase of refolding of those proteins.

More complex models have been proposed in which the isomerization process may occur in a collapsed intermediate or native-like state as well as in the unfolded state. Prevention of the initial collapse may require the simultaneous occurrence in the unfolded state of a number of incorrect proline isomers.

The possibility that these limitations on rates of folding may not occur *in vivo* has been raised by the discovery of an enzyme which catalyses proline isomerization.

CATALYSIS OF PROLINE ISOMERIZATION BY PEPTIDYL PROLYL *CIS–TRANS* ISOMERASE

The enzyme peptidyl prolyl *cis–trans* isomerase (PPI) was first purified from porcine kidney (Fischer, Bang and Mech, 1984) and shown to catalyse the *cis–trans* isomerization of proline imide bonds in peptides. Subsequently, PPI has been shown to catalyse the folding of several proteins *in vitro* (Lang, Schmid and Fischer, 1987; Kiefhaber *et al.*, 1990b). However, its role *in vivo* remains to be fully established. PPI is also identical to cyclophilin (Fischer, Bang and Berger, 1989; Takahashi, Hayano and Suzuki, 1989), a protein which is inhibited by the immunosuppressive agent, cyclosporin A. Whether the immunosuppressive activity of cyclosporin A is related to inhibition of PPI is not yet clear. In folding studies with PPI, it has been possible to catalyse the slow folding phase of immunoglobulin light chain (Lang, Schmid and Fischer, 1987). The extent of acceleration of folding is dependent on PPI concentration: a molar ratio of PPI:light chain of 1:20 leads to a twofold increase in the rate of slow refolding. At high concentrations of PPI, the slow folding reaction coincides with the fast phase. With RNase A, partially folded intermediates are formed rapidly and more than one proline residue is made inaccessible to the enzyme. In this case, PPI does not catalyse slow refolding (Lang, Schmid and Fischer, 1987). Interestingly, the presence of the enzyme can alter the folding pathway of a protein. Unfolded RNase T1 can enter two possible folding pathways, due to the presence of more than one incorrect proline isomer. Kiefhaber *et al.* (1990b) have postulated that, depending upon which one of two particular prolines isomerizes first, RNase T1 will refold through either of two pathways, one slower than the other. Addition of PPI shifted the relative amount of protein refolding from the slower to the faster pathway. The activity of PPI is decreased by the presence of small concentrations of guanidine hydrochloride (GuHCl) and by a decrease in pH below 7 (Lang, Schmid and Fischer, 1987) as well as by thiol blocking agents (Takahashi, Hayano and Suzuki, 1989). Thus any potential use of PPI in folding *in vitro* would need to take these factors into account. The susceptibility of the enzyme to thiol reagents has been taken as evidence that it exerts its catalytic effect by acting as a nucleophile (Fischer, Bang and Berger, 1989). However, more recent evidence argues against this and suggests that the enzyme mechanism involves distortion of the substrate C–N amide bond (Harrison and Stein, 1990).

THE REFOLDING OF DISULPHIDE BONDED PROTEINS

The major rate-determining step in the *in vitro* refolding of disulphide-bridge-containing proteins is the formation of those bonds. The mechanism and order of their formation has been studied intensively and, in the case of bovine pancreatic trypsin inhibitor (BPTI) in particular, has led to a detailed understanding of the folding pathway (Creighton, 1978, 1984; Kim and Baldwin, 1990; Weissman and Kim, 1991). As well as providing information about the mechanism of protein folding, these studies are also of biotechno-

logical significance, as disulphide-containing proteins may be produced in prokaryotic organisms which lack the appropriate environment to oxidize the cysteines. The recombinant protein in these cases has to be folded and the correct disulphides formed *in vitro* following extraction from the cell. It has been established that the information and thermodynamic drive to bring about the correct folding and disulphide bond formation reside in the primary sequence of a protein (Anfinsen, 1967).

The oxidative folding of BPTI, which contains three disulphide bonds, proceeds through a defined pathway, outlined in *Figure 2*. The key features of this pathway can be summarized thus:

1. Although potentially a very large number of disulphide-bonded intermediates could form, only a limited number of these are observed during folding.
2. Initially, a single species with a disulphide bond between cysteines 30 and 51 is formed, and is present in all subsequent folding steps.
3. Intermediates containing non-native disulphide bonds are formed during folding.
4. The rate-determining step of folding involves a reshuffling reaction in which an intermediate containing two native disulphide bonds is formed from intermediates containing one native and one non-native disulphide bond.

A recent study by Weissman and Kim (1991) shows that only intermediates containing native disulphide bonds populate the folding pathway of BPTI. Although rearrangements of disulphide bonds occur, it is a random process. An NMR study of the folding intermediates showed that those containing native disulphide bonds also possessed some native-like structure, and that non-native disulphide bonds do not contribute to the acquisition of native structure (States *et al.*, 1987). The relation between the formation of disulphide bonds and the stabilization of non-covalent secondary structure intermediates is illustrated by the stabilization by a single disulphide of a folding intermediate in a peptide from BPTI (Kim and Baldwin, 1990).

The approach used to characterize the folding of BPTI has also been attempted in studying the folding of RNase A. Whereas a folding pathway could clearly be demonstrated with BPTI, the results with RNase A were much more complicated and an ordered folding pathway could not be discerned. Two-dimensional gel electrophoresis of proteolytic digests of the intermediates showed that a great many disulphide-bonded species are formed, none of which seems to predominate (Creighton, 1979). With one exception, they are all inactive and the polypeptide chain is unfolded (Galat *et al.*, 1981). However, one intermediate has been characterized which has enzymic activity and a native-like conformation (Creighton, 1980a). It contains three out of the four native disulphide bonds. The missing disulphide bond, which is between cysteines 40 and 95, is located at one extremity of the folded molecule, far from the active site. Clearly, the greatly increased number of possible combinations that a protein containing four disulphide bonds has in comparison to one containing three such bonds (Anfinsen and

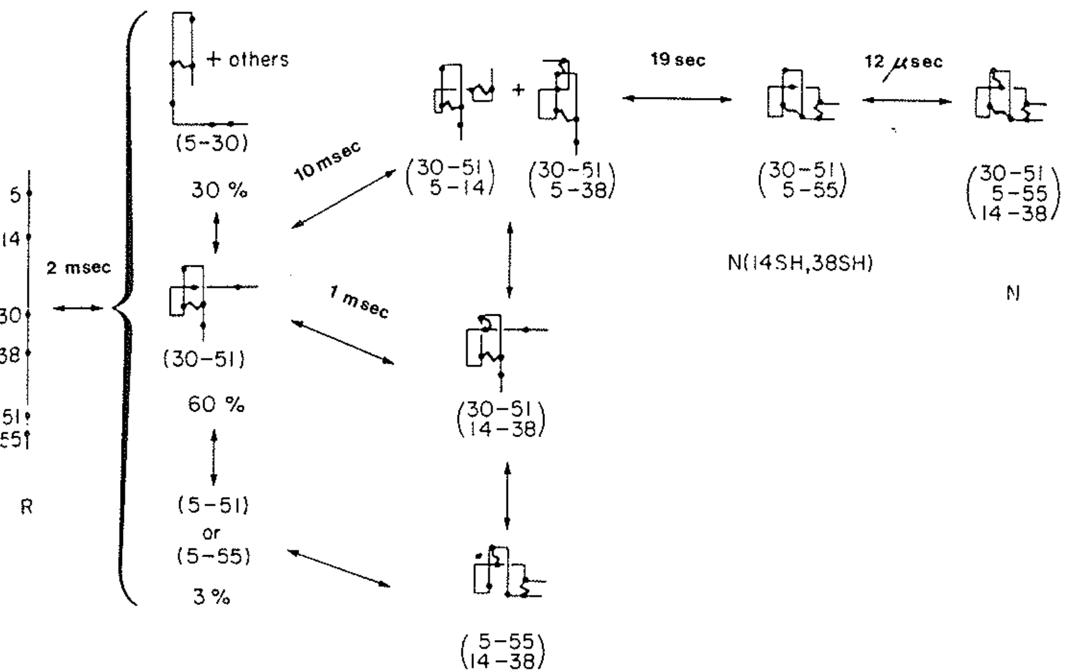
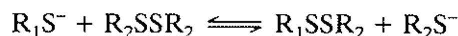


Figure 2. The refolding pathway of a disulphide-bonded protein. Bovine pancreatic trypsin inhibitor (BPTI) contains six cysteine residues, all of which are involved in disulphide bonds in the native state. The polypeptide chain is represented schematically as a solid black line with the positions of the cysteines indicated. Also indicated are the cysteine residues involved in disulphide bond formation in the intermediates, as well as the free thiol groups in native-like conformations (N). The time of formation of an intermediate from the species preceding it in the pathway is shown. From the reduced protein (R) a non-random distribution of one-disulphide-bonded intermediates is formed after an initial rapid equilibration. The relative abundance of each of the one-disulphide-bonded species is indicated. The predominant intermediate (30–51) readily forms three other species containing a second disulphide bond. Two of these intermediates (30–51, 5–14; and 30–51, 5–38) have similar kinetic properties. The next stage in the pathway is the slowest, involving arrangement of disulphide bonds to form a native-like intermediate. (From Creighton and Goldenberg, 1984.)

Scheraga, 1975) may reflect the greater number of intermediates observed experimentally.

Inter-subunit disulphide bonds can be formed *in vitro* as, for example, in immunoglobulins (Kishida, Azuma and Hamaguchi, 1976).

In general, the method adopted to study disulphide bond formation is an exchange reaction of the reduced protein with a thiol reagent:



$$R_1S^- + R_2SSR_2 \rightleftharpoons R_1SSR_2 + R_2S^-$$

This occurs spontaneously at alkaline pH and is readily reversible. In the first step of the reaction, a disulphide bond is formed between the thiol reagent and a protein thiol, to yield a mixed disulphide. Then, a second thiol group of the protein comes into close proximity with the first cysteine residue, displacing the thiol reagent and forming an intramolecular disulphide bond. The intermediates formed in this reaction may then be trapped. Potential thiol reagents have been discussed by Creighton (1978, 1989). Disulphide bond formation can also be promoted by air oxidation and, although this is usually a slow process, it is economical on a large scale and can lead to good yields.

THE ROLE OF PROTEIN DISULPHIDE ISOMERASE

Even under the most favourable conditions, the half-time of reactivation of fully reduced RNase A, which in the native state contains four disulphide bonds, is approximately 20 min (Epstein *et al.*, 1962). The finding that the presence of a microsomal fraction from rat liver could reduce the half-time to less than 5 min (Goldberger, Epstein and Anfinsen, 1963), led to the proposal that disulphide bond formation was catalysed *in vivo*. The enzyme responsible has been purified and characterized as protein disulphide isomerase (PDI). There is a strong correlation between tissues synthesizing disulphide-bonded proteins and PDI activity (Freedman, 1984; Freedman, Brockway and Lambert, 1984) and it is located within the lumen of the endoplasmic reticulum (Freedman *et al.*, 1989). Strong support for an *in vivo* role of PDI in catalysing disulphide bond formation stems from cross-linking studies that have shown PDI to be associated with immunoglobulins (Roth and Pierce, 1987). The strongest evidence, however, was the demonstration that addition of PDI-depleted microsomes to an *in vitro* translation system resulted in defective formation of protein disulphide bonds which could be reversed by reconstitution of the microsomes with PDI (Bulleid and Freedman, 1988).

Three kinds of experiment support the concept that PDI catalyses the *exchange* of disulphides, viz the formation of native disulphide bonds in the presence of a mild oxidant, the isomerization of protein containing 'scrambled' disulphide bonds and the net reduction by PDI of disulphide bonds under appropriate reducing conditions. *In vitro*, the effects of PDI have been tested on the refolding of BPTI. The same disulphide-bonded intermediates of the folding pathway were formed in the same order, with the rate-determining stages of folding being accelerated (Creighton, Hillson and

Freedman, 1980). Chemical modification studies have shown that a thiol group participates in the reactions catalysed by the enzyme (Lambert and Freedman, 1984). The enzyme is dimeric, each subunit consisting of two active-site thiols which act independently of each other (Freedman *et al.*, 1989). The cDNA sequences of four PDI variants (Parkkonen, Kivirikko and Pihlajaniemi, 1988) show a high degree of conservation and internal duplication: two domains within the polypeptide chain show very strong homology to thioredoxin, which participates in thiol-disulphide oxido-reductions. It seems likely that PDI is not only structurally very similar to thioredoxin, but exhibits similar active site chemistry.

In addition, PDI is almost identical in sequence to the β-subunit of prolyl-4-hydroxylase, which catalyses the hydroxylation of proline residues in nascent procollagen polypeptides (Pihlajaniemi *et al*, 1987). Both proteins have PDI activity (Koivu *et al.*, 1987) and are encoded by one single copy in the human genome (Pihlajaniemi *et al.*, 1987). The participation of PDI in other co-translational events has been proposed, including *N*-glycosylation (discussed by Freedman, 1989). However, the mechanism of action of the enzyme and its regulation in these processes is far from clear at present. There is some debate concerning the ability of PDI to catalyse disulphide bond formation in every disulphide-bonded protein. However, it is most probable that the successful application of PDI in renaturation depends upon optimizing the refolding conditions for any given case and not to an intrinsic limitation of the catalytic abilities of the enzyme (Freedman, personal communication).

THE FOLDING AND ASSOCIATION OF MULTISUBUNIT PROTEINS

Studies on the mechanisms of folding have concentrated largely on small, single-domain proteins in order to elucidate the fundamental mechanisms of secondary structure formation and collapse. Many recombinant proteins are also of this nature but already more complex multidomain and multisubunit proteins are being expressed. It has been shown that domains in a 'dumb-bell' type protein, like the immunoglobulin light chain, can fold independently; there is some evidence that, where the domains are closely associated as in penicillin acylase, the N-terminal domain can fold and the next domain flicker into the native-like conformation and associate (Lindsay and Pain, 1990, 1991). In this section we shall consider the assembly of independent subunits to form more complex proteins: most of the detailed work has been carried out on proteins containing identical subunits.

The chief findings from studies of the assembly of multisubunit proteins can be summarized thus:

1. Rapid folding of the monomeric subunits occurs to a structural form capable of apparently specific interaction with other subunits.
2. The folding reaction is not separable from association, and reorganization of the subunits takes place after association has occurred.

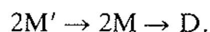
3. The rate-limiting steps in the assembly pathway may vary from protein to protein.

The importance of the specificity of interactions in enabling productive association to occur was demonstrated long ago by the failure to generate hybrid enzymes (Cook and Koshland, 1969). This principle is supported by the accumulated data for the rate constants of association steps during assembly, which are in the range 10^4–10^5 mol $l^{-1}s^{-1}$ (Jaenicke, 1987). This is slower, by a factor of at least 10^3, than those for diffusion-controlled association processes. Structural characterization of the monomeric intermediates is limited, although they are known to possess considerable secondary structure. An equilibrium intermediate of the homodimeric, aspartate aminotransferase from *E. coli* has been studied recently, and shown to possess the properties of the molten globule state (Herold and Kirschner, 1990).

In studying assembly, spectroscopic techniques used in the study of monomeric proteins have been employed and others have been developed. The most important of these has been chemical cross-linking, in which intermediates of assembly are trapped, separated and characterized by SDS-PAGE and quantified by gel scanning. The association pathway has also been followed using limited proteolysis as a means of quenching. Association has been followed after dissociating the oligomer by denaturant, high pressure or low temperature. The presence of ligands during association can have either a stabilizing or a destabilizing effect: there is no general rule. Detailed discussion of the various methods has been provided by Jaenicke and Rudolph (1986, 1989).

The study of association *in vitro* is complicated by competition with aggregation which is favoured at higher protein concentrations. In many cases, recovery is much less than 100%. *Figure 3* shows that at a polypeptide-chain concentration of 1 μM, regain of activity of porcine muscle lactate dehydrogenase (LDH) is negligible, whereas at a concentration of 0.01 μM, it is close to 100%. The practical implications of this finding are clear: that in a production scheme for multisubunit proteins, the working limits of the polypeptide-chain concentration range would need to be established. In addition, techniques of renaturation need to be devised to overcome the problems raised by the low protein concentrations generally found to be necessary.

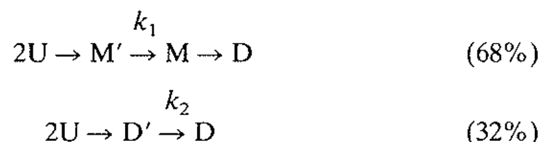
The assembly of multisubunit proteins has been comprehensively reviewed (Jaenicke, 1987). Here, only some of the findings from these studies will be discussed, beginning with dimeric proteins and progressing to structures of increasing complexity. With most dimers, the folding and association of monomers is found to obey a sequential uni-bimolecular mechanism, with rate-determining folding and association steps:



$$2M' \rightarrow 2M \rightarrow D.$$

In the case of cytoplasmic malate dehydrogenase, fast folding into struc-

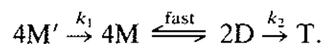
tured monomeric intermediates occurs. The formation of the native enzyme is governed by two parallel steps, one much slower than the other (Rudolph, Fuchs and Jaenicke, 1986). The recovery of secondary structure is a multistep process. Sixty-four per cent of the secondary structure, as determined by far UV CD measurements, is regained in the initial fast phase, the remainder in two further reactions involving both dimer formation and reshuffling within the dimer. Cross-linking revealed the presence of a non-native dimeric intermediate. The formation of the native protein from this intermediate was responsible for the very slow phase. To account for the observed results, these workers put forward the following mechanism in which the native state is formed from two different pathways:



$$2U \rightarrow M' \xrightarrow{k_1} M \rightarrow D \qquad (68\%)$$

$$2U \rightarrow D' \xrightarrow{k_2} D \qquad (32\%)$$

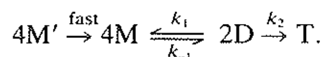
The most exhaustively studied tetramer has been porcine muscle LDH. The kinetic mechanism shows a very rapid monomer–dimer equilibrium, which is close to the rate limit imposed by diffusion (Hermann, Jaenicke and Rudolph, 1981). Dimer formation is determined by folding of the monomers, which is a first-order process (Zettlmeissl, Rudolph and Jaenicke, 1982):



$$4M' \xrightarrow{k_1} 4M \xrightleftharpoons{\text{fast}} 2D \xrightarrow{k_2} T.$$

Association of LDH was quenched by partial proteolysis (Girg, Rudolph and Jaenicke, 1981), where it was arrested at the dimer stage from which the tetramer could not be reconstituted. An amino-terminal decapeptide, cleaved by the proteolysis treatment, is essential for tetramer formation. Though the dimer is inactive, it can bind NAD^+ (Rudolph, 1985), so it must possess native structure. However, activity could be induced in the dimer by the presence of sulphate (Girg, Jaenicke and Rudolph, 1983). The kinetics of formation of this 'trapped' dimeric intermediate were the same as those of dimer formation in the native tetramer (Opitz *et al.*, 1987).

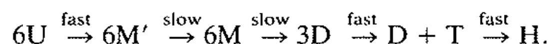
The association pathway of phosphoglycerate mutase from yeast also shows a monomer–dimer equilibrium (Hermann *et al.*, 1983). The monomeric and dimeric intermediates also exhibit partial activity (Hermann, Jaenicke and Price, 1985). The association pathway of this enzyme can be summarized thus:



$$4M' \xrightarrow{\text{fast}} 4M \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2D \xrightarrow{k_2} T.$$

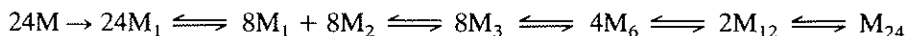
The assembly of homohexameric bovine liver uridine diphosphoglucose dehydrogenase is characterized by sequential uni-bimolecular kinetics, with the rate-determining steps of the reaction occurring at the monomer level. The hexamer is generated by dimer associating with tetramer in the following

scheme (Jaenicke, Rudolph and Feingold, 1986):



$$6U \overset{\text{fast}}{\longrightarrow} 6M' \overset{\text{slow}}{\longrightarrow} 6M \overset{\text{slow}}{\longrightarrow} 3D \overset{\text{fast}}{\longrightarrow} D + T \overset{\text{fast}}{\longrightarrow} H.$$

The association pathway of the 24-subunit iron storage protein, ferritin, was shown to proceed via formation of a trimer (Gerl and Jaenicke, 1988), as follows:



$$24M \rightarrow 24M_1 \rightleftharpoons 8M_1 + 8M_2 \rightleftharpoons 8M_3 \rightleftharpoons 4M_6 \rightleftharpoons 2M_{12} \rightleftharpoons M_{24}$$

A practical consideration is the fact that specific association is always in competition with aggregation that is usually irreversible. This is illustrated in *Figure 3*, where the concentrations required for examination of the process of assembly are seen to be low and impractically low when the aim is to produce significant quantities of protein from the unfolded state. This competition can occur even *in vivo*, as shown for the wild-type tailspike protein (King, 1986), but is nevertheless largely avoided by the concentration of free monomer being kept very low. This may be brought about by rapid association immediately following biosynthesis and folding to a collapsed state, and also by the involvement of chaperone proteins (*see* below). In an *in vitro* assembly process, therefore, processes have to be developed to ensure that the concentration of monomer at any time is kept low.

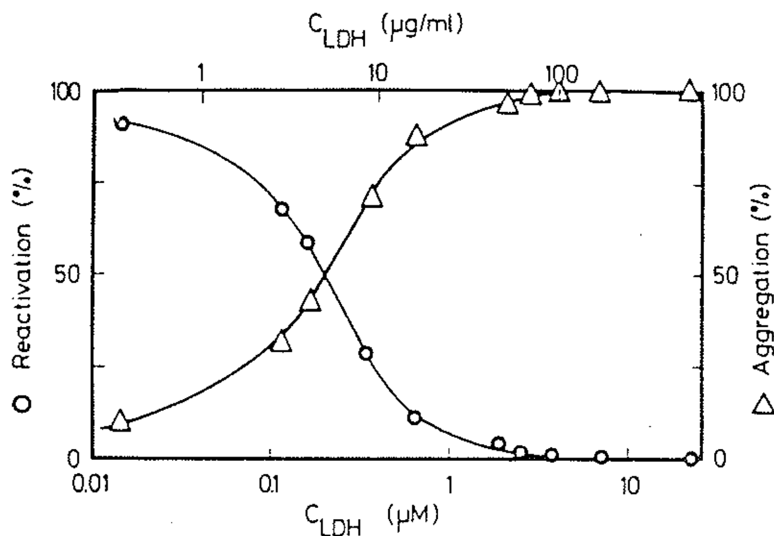


Figure 3. The effect of protein concentration upon the *in vitro* reactivation of a multisubunit protein. Reactivation (O) of acid-denatured lactate dehydrogenase competes with the formation of aggregates (△), in which the protein is inactive. At low protein concentrations, the pathway leading to aggregation is not a major one, but it becomes more favoured as the concentration increases. (From Zettmeissl, Rudolph and Jaenicke, 1979.)

CHAPERONINS AND THE ART OF AVOIDING STICKY SITUATIONS

Currently, there is much interest in the role of a family of highly conserved proteins known as chaperonins. They are involved in a wide range of cellular processes, such as translocation of polypeptide chains across membranes and the assembly of multisubunit proteins. These functions require that the chaperone protein interacts with the polypeptide chain before it attains its native confirmation. What is lacking at the moment is a precise knowledge of the role of chaperonins in the folding process *in vivo*. Recent experimental evidence shows that hsp70 strongly interacts with the polypeptide chain as it is being synthesized (Beckmann, Mizzen and Welch, 1990). They are commonly referred to as catalysts of folding. Whether they are indeed true catalysts, which accelerate the rate of folding, is not clear. Hsp70 has been proposed to act as an 'unfoldase', using the energy of ATP hydrolysis to disrupt incorrectly formed interactions within the polypeptide chain (Pelham, 1986). However, the emerging view of the mechanism of action of other chaperonins would appear to contradict this hypothesis (Goloubinoff *et al.*, 1989; Buchner *et al.*, 1991; Mendoza, Lorimer and Horowitz, 1991).

Recent reviews include those by Georgopoulos and Ang (1990), Gething and Sambrook (1990), Hemmingsen (1990), Horwich, Neupert and Hartl (1990), and Wiech, Stuart and Zimmerman (1990). Here, we will briefly summarize findings directly relevant to protein folding and assembly, and consider their potential impact in biotechnology.

In the cell, chaperonins perform their roles in a co-ordinated fashion with various proteins, including other chaperonins. ATP is frequently involved. They have a broad substrate specificity and some of them have a greater binding affinity for polypeptides than others (Kusters *et al.*, 1989; Lecker *et al.*, 1989).

Current evidence suggests that chaperonins interact with an intermediate conformational form of the polypeptide chain, and that this is the only form of the polypeptide recognized by them. The role of chaperonins has been demonstrated in the following cellular processes: the transport of the polypeptide to a site of translocation; translocation itself; enabling the polypeptide to interact with a ligand; and assembly of multisubunit proteins. It is now becoming apparent that for all of these types of interaction to occur, the polypeptide chain must be maintained in an appropriate conformational state that is less completely folded than the native state. For example, the binding of pre-β-lactamase and of chloramphenicol acetyl transferase to chaperone GroEL was chased out using heat-denatured, but not native, myoglobin (Bochkareva, Lissin and Girshovich, 1988); the polypeptide chains bound to yeast hsp60 (Ostermann *et al.*, 1989), hsp70 (Deshaies *et al.*, 1988) and SecB (Weiss, Ray and Bassford, 1988) are protease sensitive. The clearest indication about the state of the bound protein, however, comes from the finding that the acid and GuHCl denatured forms of a dimeric ribulose biphosphate carboxylase (Rubisco), which differ in that the acid-denatured state still contains some secondary structure, fold through a common intermediate, Rubisco-I, that binds to GroEL (Goloubinoff *et al.*, 1989). Also, the presence

of tertiary structure was demonstrated in ProOmpA bound to SecB by fluorescence energy transfer (Lecker, Driessen and Wickner, 1990). By studying the fluorescence energy properties of a derivatized form of ProOmpA, these workers showed that SecB binds a non-native conformational state of the polypeptide chain, which is clearly not the fully unfolded form of the protein seen upon GuHCl denaturation.

Native proteins do not interact with chaperonins. Once dihydrofolate reductase acquired protease resistance, it did not bind to hsp60 (Ostermann *et al.*, 1989) and glucocorticoid receptor which binds steroid does not interact with hsp90 (Dalman *et al.*, 1989). The fact that it is a non-native 'sticky', mobile, intermediate state that binds is consistent with the fact that most chaperone proteins are relatively non-specific with respect to the proteins that can be stabilized against aggregation.

At present, the prospects for using chaperonins in the commercial production of proteins are extremely hopeful. Current work indicates that, in principle at least, there are two major ways in which chaperonins can be used in the production of commercially useful proteins. A chaperonin gene can be introduced into a host cell and co-expressed with the gene of interest (Goloubinoff, Gatenby and Lorimer, 1989). Alternatively, since many purification strategies involve solubilizing the protein from inclusion bodies in denaturant, chaperonins could be constituents of the renaturation medium (Goloubinoff *et al.*, 1989). In the second approach, they could be employed either to assemble proteins which would not otherwise fold or assemble under *in vitro* conditions, or to increase the yield of refolding and assembly.

The broad substrate specificity of most chaperonins, as well as the fact that more than one chaperonin will interact with any given protein, conceivably allows scope for optimization of a refolding strategy for a particular protein by selection of the best chaperone for a particular job. The fact that they are induced under extreme physiological conditions can possibly be exploited to oppose potentially restrictive factors, such as temperature for example, in the folding and production of active protein.

SIGNAL SEQUENCES AND THEIR INFLUENCE UPON FOLDING

The role of leader sequences as signals in targeting and secretion of proteins is well recognized. What is not fully understood at present is their structural relationship to what becomes the mature protein. The aim here is to summarize briefly what is known about their structure and their relevance to protein folding.

Signal sequences (Gierasch, 1989; Randall and Hardy, 1989) are typically 15–30 amino acids in length, very hydrophobic and lack primary sequence homology. The most highly conserved site is at the point of proteolytic cleavage, where there is a preference for amino acid residues which have relatively small side-chains. Although the other regions of the peptide lack homology, they show common distributions of residue type. There are three regions with specific characteristics within and adjacent to the signal peptide. Five to seven residues from the cleavage site, on the C-terminal side, the

sequence is generally more polar than a hydrophobic region which is immediately N-terminal to it. Typically, the latter consists of 10±3 residues. The N-terminal region contains a net positive charge.

The importance of hydrophobicity of the signal sequence was demonstrated by the finding that 20% of randomly generated peptide sequences could function as an export peptide for invertase from *Saccharomyces cerevisiae* (Kaiser *et al.*, 1987). Though the primary sequence homology between the effective peptides was low, all of them were hydrophobic.

Despite this lack of sequence homology, isolated signal peptides have similar conformational preferences: they are random in aqueous solution and adopt an α-helical structure with non-polar solvents or micelles (Gierasch, 1989).

Several lines of evidence have been advanced to support the suggestion that the signal sequence does not interact with the mature protein. Baty and Lazdunski (1979) purified an antibody raised against precursor alkaline phosphatase but which did not cross-react with the mature protein. As with the mature form of the protein, the precursor of maltose binding protein (MBP) is resistant to a variety of proteases; it also binds a non-ionic detergent, whereas the mature protein does not (Dierstein and Wickner, 1985). In addition, it can bind substrate (Ferenci and Randall, 1979). The signal peptide has, however, been shown to interact with the protein during folding (Park *et al.*, 1988). The kinetics of refolding of both the precursor and mature forms of guanidine-denatured MBP and ribose binding protein (RBP) were compared. By extrapolation of the rates of refolding to solution conditions of zero GuHCl, the refolding rate of MBP was decreased in the precursor by a factor of 40, with the relaxation time increasing from 0·7 to 28 s. The presence of the leader also decreased the rate of folding of RBP. If the signal is attached to the membrane while the protein folds, however, it is difficult to predict the effect on the folding rate.

PROTEINS ACTIVATED BY PROTEOLYSIS

A range of important proteins, including several proteolytic enzymes and protein hormones, are rendered functional by proteolysis with the loss of one or more peptides. In certain cases, such as penicillin acylase (Sizmann, Keilmann and Böck, 1990), the activation process accompanies translocation. The obvious route to producing such proteins would be to clone and express the gene for the precursor, followed by *in vitro* proteolysis. In many cases, however, the degree of specificity of cleavage obtained in the cellular environment cannot be reproduced *in vitro*. This has led to the requirement to express DNA coding for the component chains of the activated protein—usually as insoluble inclusion bodies to avoid proteolytic degradation—followed by folding and assembly to the functional protein.

One of the significant experiments originally carried out in Anfinsen's laboratory was the addition of PDI to insulin and to chymotrypsin. Both these proteins, in marked contrast to ribonuclease, became 'scrambled' with respect to their disulphide bonds, lost activity and usually aggregated. The

information in the amino acid sequence of proinsulin and of chymotrypsinogen that led to the correct folding of these precursors was apparently lost in the cleaved and functional forms (Anfinsen, 1967).

The ability to reconstitute insulin from isolated A and B chains at yields up to 60% (Frank and Chance, 1983) has led to a reconsideration of this explanation. The assembly of active staphylococcal nuclease from complementary and overlapping peptides (Anfinsen, 1972) confirmed that information for folding resides in smaller units than the complete amino acid sequence. The view presented here is that smaller units of a protein, although intrinsically capable of folding, attain a conformation that is both less stable and more 'sticky' due to the exposure of sites that would normally be satisfied by interaction with neighbouring folding units. Thus the usual competition with an aggregation reaction (Pain, 1977)

$$N \rightleftharpoons D \rightarrow X$$

leads to low yields of correctly folded and assembled protein.

Penicillin acylase consists of two chains, α and β, of M_r 23 800 and 62 200, respectively, from proteolytic cleavage of a precursor, of M_r 92 000. The isolated and unfolded β-chain aggregates strongly on refolding, even in relatively high concentrations of urea, while the originally N-terminal α-chain refolds in high yield to a compact globular state that closely resembles its conformation in the active enzyme (Lindsay and Pain, 1990). Active enzyme can be regenerated in up to 60% yield. The yield can be increased by adding unfolded β-chain to folded α-chain, with higher activities being obtained by increasing the molar ratio of α to β-chains (*Figure 4*). This leads to the conclusion that the β-chain is capable of folding to a native-like conformation that can be stabilized by interaction with a folded α-chain. It is also concluded that the folded β-chain must meet an α-chain before meeting another β-chain in order to assemble correctly and avoid aggregation (Lindsay and Pain, 1991).

These experiments, together with those on the high-yield refolding and assembly of cross-linked insulin (Tang and Tsou, 1990), lead to the proposal that the intervening peptide in the precursor acts primarily to increase the probability of the two chains meeting and thus avoiding aggregation. Insertion of residues into the intervening peptide of penicillin acylase *in vivo* slowed but did not prevent processing (Sizmann, Keilmann and Böck, 1990), again supporting a primarily non-conformational role for the intervening peptide.

A second type of activation process involves excision of an N-terminal peptide, ranging from 44 residues for cathepsin D to 290 residues for activin A. For cathepsin D (Pain, Lah and Turk, 1985), α-lytic protease (Silen and Agard, 1989), subtilisin (Zhu *et al.*, 1989), activin A and transforming growth factor (TGF-β1) (Gray and Mason, 1990), however, the presence of the pro-peptides is an absolute requirement for folding. It is particularly interesting that the larger pro-peptides, when co-expressed with but not covalently linked to the mature protein, still enable the protein to fold. These pro-

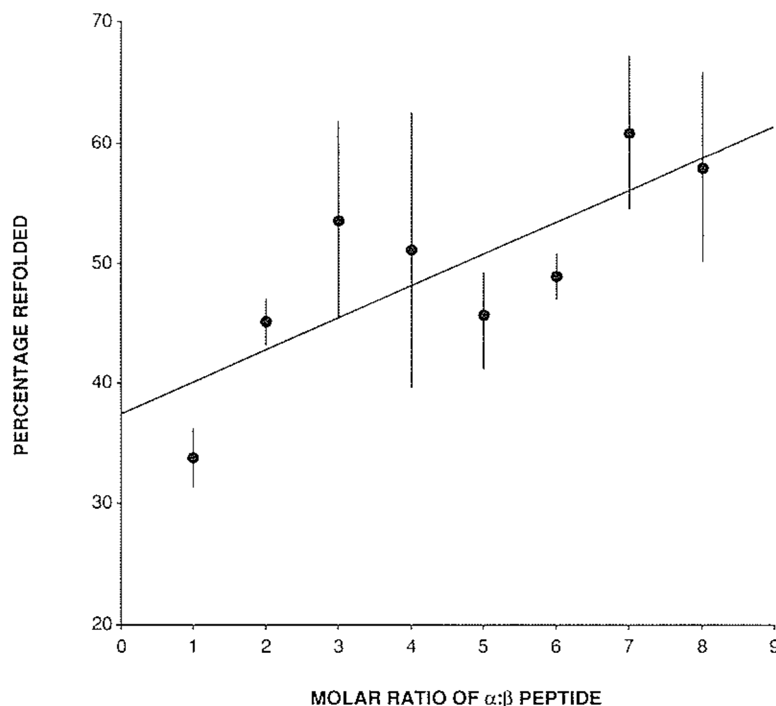


Figure 4. Recovery of active penicillin acylase depends on the concentration of α-peptide. Isolated β-peptide unfolded in 8 M urea was diluted into varying concentrations of previously refolded α-peptide in borate buffer, pH 6·5 at 5·7°C. The percentage recovery of PA activity based on the constant concentration of β-peptide ($6{\cdot}4 \times 10^{-8}$ mol l^{-1}) is plotted against the mole ratio of α-peptide to β-peptide. Vertical bars indicate standard deviations for the three determinations. It should be noted that addition of unfolded α-peptide to β-peptide in folding conditions (i.e. absence of urea) results in zero yield of PA activity (Lindsay and Pain, 1991).

peptides therefore may act like disposable but genetically economical chaperonins that protect the intermediate states from aggregation during folding but are not important for the function or stability of the folded protein (Silen and Agard, 1989).

Unresolved questions still remain with mature human renin which has not been renatured from inclusion bodies in *E. coli*, necessitating the expression of pro-renin (Sharma *et al.*, 1987). In mammalian cells, however, both pro-renin and mature renin can be expressed in folded forms (Harrison *et al.*, 1989).

A COMPARISON OF PROTEIN FOLDING *IN VITRO* WITH FOLDING *IN VIVO*

In vitro studies of protein folding start with the completed polypeptide chain already in solution, in which the refolding conditions are applied along the length of the chain simultaneously. *In vivo*, however, the possibility exists that folding is to some extent co-translational. The rate of chain elongation in rabbit haemoglobin was estimated to be approximately two amino acids per second (Dintzis, 1961). This is not inconsistent with experimentally observed

rates of refolding *in vitro*, which range from seconds to decades of minutes. Also, as has already been discussed, secondary structure formation is very rapid *in vitro* and does not require a great length of the polypeptide chain. One possible limitation to the degree of co-translational folding is the finding that during translation not all the amino acids added to the growing chain are immediately available for folding: 30–35 of the most recently added residues are estimated to be shielded by the ribosome (Malkin and Rich, 1967).

The question that must be addressed is, to what extent and in what form does protein folding occur during translation?

There are more complex ways in which folding *in vivo* may differ from that *in vitro*. It is not separable from other cellular processing events such as proteolytic activation, translocation and glycosylation. The assembly of multimeric proteins requires both spatial and temporal separation, for example during the folding and association of immunoglobulin chains (Bergman and Kuehl, 1979). Enzymes and accessory proteins may also facilitate folding, as described earlier.

Despite these differences, it has always been assumed that the experimentally verifiable *principles* of folding *in vitro* will apply *in vivo*. For example, a partially synthesized polypeptide still attached to the ribosome will, to the extent that it is free from the ribosome surface, be able to flicker into secondary structure. Further, a significant number of non-polar side-chains in the nascent chain will be expected to extend a hydrophobicity-based drive to a globule-like state. A reasonable proposal thus emerges that the globule, with increasingly stabilized secondary structure, could grow like an oil drop until finally released, to undergo any final shuffling to the native state. In multidomain polypeptides, completed domains could fold while following domains are still being synthesized.

The major question to be addressed is therefore to what extent findings from the study of protein folding *in vitro* are consistent with what is known about folding *in vivo*. Though the information obtained is relatively limited, *in vivo* studies substantiate *in vitro* findings mentioned earlier. The importance of secondary structure formation early on in the folding pathway has been demonstrated, as well as the independent folding of domains and also the occurrence of aggregation.

Studies on the tailspike protein (TSP) of phage λ-22 of *Salmonella typhimurium* have been among the most revealing in this respect (King, 1986; King, Haase and Yu, 1987). The folding of temperature-sensitive mutant forms of the protein proceeds at permissive temperatures to produce functional proteins at least as stable as the wild type. Raising the temperature, however, favoured aggregation rather than continued folding and specific association, and little or no active protein was obtained. Analysis of the mutations revealed that in some cases glycines were substituted by bulky, charged residues. Glycines commonly occur in β-turns, which is particularly significant in view of the mainly β-sheet structure of TSP. In other temperature-sensitive mutants, threonine was replaced by amino acids lacking a hydroxyl group. Threonines are commonly involved in internal hydrogen bonding, the loss of which may destabilize folding intermediates.

The independent folding of domains *in vivo* has been demonstrated, for example, by creating inter-species hybrids of yeast and human PGK (Mas *et al.*, 1986). These chimeras had catalytic properties very similar to those of the parent proteins, despite a 35% difference in amino acid sequence. Intermediates of assembly have been detected with immunoglobulins (Bergman and Kuehl, 1979). In one mouse myeloma cell line a completed light chain was covalently associated with a nascent heavy chain, whereas in another cell line association proceeded through the formation of a heavy chain dimer. The minimum size of nascent heavy chain detected as an association intermediate was 38 kDa, enough to allow synthesis and folding of the V_H and C_H1 domains that, with light chain, constitute the disulphide-bonded F_{ab} fragment. That aggregation may occur during folding *in vivo* as well as *in vitro* has been shown with wild type as well as mutant forms of TSP (Haase-Pettingell and King, 1988). A mutant form of α_1-antitrypsin known as the Z-form, forms intracellular inclusions, with only 15% of the protein being secreted in an active form (Carrell, 1986). The remainder is blocked at the final processing stage in the secretory pathway. In this mutant, a conserved glutamate which forms an important salt bridge is lost. Urea gradient gel electrophoresis has shown that the Z-protein, like the normal, wild-type protein, folds rapidly to the compact intermediate, which then folds to the native state much more slowly than the wild-type inhibitor (Brind *et al.*, submitted for publication).

No significant inconsistencies have yet emerged between what is known about *in vitro* and *in vivo* folding, but more work is required before we can be certain about the details of the latter process.

Protein folding in biotechnology

THE PRODUCTION OF RECOMBINANT PROTEINS

The genetics of *E. coli* expression have lent themselves well to the exploitation of this organism for the production of foreign proteins (reviewed by Marston, 1986). However, advances in the technology of genetic manipulation and understanding of the processes (e.g. Reznikoff and Gold, 1986; Kingsman and Kingsman, 1988) have allowed molecular biologists to capitalize on the secretion pathways of yeast, mammalian cells, fungi and other bacteria for the production of recombinant protein (Harris, 1990). Discussion of the merits of these different systems lies outside the scope of this review.

Two routes of protein production are available in *E. coli*. The use of various leader sequences can direct secretion of soluble protein which, in the periplasmic space, is correctly folded and contains the correct disulphide pairings. Examples of the successful use of this method are the alkaline phosphatase signal sequence to secrete BPTI (Bermann-Marks *et al.*, 1986) and the OmpA leader to produce human growth hormone at 10–15 μg hGH/A_{600} *E. coli* cells in the periplasmic space (Hsiung, Mayne and Becker, 1986). The second route is the intracellular high expression of foreign protein, often leading to the formation of refractile inclusion bodies within the cell

(Marston, 1986). These inclusion bodies or granules are proteinaceous and are visible by dark-field microscopy (Schoner, Ellis and Schoner, 1985), providing an attractive route for the concentration of large amounts of protein.

Although the secretion route produces folded, active protein, the levels of expression tend to be lower than those obtained using the alternative intracellular route. Also, with the former, purification from the other periplasmic proteins is required. The inclusion body route provides very high protein expression levels, in some cases greater than 25% of the total cytoplasmic protein (Hartley and Kane, 1988), in a discrete body that is a distinct aid to further purification (Marston, 1986 and references within) and which provides protection from proteolytic degradation of foreign (Kitano *et al.*, 1987) and labile (Cheng *et al.*, 1981) protein by the host. However, these inclusions require solubilization, refolding and, where appropriate, re-oxidation to the native state.

THE NATURE OF INCLUSION BODIES

The physical process of inclusion-body formation is a subject arousing much interest (Schein and Noteborn, 1988; Mitraki and King, 1989; Schein, 1989). An understanding of the mechanism is desirable from a commercial aspect as it would allow the optimization of production and recovery. The presence of plasmids, ribosomal RNA and RNA polymerase within inclusion bodies indicates that they form immediately after translation of the protein (Hartley and Kane, 1988). It has been demonstrated that in the prochymosin inclusion body some of the aggregated protein contains intermolecular disulphide bonds, but that the majority of it is in a reduced form (Schoemaker, Brasnett and Marston, 1985). The reducing environment of the *E. coli* cytoplasm (Fahey, Hunt and Windham, 1977) will obviously inhibit disulphide formation, though recent findings argue that BPTI can oxidize in the *E. coli* cytoplasm (Nilsson *et al.*, 1989). It is probable that the formation of these covalent disulphide linkages in inclusion bodies occurs on exposure to air during the cell rupturing phase at the start of isolation (Marston, 1986; Mitraki and King, 1989). However, the presence of reducing microenvironments within the dense aggregate has not been ruled out (Schoemaker, Brasnett and Marston, 1985). Certainly, the periplasmic space provides the oxidizing environment for disulphide formation in secreted proteins (Hsiung, Mayne and Becker, 1986).

Attempts have been made to monitor the formation of the inclusion bodies using single-cell light-scattering in order to quantitate accumulation of product (Wittrup *et al.*, 1988). Schoner, Ellis and Schoner (1985) observed that small granules form first only at one pole of the cell. Subsequently, these granules enlarge and this is followed by the appearance of granules at the opposite pole of the cell and then finally between the poles. The polar and sub-polar position of granules has been noted for other proteins (Kitano *et al.*, 1987).

No obvious mechanism was immediately discernible for the aggregation of

proteins into inclusion bodies. Kane and Hartley (1988) have examined data on a number of protein inclusions in *E. coli* K12 and found no correlation between host strain or protein hydropathy and granule formation. Similarly, the rate of production of protein and cellular concentration of protein show no direct correlation with the appearance of intracellular protein aggregates (Schein and Noteborn, 1988), though they do show a temperature effect on formation. The protein A-β-galactosidase fusion shows the yield of inclusion bodies to be dependent on growth conditions (Stransberg, Veide and Enfors, 1987). More recently the tight binding of an acidic component from *E. coli* to the expressed heterologous protein has been implicated as a factor in promoting protein aggregation to an inclusion body (Darby and Creighton, 1990).

A very basic model was postulated for refractile-body formation, based on protein conformation (Hartley and Kane, 1988). As more evidence becomes available, it is increasingly held that inclusion bodies form through 'specific aggregation' of protein folding intermediates, for example in tailspike protein (King, 1986; Haase-Pettingell and King, 1988) and bGH (Brems *et al.*, 1986, 1987; Havel *et al.*, 1986; Brems, 1988). The work reviewed by Mitraki and King (1989) indicates that exposed hydrophobic surfaces in intermediate states play a crucial role in protein association, as illustrated for bGH (cited above) and PGK (Mitraki *et al.*, 1987). An understanding of the factors influencing formation and stability of intermediate states during protein folding (discussed in preceding sections) is thus crucial to controlling, and therefore optimizing, the formation of inclusion bodies. It is interesting to note, however, that pro-subtilisin, which aggregates in the periplasmic space with a certain expression vector, is produced with a sixteenfold increase in concentration of soluble pro-protein when lower inducer concentration and growth temperature are used (Takagi *et al.*, 1988).

The requirement to recover active protein from insoluble aggregates resulted in the 'art' of protein folding becoming critical for the initial development of the biotechnology industry (Goldberg, 1985; King, 1986; Pain, 1987). The formation of insoluble aggregates in bacteria is beginning to be seen as a consequence of the mechanism by which proteins fold *in vitro*, with solution conditions, i.e. pH, ionic strength and temperature, affecting the stability of the intermediates, particularly the kinetic 'molten globule' that accumulates during folding (*see* p. 50).

SOLUBILIZATION OF PROTEINS FROM INCLUSION BODIES

Isolation of the inclusion body

The first step in recovery of protein from an inclusion body is to release it from the cell. Cells harvested by centrifugation are resuspended in a suitable buffer and then lysed. Typically, cell lysis is achieved using a French press, by sonication or by enzymes such as lysozyme. Initial centrifugation at relatively low speed is then used to sediment the inclusions from the cellular debris

which remains in the supernatant. Speeds in the range 500–5000 *g* are quoted in a broad patent (Jones, Olson and Shire, 1985), though speeds in the range of 5000–12 000 *g* are used by other workers (Marston, 1986). The use of 'kill' treatment on cells to enhance the amount of precipitated protein before lysis has also been used (Jones, Olson and Shire, 1985). Quoted 'killing' procedures are acid or heat treatment or, alternatively, exposure to a non-polar organic solvent. A number of published protocols kill the cells prior to recovery of inclusion bodies (Furman *et al.*, 1987; Weir and Sparks, 1987).

Solubilization of the inclusion body

Washing the bGH inclusion body (Schoner, Ellis and Schoner, 1985) with 0·5 M, 1 M and 2 M urea has been shown to remove large amounts of contaminating *E. coli* protein from the spun pellet without significantly solubilizing the granule. Increasing the concentration of urea to 5 M with Triton X-100 included does not significantly increase purity within the granule. Further purification can only be achieved after complete disassociation of the aggregate. Langley *et al.* (1987a) washed bGH granules with a mixture of EDTA, lysozyme and deoxycholate to extract lipid, nucleic acid, lipopolysaccharide and protein contaminants. Triton X-100 washes have been used successfully to remove contaminants from the prochymosin inclusion body (Marston *et al.*, 1984). The ability to attain up to 90% protein purity in two simple steps following cell lysis makes the intracellular granule expression route an extremely useful aid to purification.

Marston (1986) quotes a range of published conditions and denaturants for solubilizing various protein inclusions. These methods include the use of alkaline pH, detergents and organic solvents, as well as urea and GuHCl. A patent series from Genentech covers solubilization, purification and recovery of heterologous protein (a heterologous protein is defined as one not normally produced in the host, or which is normally produced only in small quantity) from refractile bodies using strong denaturants, buffer solutions and molecular sieving or centrifugation (Olson, 1985; Builder and Ogez, 1985; Olson and Pai, 1985). A further patent application (Jones, Olson and Shire, 1985) for a process for 'Purification and activity assurance of precipitated heterologous proteins' states the use of either various strong denaturants (urea is classed as weak denaturant), with guanidine salts or sodium thiocyanate at 6–9 M concentration being preferred, or detergents (Triton, SDS) in the 0·01–2·0% range. The next aspect of the patent is covered later in the two-step renaturation procedure, exchanging strong for weak denaturant. Some of the protein samples quoted are pro-renin, capsid protein of foot-and-mouth disease (FMD) virus and porcine growth hormone. In the case of the fusion VP1 protein of FMD, 8 M urea over a range of pH values failed to fully solubilize inclusions (Shire *et al.*, 1984). A number of proteins have been expressed using the CIIFX fusion protein (a factor X cleavable fusion) and the inclusion bodies formed are only soluble in solutions such as 8 M urea or 6 M GuHCl (Nagai, Thorgerson and Luisi, 1988).

Thiol reagents do not usually need to be present for solubilization of

inclusions but, as will be discussed later, cysteine residues are an important consideration during refolding. Interleukin-4 does not appear to be released from granules by 1% SDS or urea; but 5 M GuHCl, pH 8·0 with 2 mM reduced and 0·2 mM oxidized glutathione, produces soluble protein which can be refolded (Kimmenade *et al.*, 1988). Platelet-derived growth factor-β (PDGF-β) is produced in granules in *E. coli* (Hoppe, Weich and Eichner, 1989) which are dissolved in 50 mM TRIS pH 7·8, 2% SDS, 2% mercapto-ethanol at 37°C, followed by the addition of acetone to reprecipitate the protein. After cleavage of the fusion protein with cyanogen bromide, the protein is resuspended in 6 M GuHCl and sulphitolysed to protect thiol groups.

A novel procedure for solubilizing inclusion bodies involves the use of an ion-exchange resin (Hoess *et al.*, 1988). Addition of ion-exchange resin to bacterial lysates containing inclusions of SV40 Th peptide and IL-2 show recovery of soluble, active protein from the granule. The resins S- and Q-Sepharose have been used, and stable binding to the resin is not a prerequisite for solubilization. The ability of a charged resin to remove protein from an aggregate, and for the solubilized protein to fold to an active conformation, raises some interesting questions about the forces holding inclusion bodies together. Soluble renin has been obtained from inclusions using repeated French press passages in the absence of denaturant (Sharma, 1986). This method does not completely solubilize the pellet and presumably removes only weakly associated protein.

Two further procedures have recently been patented for the recovery of active protein from inactive forms. In one patent, inclusion bodies are dissolved with the aid of a surfactant, then exchanged with an aqueous solvent to allow refolding, with contact to a second solution containing weak denaturant to prevent precipitation (Patroni, 1990). The second method uses a molecular sieve to convert solubilized denatured protein into a biologically active form (Hermann, 1990).

REFOLDING PROTEINS FROM INCLUSION BODIES

Once solubilized, the recombinant protein may be further purified in its unfolded (reduced if necessary) form to prevent any foreign proteins interfering in the refolding process. In the case of PDGF-β the protein is chromatographed in GuHCl over Sephacryl S-200, then reverse-phase HPLC prior to reoxidation (Hoppe, Weich and Eichner, 1989). Insulin-like growth factor II is purified after sulphitolysis using mono-Q FPLC (assisted by the negative charges from the sulphitolysed cysteines) prior to renaturation (Furman *et al.*, 1987). In the case of recombinant human IFN, a mixture of ion-exchange, gel filtration and reverse phase chromatography is used (Arakawa, 1986). The wealth of chromatographic media available allows a great deal of choice, depending on individual protein requirements. Genetically engineered fusions of affinity (Hopp *et al.*, 1988) and charged tails (Sassenfeld and Brewer, 1984) have been used successfully to improve purification of recombinant proteins in different expression systems.

The basic methods for refolding a protein have recently been set out (Jaenicke and Rudolph, 1989). Here, the techniques and parameter variables involved in refolding a recombinant protein from an inclusion body will be illustrated using some specific examples. An obvious way to refold proteins is to reduce the denaturant concentration by dilution with buffer, diafiltration or by dialysis. The drawback here is that precipitation of protein can occur (Jones, Olson and Shire, 1985) resulting in low yields. Aggregation during refolding can be exacerbated at high protein concentrations (London, Skrzynia and Goldberg, 1974) and critical concentrations are therefore often quoted (Marston, 1986; Furman *et al.*, 1987; Kimmenade *et al.*, 1988). In the case of patent applications for the production of insulin by combining recombinant A and B chains, broad concentration ranges are claimed but a specific range preferred (Chance and Hoffman, 1981a,b).

USE OF DENATURANT EXCHANGE

The use of intermediate denaturing conditions during refolding suggested a dependence of yield on denaturant concentration (Chan *et al.*, 1973). A detailed study of the folding of tryptophan synthase (London, Skrzynia and Goldberg, 1974) demonstrated that on reduction first to an intermediate urea concentration followed by dialysis to zero urea, a distinct minimum yield was obtained at 3 M urea (though the protein concentration was also critical).

The problem of protein solubility during refolding from inclusion bodies has been successfully addressed using buffer exchange from a 'strong' to a 'weak' denaturant (as defined in the preceding section) prior to complete removal of denaturant (Jones, Olson and Shire, 1985). An aspect of the patent involves solubilizing the inclusion body and 'exchange against a comparable concentration of denaturant whose chaotropic properties are inherently weaker (e.g. GuHCl to urea) or by dilution to (if the solubility characteristics of the subject protein will permit) a decreased concentration of the same strong denaturant'. The approach has been used with human, bovine and porcine growth hormone, urokinase, pro-renin and Sarc protein. A further aspect of the patent involves the use of this technique for maintenance of solubility during oxidation of disulphide bonds, as discussed in a subsequent section.

SOLVENT SYSTEMS

Much of the work investigating the effects of pH on protein structure has been concerned with studying the acid and alkali unfolding transitions of protein. Often pH is considered in relation to thiol exchange reactions during disulphide bond formation. The effects of pH on the refolding of recombinant proteins have been recognized as important (Marston, 1986) and a pH drop from 10·7 to 8 is used to refold prochymosin purified from inclusion bodies (Marston *et al.*, 1988). In the cases of IL-1β (Craig *et al.*, 1987) and IFN-γ (Mulkerrin and Wetzel, 1989) unfolding at certain pH values is irreversible. Tumour necrosis factor shows no dependence of yield of refolding over a

broad range of pH (Craig, Wingfield and Pain, 1987). The removal of acidic components tightly complexed to unfolded protein obtained from inclusion bodies may be a consideration in refolding protocols (Darby and Creighton, 1990).

In the case of β-lactamase, which has been well studied, pH dependence of both the refolding pathway and kinetics have been observed to be associated specifically with the titration of an anomalous carboxyl and a histidine residue, respectively (Craig, 1986; Craig, Wingfield and Pain, 1987). Refolding of *iso*-2-cytochrome *c* at high pH (Nall, 1986) leads to the formation of an alkaline form of the protein which is more stable than the native state. This alkaline state is formed via non-native structural intermediates.

The effects of ionic strength during refolding of proteins must first be addressed from the point of preferential hydration of protein molecules (Timasheff and Arakawa, 1989; Arakawa, Bhat and Timasheff, 1990). In the case of IFN (Hsu and Arakawa, 1985) lower ionic strength gives better recovery of non-aggregated protein during refolding. A patent based on this for production of IFN prefers ammonium acetate to phosphate buffer at low ionic strength, giving an improved ratio of non-aggregated to aggregated protein (Arakawa, 1986). The refolding pathway in the model BPTI system appears to be independent of ionic strength (Creighton, 1980b) but is dependent on the type of ion according to the Hofmeister series of salt stabilization of native structure.

On the whole, lower ionic strengths are used in refolding processes, although high ionic strength has been used to fold protein spontaneously in the presence of intermediate denaturant concentrations. It has been shown (Mitchinson and Pain, 1985) that unfolded β-lactamase in urea will fold to the native state on addition of ammonium sulphate. The proposed mechanism involves sulphate changing the degree of preferential solvation of the protein surface by water and thereby weakening the denaturing effect. The structure-stabilizing effect of sulphate is a recognized parameter when considering protein folding protocols (Jaenicke and Rudolph, 1989). Concentrations of MES buffer salt in the range 0·25–1·0 M have been demonstrated to have a similar (though not identical) effect to ammonium sulphate on inducing the folding of intermediate states in denaturing conditions (Craig, 1986).

An appreciation of solvation effects is important not only in refolding proteins but also in the handling and formulation of secreted folded molecules. This is demonstrated for tissue plasminogen activator (Duffy, Prior and Scott, 1990) where an arginine/citric acid buffer is required to obtain concentrated protein solutions for intravenous administration.

The effect of ligands is another factor to be examined in attempts to refold recombinant proteins. In the case of recombinant horseradish peroxidase (Smith *et al.*, 1990) a number of factors have been found to influence folding. The presence of calcium ions during folding appears to be critical, probably by stabilizing a native-like conformation around the haem-binding site and thereby influencing the equilibrium. Other factors involved are the presence of haem, the oxidation conditions and the urea concentration.

The non-ionic detergent, lauryl maltoside, has been used to refold

rhodanese successfully from GuHCl (Tandon and Horowitz, 1986, 1989). The use of detergents provides a further option for maintaining solubility during folding.

Schein (1990) has recently examined factors affecting the solubility of proteins, and the prospect of modifying solubility by protein engineering is discussed.

RE-OXIDATION OF DISULPHIDE BONDS

The methodology and protocols to adopt for formation and characterization of disulphide bonds in proteins are set out by Creighton (1989). Oxidation of cysteine residues between molecules during isolation and purification from inclusions can lead to aggregation. Protection of thiol groups using the sulphitolysis reaction is used during recombinant insulin production (Chance and Hoffman, 1981a,b), recombinant IGF-II (Furman *et al.*, 1987) and urokinase (Jones, Olson and Shire, 1985). Sulphitolysis involves classic chemical reactions using sodium sulphite in the presence of sodium tetrathionate to protect free thiol groups. Cyanogen bromide cleavage of fusion protein constructs at methionine residues in the linker region can lead to modification of thiol groups. The addition of thiosulphate to the cleavage reaction mixture is found to prevent these modifications (Di Marchi, 1984) and, as a bonus, the thiol groups readily sulphitolyse when exposed to sodium sulphite only. This reduces production costs as the expensive sodium tetrathionate is not required.

The promotion of correct intramolecular disulphide pairing during refolding is the next step. Air oxidation is slow, and when groups have been protected, addition of excess cysteine (Furman *et al.*, 1987) or mixtures of reduced and oxidized thiol reagents (Creighton, 1989) can be used. The exchange reactions are quenched either by a rapid pH drop to below 6 and/or desalting to remove the reagent. Aggregation often occurs, and wrongly paired disulphides are found which are separated from the correct product using gel filtration (Langley *et al.*, 1987a), reverse phase (Furman *et al.*, 1987) and ion-exchange (Smith *et al.*, 1989) chromatography or combinations thereof. Copper chloride and (o-phenanthroline)$_2Cu^{2+}$ have been used to selectively oxidize IFN-β and IL-2 solubilized in low chaotrope conditions at pH 5·5–9 (Koths and Halennbeck, 1984). Examples where low chaotrope concentrations have been used to maintain solubility during oxidative refolding are PDGF-β (Hoppe, Weich and Eichner, 1989) and other proteins (Jones, Olson and Shire, 1985). The re-oxidation of cysteine residues in reduced bovine growth hormone maintained in a stable state with native-like secondary structure (in 4·5 M urea) has been examined (Holzman, Brems and Dougherty, 1986). Refolding and re-oxidation of the monomeric species was found to occur with a yield of around 90%.

A method has been claimed (Seely and Yang, 1988) where a reduced protein which is maintained in its unfolded form is reacted with excess cysteine for 12–36 h, allowing native disulphide bonds to form, resulting in reduced levels of aggregation and inactive protein formation prior to subse-

quent refolding. Langley *et al.* (1987a) also oxidized native disulphides in GuHCl-denatured bGH using air oxidation over 20 h. The bGH was subsequently refolded by removal of GuHCl after separation from aggregates.

The procedure of refolding proteins bound non-covalently to a solid column matrix (Creighton, 1986) has a potential application in this field. For example, the binding of urea-unfolded protein to an FPLC ion-exchange column with strict control of the concentrations of urea and reduced/oxidized thiol reagent allows tight control over the oxidation of disulphides and refolding. The fact that the protein is held close to the column reduces the problem of aggregation. The folded protein can be eluted using normal chromatographic methods and the system allows for easy recycling of incorrectly folded material.

CHARACTERIZATION OF THE FINAL PRODUCT

Following its production, the folded, active recombinant protein requires characterization before clinical trials, etc. The N-terminal methionine residue is not always completely processed during production in *E. coli*, which can lead to heterogeneous preparations. The slight differences in charge between protein with and without N-terminal Met has been exploited to obtain homogeneous preparations of recombinant IL-1β and bGH (Wingfield *et al.*, 1987a) and IL-2 (Yamada *et al.*, 1986). A specific methionine aminopeptidase has been characterized which may have an application in removing unprocessed methionine from the N-terminus of recombinant proteins (Wingfield *et al.*, 1989). Aminopeptidase P has been successfully used to remove the N-terminal methionine from recombinant interleukin-6. Other charge heterogeneity can occur in recombinant proteins due to deamidation, and genetic engineering of the molecule has been successfully used to stabilize IL-1α against this (Wingfield *et al.*, 1987b).

Many recombinant proteins that have been produced in *E. coli* are glycosylated in their authentic form. Parekh *et al.* (1989) recently reviewed the biological significance of *N*-glycosylation. The conformational stability and folding of granulocyte macrophage colony stimulating factor does not appear to be affected by glycosylation (Wingfield *et al.*, 1988). Glycosylation may be involved in preventing aggregation of partially unfolded intermediates during folding (Schulke and Schmid, 1988) and lack of sugar chains may be the cause of low yields of refolded protein (due to aggregation) obtained from some inclusion bodies (for example, horseradish peroxidase C; Smith *et al.*, 1990). Either the lack of, or the incorrect, glycosylation of vesicular stomatitis virus G proteins has been implicated in the formation of intermolecular disulphide-bonded aggregates of the protein (Machamer and Rose, 1988). Glycosylation of α_1-antitrypsin increases its solubility and resistance to aggregation without affecting the thermodynamic stability (L. Powell and R.H. Pain, submitted for publication).

The purity of proteins to be used as pharmaceuticals must be assessed and the procedures for toxicology, pyrogen testing, etc. have been outlined with specific reference to tPA (Anicetti, Keyt and Hancock, 1989). For approval

of a protein product, chemical identity with the authentic natural product is required. In some cases structural and conformational identity is required, and the patent for IFN production (Arakawa, 1986) includes biophysical characterization of the recombinant product using CD. The range of biophysical techniques available, such as CD, NMR, analytical ultracentrifugation and fluorescence, have all been used to characterize a number of recombinant proteins, such as TNF (Wingfield, Pain and Craig, 1987), bGH (Langley *et al.*, 1987b), IL-1β (Craig *et al.*, 1987), IGF-I (Elliot *et al.*, 1990) and a genetically engineered, cysteine-free analogue of basic fibroblast growth factor (Arakawa *et al.*, 1989). Once characterized the analysis serves as proof of identity and also as a baseline for production batch quality assurance. The value of careful product analysis is illustrated in the case of IGF-II, where folded protein containing a mis-matched disulphide has been isolated after renaturation and re-oxidation (Smith *et al.*, 1989). The protein is active but is not recognized by the major serum-binding protein. As the binding protein controls serum levels of the growth factor, this could cause unwanted biological side-effects upon administration.

FUTURE PROSPECTS

The whole area of bacterial expression of protein allows great scope for optimization and improvement. The question of whether to follow the soluble or insoluble production route remains open for debate (Schein, 1989). The factors of cost, ease of purification, level of expression and overall recovery must all be addressed.

The successful co-expression of GroE with Rubisco in *E. coli*, resulting in subunit assembly of the protein (Goloubinoff, Gatenby and Lorimer, 1989), suggests a possible application for chaperonins in high-expression production systems. The implications of chaperonin-assisted polypeptide folding for production of functional heterologous proteins in bacteria are reviewed by Gatenby, Viitanen and Lorimer (1990) and discussed further by Geisow (1991). The possibility of mimicking other facets of mammalian expression in bacterial systems (Schein, 1989) is a further consideration in optimizing the production process. Combined with elucidation of the role of the compact, 'molten globule' state and the chaperone families of proteins in the protein-folding story there is ample opportunity for further capitalizing on bacterial expression.

Acknowledgements

The authors wish to acknowledge the support of the Science and Engineering Research Council, particularly for funding for the Service in Solution Conformation and associated equipment. We also acknowledge the support for Roman Hlodan by ICI Pharmaceuticals. We are indebted to Susan Lee for her attention to detail in preparing the manuscript.

References

ADAMS, B., BURGESS, R.J. AND PAIN, R.H. (1985). The folding and mutual interactions of the domains of yeast 3-phosphoglycerate kinase. *European Journal of Biochemistry* **152**, 715–720.

ANFINSEN, C.B. (1967). The formation of the tertiary structure of proteins. *Harvey Lectures* **61**, 95–116.

ANFINSEN, C.B. (1972). The formation and stabilisation of protein structure. *Biochemical Journal* **128**, 737–754.

ANFINSEN, C.B. AND SCHERAGA, H.A. (1975). Experimental and theoretical aspects of protein folding. *Advances in Protein Chemistry* **29**, 205–300.

ANICETTI, V.R., KEYT, B.A. AND HANCOCK, W.S. (1989). Purity analysis of protein pharmaceuticals produced by recombinant DNA technology. *Trends in Biotechnology* **7**, 342–349.

ARAKAWA, T. (1986). *Method for purifying interferon and composition of matter produced thereby.* European Patent 0, 199, 568.

ARAKAWA, T., BHAT, R. AND TIMASHEFF, S.N. (1990). Why preferential hydration does not always stabilise the native structure of globular proteins. *Biochemistry* **29**, 1924–1931.

ARAKAWA, T., HSU, Y.R., SCHIFFER, S.G., TSAI, T.B., CURLESS, C. AND FOX, G.M. (1989). Characterisation of a cysteine-free analog of recombinant human basic fibroblast growth factor. *Biochemical and Biophysical Research Communications* **161**, 335–341.

BATY, D. AND LAZDUNSKI, C. (1979). An anti-(signal peptide) antibody: purification, properties and use as a conformational probe. *European Journal of Biochemistry* **102**, 503–507.

BECKMANN, R.D., MIZZEN, L.A. AND WELCH, W.J. (1990). Interaction of Hsp70 with newly synthesised proteins: implications for protein folding and assembly. *Science* **248**, 850–853.

BERGMAN, L.W. AND KUEHL, W.M. (1979). Formation of intermolecular disulphide bonds on nascent immunoglobulin polypeptides. *Journal of Biological Chemistry* **254**, 5690–5694.

BERMANN-MARKS, C., VASSER, M., NG, P., HENZEL, W. AND ANDERSON, S. (1986). Production of native correctly folded bovine pancreatic trypsin inhibitor by *E. coli*. *Journal of Biological Chemistry* **261**, 7115–7118.

BLOND, S. AND GOLDBERG, M. (1987). Partly native epitopes are already present on early intermediates in the folding of tryptophan synthase. *Proceedings of the National Academy of Sciences USA* **84**, 1147–1151.

BLOND-ELGUINDI, S. AND GOLDBERG, M.E. (1990). Kinetic characterisation of early immunoreactive intermediates during the refolding of guanidine-unfolded *Escherichia coli* tryptophan synthase β_2 subunits. *Biochemistry* **29**, 2409–2417.

BOCHKAREVA, E.S., LISSIN, N.M. AND GIRSHOVICH, A.S. (1988). Transient association of newly synthesised unfolded proteins with the heat shock GroEL protein. *Nature* **336**, 254–257.

BRANDTS, J.F., BRENNAN, M. AND LIN, L.-N. (1977). Unfolding and refolding occur much faster for a proline-free protein than for most proline containing proteins. *Proceedings of the National Academy of Sciences USA* **74**, 4178–4181.

BRANDTS, J.F., HALVORSON, H.R. AND BRENNAN, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis–trans* isomerism of proline residues. *Biochemistry* **14**, 4953–4963.

BREMS, D.N. (1988). Solubility of different folding conformers of bovine growth hormone. *Biochemistry* **27**, 4541–4546.

BREMS, D.N., PLAISTED, S.M., KAUFFMAN, E.W. AND HAVEL, H.A. (1986). Characterisation of an associated equilibrium folding intermediate of bovine growth hormone. *Biochemistry* **25**, 6539–6543.

BREMS, D.N., PLAISTED, S.M., KAUFFMAN, E.W., LUND, M. AND LEHRMAN, S.R.

(1987). Helical formation in isolated fragments of bovine growth hormone. *Biochemistry* **26**, 7774–7778.

BUCHNER, J., SCHMIDT, H. FUCHS, M., JAENICKE, R., RUDOLPH, R., SCHMID, F.X. AND KIEFHABER, T. (1991). GroE facilitates refolding of citrate synthase by suppressing aggregation. *Biochemistry* **30**, 1586–1591.

BUILDER, S.E. AND OGEZ, J.R. (1985). *Purification of heterologous protein from host cells by solubilising and using molecular sieve or centrifugation.* US Patent 4,511,502.

BULLEID, N.J. AND FREEDMAN, R.B. (1988). Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes. *Nature* **335**, 649–651.

BYCHKOVA, V.E., PAIN, R.H. AND PTITSYN, O.B. (1988). The 'molten globule' state is involved in the translocation of proteins across membranes? *FEBS Letters* **238**, 231–234.

BYCROFT, M., MATOUSCHEK, A., KELLIS, J.T., SERRANO, L. AND FERSHT, A.R. (1990). Detection and characterisation of a folding intermediate in barnase by NMR. *Nature* **346**, 488–490.

CARRELL, R.W. (1986). α1-Antitrypsin: molecular pathology, leukocytes, and tissue damage. *Journal of Clinical Investigation* **78**, 1427–1431.

CHAN, W.W.C., MORT, J.S., CHONG, D.K.K. AND MACDONALD, P.D.M. (1973). Studies on protein subunits III. Kinetic evidence for the presence of active subunits during the renaturation of muscle aldolase. *Journal of Biological Chemistry* **248**, 2776–2784.

CHANCE, R.E. AND HOFFMAN, J.A. (1981a). *Process for producing an insulin.* European Patent 0,037,256.

CHANCE, R.E. AND HOFFMAN, J.A. (1981b). *Process for producing an insulin.* US Patent 4,421,685.

CHENG, Y.E., KWOH, D., KWOH, T.J., SOFTVEDT, B.C. AND ZIPSER, D. (1981). Stabilisation of a degradable protein by its overexpression in *E. coli. Gene* **14**, 121–130.

CHRISTENSEN, H.C. AND PAIN, R.H. (1991). Molten globule intermediates and protein folding. *European Biophysics Journal* **19**, 221–229.

COOK, R.A. AND KOSHLAND, D.E., JR (1969). Specificity in the assembly of multisubunit proteins. *Proceedings of the National Academy of Sciences USA* **64**, 247–254.

CRAIG, S. (1986). Critical residues in the folding of β-lactamase from *Staphylococcus aureus.* PhD thesis, University of Newcastle upon Tyne, UK.

CRAIG, S., WINGFIELD, P. AND PAIN, R.H. (1987). Different effects of pH in the refolding of proteins. *Biochemical Society Transactions* **15**, 1146–1147.

CRAIG, S., SCHMEISSNER, U., WINGFIELD, P. AND PAIN, R.H. (1987). Conformation, stability and folding of interleukin 1β. *Biochemistry* **26**, 3570–3576.

CREIGHTON, T.E. (1978). Experimental studies of protein folding and unfolding. *Progress in Biophysics and Molecular Biology* **33**, 231–298.

CREIGHTON, T.E. (1979). Intermediates in the refolding of reduced ribonuclease A. *Journal of Molecular Biology* **129**, 411–431.

CREIGHTON, T.E. (1980a). A three-disulphide intermediate in refolding of reduced ribonuclease A with a folded conformation. *FEBS Letters* **118**, 283–288.

CREIGHTON, T.E. (1980b). Role of the environment in the refolding of reduced pancreatic trypsin inhibitor. *Journal of Molecular Biology* **144**, 521–550.

CREIGHTON, T.E. (1984). Pathways and mechanisms of protein folding. *Advances in Biophysics* **18**, 1–20.

CREIGHTON, T.E. (1986). *A process for the production of a protein.* International Patent WO86/05809.

CREIGHTON, T.E. (1989). Disulphide bonds between cysteine residues 155–167. In *Protein Structure, a Practical Approach* (T.E. Creighton, Ed.). IRL Press, Oxford.

CREIGHTON, T.E. AND GOLDENBERG, D.P. (1984). Kinetic role of metastable native-like two disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. *Journal of Molecular Biology* **179**, 497–526.
CREIGHTON, T.E., HILLSON, D.A. AND FREEDMAN, R.B. (1980). Catalysis by protein-disulphide isomerase of the unfolding and refolding of proteins with disulphide bonds. *Journal of Molecular Biology* **142**, 43–62.
DALMAN, F.C., BRESNICK, E.H., PATEL, P.D., PERDEW, G.H., WATSON, S.J. AND PRATT, W.B. (1989). Direct evidence that the glucocorticoid receptor binds to hsp90 at or near the termination of receptor translation *in vitro*. *Journal of Biological Chemistry* **264**, 19815–19821.
DARBY, N.J. AND CREIGHTON, T.E. (1990). Folding proteins. *Nature* **344**, 715–716.
DESHAIES, R.J., KOCH, B.D., WERNER-WASHBURNE, M., CRAIG, E.A. AND SCHEKMAN, R. (1988). A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. *Nature* **332**, 800–805.
DIERSTEIN, R. AND WICKNER, W. (1985). The leader region of pre-maltose binding proteins binds amphiphiles. *Journal of Biological Chemistry* **260**, 15919–15924.
DI MARCHI, R.D. (1984). *Process for inhibiting undesired thiol reactions during cyanogen bromide cleavage of peptides*. US Patent 4,451,396.
DINTZIS, H.M. (1961). Assembly of the peptide chains of hemoglobin. *Proceedings of the National Academy of Sciences USA* **47**, 247–261.
DOBSON, C.M. AND EVANS, P.A. (1988). Trapping folding intermediates. *Nature* **335**, 666–667.
DUFFY, S.A., PRIOR, C.P. AND SCOTT, R.W. (1990). *Method to solubilise tissue plasminogen activator*. US Patent 4,898,826.
ELLIOTT, S., FAGIN, K.D., NARHI, L.O., MILLAR, J.A., JONES, M., KOSKI, R., PETERS, M., HSIEH, P., SACHDEV, R., ROSENFELD, R.D., ROHDE, M.F. AND ARAKAWA, T. (1990). Yeast-derived recombinant human insulin-like growth factor I: production, purification and structural characterisation. *Journal of Protein Chemistry* **9**, 95–104.
EPSTEIN, C.J., GOLDBERGER, R.F., YOUNG, D.M. AND ANFINSEN, C.B. (1962). A study of the factors influencing the rate and extent of enzymic reactivation during reoxidation of reduced ribonuclease. *Archives of Biochemistry and Biophysics* Supplement **1**, 223–231.
FAHEY, R.C., HUNT, J.S. AND WINDHAM, G.C. (1977). On the cysteine and cystine content of proteins. Differences between intracellular and extracellular proteins. *Journal of Molecular Evolution* **10**, 155–160.
FERENCI, T. AND RANDALL, L.L. (1979). Precursor maltose-binding protein is active in binding substrate. *Journal of Biological Chemistry* **254**, 9979–9981.
FISCHER, G., BANG, E. AND BERGER, H. (1989). Kinetic β-deuterium isotope effects suggest a covalent mechanism for the protein folding enzyme peptidylprolyl *cis/trans*-isomerase. *FEBS Letters* **250**, 267–270.
FISCHER, G., BANG, H. AND MECH, C. (1984). Detection of enzyme catalysis for *cis–trans*-isomerisation of peptide bonds using proline-containing peptides as substrates. *Biomedica Biochimica Acta* **43**, 1101–1111.
FISCHER, G., WITTMAN-LIEBOLD, B., LANG, K., KIEFHABER, T. AND SCHMID, F.X. (1989). Cyclophilin and peptidyl-prolyl *cis–trans*-isomerase are probably identical proteins. *Nature* **337**, 476–478.
FRANK, B.H. AND CHANCE, R.E. (1983). Two routes for producing human insulin utilising recombinant DNA technology. *Münchner Medizinischer Wochenschrift* **125** (Suppl. 1), S14–20.
FREEDMAN, R.B. (1984). Native disulphide bond formation in protein biosynthesis: evidence for the role of protein disulphide isomerase. *Trends in Biochemical Sciences* **9**, 438–441.
FREEDMAN, R.B. (1989). Protein disulphide isomerase: multiple roles in the modification of nascent secretory proteins. *Cell* **57**, 1069–1072.
FREEDMAN, R.B., BROCKWAY, B.E. AND LAMBERT, N. (1984). Protein disulphide-

isomerase and the formation of native disulphide bonds. *Biochemical Society Transactions* **12**, 929–932.

FREEDMAN, R.B., BULLEID, N.J., HAWKINS, H.C. AND PAVER, J.L. (1989). Role of protein disulphide-isomerase in the expression of native proteins. *Biochemical Society Symposium* **55**, 167–192.

FURMAN, T.C., EPP, J., HSUING, H.M., HOSKINS, J., LONG, G.L., MENDELSOHN, L.G., SCHONER, B., SMITH, D.P. AND SMITH, M.C. (1987). Recombinant human insulin-like growth factor II expressed in *Escherichia coli*. *Bio/Technology* **5**, 1047–1051.

GALAT, A., CREIGHTON, T.E., LORD, R.C. AND BLOUT, E.R. (1981). Circular dichroism, Raman spectroscopy, and gel filtration of trapped folding intermediates of ribonuclease. *Biochemistry* **20**, 594–601.

GAREL, J.-R. AND BALDWIN, R.L. (1973). Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proceedings of the National Academy of Sciences USA* **70**, 3347–3351.

GATENBY, A.A., VIITANEN, P.V. AND LORIMER, G.M. (1990). Chaperonin assisted polypeptide folding and assembly: implications for the production of functional proteins in bacteria. *Trends in Biotechnology* **8**, 354–358.

GEISOW, M.J. (1991). Do-it-yourself molecules: chaperonin self assembly. *Trends in Biotechnology* **9**, 45.

GEORGOPOULOS, C. AND ANG, D. (1990). The *Escherichia coli groE* chaperonins. *Seminars in Cell Biology* **1**, 19–25.

GERL, M. AND JAENICKE, R. (1988). Self-assembly of apoferritin from horse spleen after reversible chemical modification with 2,3-dimethylmaleic anhydride. *Biochemistry* **27**, 4089–4096.

GETHING, M.-J. AND SAMBROOK, J. (1990). Transport and assembly processes in the endoplasmic reticulum. *Seminars in Cell Biology* **1**, 65–72.

GIERASCH, L.M. (1989). Signal sequences. *Biochemistry* **28**, 923–929.

GIRG, R., JAENICKE, R. AND RUDOLPH, R. (1983). Dimers of porcine skeletal muscle lactate dehydrogenase produced by limited proteolysis are enzymically active in the presence of stabilising salt. *Biochemistry International* **7**, 433–441.

GIRG, R., RUDOLPH, R. AND JAENICKE, R. (1981). Limited proteolysis of porcine-muscle lactic dehydrogenase by thermolysin during reconstitution yields dimers. *European Journal of Biochemistry* **119**, 301–305.

GOLDBERG, M.E. (1985). The second translation of the genetic message: protein folding and assembly. *Trends in Biochemical Sciences* **10**, 388–391.

GOLDBERGER, R.F., EPSTEIN, C.J. AND ANFINSEN, C.B. (1963). Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *Journal of Biological Chemistry* **238**, 628–635.

GOLOUBINOFF, P., GATENBY, A. AND LORIMER, G.H. (1989). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**, 44–47.

GOLOUBINOFF, P., CHRISTELLER, J.T., GATENBY, A.A. AND LORIMER, G.H. (1989). Reconstitution of active dimeric ribulose biphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* **342**, 884–889.

GOODMAN, E.M. AND KIM, P.S. (1989). Folding of a peptide corresponding to the α-helix in bovine pancreatic trypsin inhibitor. *Biochemistry* **28**, 4343–4347.

GOTO, Y. AND HAMAGUGCHI, K. (1982). Unfolding and refolding of the constant fragment of an immunoglobulin light chain. *Journal of Molecular Biology* **156**, 891–910.

GRAY, A.M. AND MASON, A.J. (1990). Requirement for activin A and transforming growth factor-β1 pro-regions in homodimer assembly. *Science* **247**, 1328–1330.

HAASE-PETTINGELL, C.A. AND KING, J. (1988). Formation of aggregates from a thermolabile *in vivo* folding intermediate in P22 tail spike maturation. *Journal of Biological Chemistry* **263**, 4977–4983.

HARRIS, T.R. (1990). *Protein Production by Biotechnology*. Elsevier Scientific, Barking.

HARRISON, R.K. AND STEIN, R.L. (1990). Mechanistic studies of peptidyl prolyl *cis–trans* isomerase: evidence for catalysis by distortion. *Biochemistry* **29**, 1684–1689.

HARRISON, T.M., CHIDGEY, M.A.J., BRAMMAR, W.J. AND ADAMS, G.J. (1989). The pro-peptide is not necessary for active renin secretion from transfected mammalian cells. *Proteins: Structure, Function, Genetics* **5**, 259–265.

HARTLEY, D.L. AND KANE, J.F. (1988). Properties of inclusion bodies from recombinant *Escherichia coli*. *Biochemical Society Transactions* **16**, 101–102.

HAVEL, H.A., KAUFFMAN, E.W., PLAISTED, S.M. AND BREMS, D.N. (1986). Reversible self-association of bovine growth hormone during equilibrium unfolding. *Biochemistry* **25**, 6533–6538.

HEMMINGSEN, S.M. (1990). The plastid chaperonin. *Seminars in Cell Biology* **1**, 47–54.

HERMANN, R. (1990). *Production of biologically active protein from inactive form – by denaturing and contacting with a molecular sieve*. US Patent 4,965,344.

HERMANN, R., JAENICKE, R. AND PRICE, N.C. (1985). Evidence for active intermediates during the reconstitution of yeast phosphoglycerate mutase. *Biochemistry* **24**, 1817–1821.

HERMANN, R., JAENICKE, R. AND RUDOLPH, R. (1981). Analysis of the reconstitution of oligomeric enzymes by cross-linking with glutaraldehyde: kinetics of reassociation of lactic dehydrogenase. *Biochemistry* **20**, 5195–5201.

HERMANN, R., RUDOLPH, R., JAENICKE, R., PRICE, N.C. AND SCOBBIE, A. (1983). The reconstitution of denatured phosphoglycerate mutase. *Journal of Biological Chemistry* **258**, 11014–11019.

HEROLD, M. AND KIRSCHNER, K. (1990). Reversible dissociation and unfolding of aspartate aminotransferase from *Escherichia coli*: characterisation of a monomeric intermediate. *Biochemistry* **29**, 1907–1913.

HOESS, A., ARTHUR, A.K., WANNER, G. AND FANNIG, E. (1988). Recovery of soluble, biologically active recombinant proteins from total bacterial lysates using ion-exchange resin. *Bio/Technology* **6**, 1214–1217.

HOLZMAN, T.F., BREMS, D.N. AND DOUGHERTY, J.J., JR (1986). Reoxidation of reduced bovine growth hormone from a stable secondary structure. *Biochemistry* **25**, 6907–6917.

HOPP, T.P., PRICKETT, K.S., PRICE, V.L., LIBBY, R.T., MARCH, C.J., CERETTI, D.P., URDAL, D.L. AND CONLON, P.J. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* **6**, 1204–1210.

HOPPE, J., WEICH, H.A. AND EICHNER, W. (1989). Preparation of biologically active platelet-derived growth factor type BB from a fusion protein expressed in *Escherichia coli*. *Biochemistry* **28**, 2956–2960.

HORWICH, A.L., NEUPERT, W. AND HARTL, F.-U. (1990). Protein-catalysed protein folding. *Trends in Biotechnology* **8**, 126–131.

HSIUNG, H.M., MAYNE, N.G. AND BECKER, G.W. (1986). High-level expression, efficient secretion and folding of human growth hormone in *Escherichia coli*. *Bio/Technology* **4**, 991–995.

HSU, Y.R. AND ARAKAWA, T. (1985). Structural studies on acid unfolding and refolding of recombinant human interferon-γ. *Biochemistry* **24**, 7959–7963.

JAENICKE, R. (1980). *Protein Folding*. Elsevier, Amsterdam.

JAENICKE, R. (1987). Folding and association of proteins. *Progress in Biophysics and Molecular Biology* **49**, 117–237.

JAENICKE, R. AND RUDOLPH, R. (1986). Refolding and association of oligomeric proteins. *Methods in Enzymology* **131**, 218–250.

JAENICKE, R. AND RUDOLPH, R. (1989). Folding proteins. In *Protein Structure: a Practical Approach* (T.E. Creighton, Ed.), pp.191–224. IRL Press, Oxford.

JAENICKE, R., RUDOLPH, R. AND FEINGOLD, D.S. (1986). Dissociation and *in vitro* reconstitution of bovine liver uridine diphosphoglucose dehydrogenase. The paired subunit nature of the enzyme. *Biochemistry* **25**, 7285–7287.

JOHANSON, K.O., WETLAUFER. D.B., REED, R.G. AND PETERS, T. (1981). Refolding of bovine serum albumin and its proteolytic fragments. *Journal of Biological Chemistry* **256**, 445–450.

JONES, A.J.S., OLSON, K.C. AND SHIRE, S.J. (1985). *Purification and activity assurance of precipitated heterologous proteins.* US Patent 4,512,922.

KAISER, C.A., PREUSS, D., GRISAFI, P. AND BOTSTEIN, D. (1987). Many random sequences functionally replace the secretion signal sequence of yeast invertase. *Science* **235**, 312–317.

KANE, J.F. AND HARTLEY, D.L. (1988). Formation of recombinant protein inclusion bodies in *Escherichia coli. Trends in Biotechnology* **6**, 95–101.

KELLEY, R.F. AND RICHARDS, F.M. (1987). Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry* **26**, 6765–6774.

KELLEY, R.F. AND STELLWAGEN, E. (1984). Conformational transitions of thioredoxin in guanidine hydrochloride. *Biochemistry* **23**, 5095–5102.

KIEFHABER, T., QUAAS, R., HAHN, U. AND SCHMID, F.X. (1990a). Folding of ribonuclease T_1. 1. Existence of multiple unfolded states created by proline isomerisation. *Biochemistry* **29**, 3053–3061.

KIEFHABER, T., QUAAS, R., HAHN, U. AND SCHMID, F.X. (1990b). Folding of ribonuclease T_1. 2. Kinetic models for the folding and unfolding reactions. *Biochemistry* **29**, 3061–3070.

KIM, P.S. AND BALDWIN, R.L. (1990). Intermediates in the folding reactions of small proteins. *Annual Review of Biochemistry* **59**, 631–660.

KIMMENADE, A., BOND, M.W., SCHUMAKERK, J.H., LAQUOI, C. AND KASTELEIN, R.A. (1988). Expression, renaturation and purification of recombinant human interleukin 4 from *Escherichia coli. European Journal of Biochemistry* **173**, 109–114.

KING, J. (1986). Genetic analysis of protein folding pathways. *Bio/Technology* **4**, 297–303.

KING, J., HAASE, C. AND YU, M.-H. (1987). Temperature-sensitive mutations affecting kinetic steps in protein-folding pathways. In *Protein Engineering* (D.L. Oxender and C.F. Fox, Eds), pp.109–121. Alan R. Liss, New York.

KINGSMAN, S.M. AND KINGSMAN, A.J. (1988). *Gene Expression*. Blackwell Scientific Press, Oxford.

KISHIDA, F., AZUMA, T. AND HAMAGUCHI, K. (1976). Formation of interchain disulphide bonds in Bence Jones proteins and immunoglobulins. *Journal of Biochemistry* **79**, 91–105.

KITANO, K., FUJIMOTO, S., NAKAO, M., WATANABE, T. AND NAKAO, Y. (1987). Intracellular degradation of recombinant proteins in relation to their location in *Escherichia coli* cells. *Journal of Biotechnology* **5**, 77–86.

KOIVU, J., MYLLYLÄ, R., HELAAKOSKI, T., PIHLAJANIEMI, T., TASANEN, K. AND KIVIRIKKO, K.I. (1987). A single polypeptide acts both as the β-subunit of prolyl 4-hydroxylase and as a protein disulphide-isomerase. *Journal of Biological Chemistry* **262**, 6447–6449.

KOTHS, K.E. AND HALENNBECK, R.F. (1984). *Promoting disulphide bond formation in recombinant proteins using oxidation promoters containing copper cation*. US Patent 4,572,798.

KUSTERS, R., DE VRIJE, T., BREUKINK, E. AND DE KRUIFF, B. (1989). SecB protein stabilises a translocation-competent state of purified prePhoE protein. *Journal of Biological Chemistry* **264**, 20827–20830.

KUWAJIMA, K. (1989). The molten globule state as a clue for understanding the folding and co-operativity of globular-protein structure. *Proteins: Structure, Function and Genetics* **6**, 87–103.

LAMBERT, N. AND FREEDMAN, R.B. (1984). Chemical modification of essential residues in purified rat liver protein disulphide-isomerase. *Biochemical Society Transactions* **12**, 1042.

LANG, K., SCHMID, F.X. AND FISCHER, G. (1987). Catalysis of protein folding by prolyl isomerase. *Nature* **329**, 268–270.

LANGLEY, K.E., BERG, T.F., STRICKLAND, T.W., FENTON, D.M., BOONE, T.C. AND WYPYCH, J. (1987a). Recombinant-DNA-derived bovine growth hormone from *Escherichia coli*. 1. Demonstration that the hormone is expressed in reduced form, and isolation of the hormone in oxidised, native form. *European Journal of Biochemistry* **163**, 313–321.

LANGLEY, K.E., LAI, P.H., WYPYCH, J., EVERETT, R.R., BERG, T.F., KRABILL, L.F., DAVID, J.M. AND SOUZA, L.M. (1987b). Recombinant-DNA-derived bovine growth hormone from *Escherichia coli*. 2. Biochemical, biophysical, immunological and biological comparison with the pituitary hormone. *European Journal of Biochemistry* **163**, 323–330.

LECKER, S.H., DRIESSEN, A.J.M. AND WICKNER, W. (1990). ProOmpA contains secondary and tertiary structure prior to translocation and is shielded from aggregation by association with SecB protein. *EMBO Journal* **9**, 2309–2314.

LECKER, S., LILL, R., ZIEGELHOFFER, T., GEORGOPOULOS, C., BASSFORD, P.J., KUMAMOTO, C. AND WICKNER, W. (1989). Three pure chaperone proteins of *Escherichia coli* – SecB, trigger factor and GroEL – form stable complexes with precursor proteins *in vitro*. *EMBO Journal* **8**, 2703–2709.

LIN, L.-N. AND BRANDTS, J.F. (1978). Further evidence suggesting that the slow phase in protein unfolding and refolding is due to proline isomerisation: a kinetic study of carp parvalbumins. *Biochemistry* **17**, 4102–4110.

LINDSAY, C.D. AND PAIN, R.H. (1990). The folding and solution conformation of penicillin G acylase. *European Journal of Biochemistry* **192**, 133–141.

LINDSAY, C.D. AND PAIN, R.H. (1991). Refolding and assembly of penicillin acylase, an enzyme composed of two polypeptide chains that result from proteolytic activation. *Biochemistry* **30**, 9034–9040.

LONDON, J., SKRZYNIA, C. AND GOLDBERG, M.E. (1974). Renaturation of *Escherichia coli* tryptophanase after exposure to 8M urea. *European Journal of Biochemistry* **47**, 409–415.

MACHAMER, C.E. AND ROSE, J.K. (1988). Vesicular stomatitis virus G proteins with altered glycosylation sites display temperature sensitive intracellular transport and are subject to aberrant intermolecular disulphide bonding. *Biochemistry* **263**, 5955–5960.

MCPHIE, P. (1980). Kinetic studies on the unfolding and refolding of pepsinogen in urea. *Journal of Biological Chemistry* **255**, 4048–4052.

MALKIN, L.I. AND RICH, A. (1967). Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *Journal of Molecular Biology* **26**, 329–346.

MARSTON, F.A.O. (1986). The purification of eukaryotic polypeptides synthesised in *Escherichia coli*. *Biochemical Journal* **240**, 1–12.

MARSTON, F.A.O., LOWE, P.A., DOEL, M.T., SCHOEMAKER, J.M., WHITE, S. AND ANGAL, S. (1984). Purification of calf prochymosin (prorenin) synthesised in *Escherichia coli* using a chromatographic method. *Bio/Techology* **2**, 800–804.

MARSTON, F.A.O., ANGAL, A., LOWE, P.A., CHAN, M. AND HILL, C.R. (1988). Scale-up of the recovery and reactivation of recombinant proteins. *Biochemical Society Transactions* **16**, 112–115.

MAS, M.T., CHEN, C.Y., HITZEMAN, R.A. AND RIGGS, A.D. (1986). Active human–yeast chimeric phosphoglycerate kinases engineered by domain interchange. *Science* **233**, 788–790.

MENDOZA, J., LORIMER, G.H. AND HOROWITZ, P.M. (1991). Intermediates in the chaperonin-assisted refolding of rhodanese are trapped at low temperature and show a small stoichiometry. *Journal of Biological Chemistry* **266**, 16973–16976.

MITCHINSON, C. AND PAIN, R.H. (1985). Effects of sulphate and urea on the stability and reversible unfolding of β-lactamase from *Staphylococcus aureus*. *Journal of Molecular Biology* **184**, 331–342.

MITRAKI. A. AND KING, J. (1989). Protein folding intermediates and inclusion body formation. *Bio/Technology* **7**, 690–697.

MITRAKI, A., BETTON, J.M., DESMADRIL, M. AND YON, J. (1987). Quasi-irreversibility in the unfolding–refolding transition of phosphoglycerate kinase induced by guanidine hydrochloride. *European Journal of Biochemistry* **163**, 29–34.

MULKERRIN, M.G. AND WETZEL, R. (1989). pH dependence of the irreversible thermal denaturation of gamma interferons. *Biochemistry* **28**, 6556–6561.

NAGAI, K., THORGERSON, H.-C. AND LUISI, B.F. (1988). Refolding and crystallographic studies of eukaryotic proteins produced in *Escherichia coli*. *Biochemistry Society Transactions* **16**, 108–110.

NALL, B.T. (1986). Native or native-like species are transient intermediates in folding of alkaline iso-2 cytochrome *c*. *Biochemistry* **25**, 2974–2978.

NILSSON, B., BERMANN-MARKS, C., HOBER, S. AND ANDERSON, S. (1989). Differential secretion in *Escherichia coli* of BPTI folding mutants. In *Prospects in Protein Engineering*, Abstract No. 40. University of Groningen, The Netherlands.

OAS, T.G. AND KIM, P.J. (1988). A model of a protein folding intermediate. *Nature* **336**, 42–48.

OLSON, K.C. (1985). *Recovery of precipitated heterologous protein from host cell culture by using buffer solution then strong denaturing solution.*. US Patent 4,518,526.

OLSON, K.C. AND PAI, R.C. (1985). *Solubilising heterologous protein in host cells by using a solution of guanidine salt or thiocyanate for denaturing.* US Patent 4,511,503.

OPITZ, U., RUDOLPH, R., JAENICKE, R., ERICSSON, L. AND NEURATH, H. (1987). Proteolytic dimers of porcine muscle lactate dehydrogenase: characterisation, folding, and reconstitution of the truncated and nicked polypeptide chain. *Biochemistry* **26**, 1399–1406.

OSTERMANN, J., HORWICH, A.L., NEUPERT, W. AND HARTL, F.-U. (1989). Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. *Nature* **341**, 125–130.

PAIN, R.H. (1977). The conformation and stability of folded globular proteins. In *Characterisation of Protein Conformation and Function* (F. Franks, Ed.), pp.19–36. Symposium Press, London.

PAIN, R.H. (1987). Protein folding for pleasure and for profit. *Trends in Biochemical Sciences* **12**, 309–312.

PAIN, R.H., LAH, T. AND TURK, V. (1985). Conformation and processing of cathepsin D. *Bioscience Reports* **5**, 957–967.

PAREKH, R.B., DWEK, R.A., EDGE, C.J. AND RADEMACHER, T.W. (1989). *N*-glycosylation and the production of recombinant glycoproteins. *Trends in Biotechnology* **7**, 117–121.

PARK, S., LUI, G., TOPPING, T.B., COVER, W.H. AND RANDALL, L.L. (1988). Modulation of folding pathways of exported proteins by the leader sequence. *Science* **239**, 1033–1035.

PARKKONEN, T., KIVIRIKKO, K.I. AND PIHLAJANIEMI, T. (1988). Molecular cloning of a multifunctional chicken protein acting as the prolyl 4-hydroxylase β-subunit, protein disulphide isomerase and a cellular thyroid-hormone-binding-protein. *Biochemical Journal* **256**, 1005–1011.

PATRONI, J.J. (1990). *Production of protein in physiologically active or native form by treating solubilised aggregate with urea or dimethyl sulphone, especially for recovery of recombinant growth hormone.* US Patent 4,966,963.

PELHAM, H.R.B. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* **46**, 959–961.

PIHLAJANIEMI, T., HELAAKOSKI, T., TASANEN, K., MYLLYLÄ, R., HUHTALA, M.-L., KOIVU, J. AND KIVIRIKKO, K.I. (1987). Molecular cloning of the β-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *EMBO Journal* **6**, 643–649.

PTITSYN, O.B. (1987). Protein folding: hypotheses and experiments. *Journal of Protein Chemistry* **6**, 273–293.

PTITSYN, O.B., PAIN, R.H., SEMISOTNOV, G.V., ZEROVNIK, E. AND RAZGULYAEV, O.I. (1990). Evidence for a molten globule as a general intermediate in protein folding. *FEBS Letters* **262**, 20–24.

RANDALL, L.L. AND HARDY, S.J.S. (1989). Unity in function in the absence of consensus in sequence: role of leader peptides in export. *Science* **243**, 1156–1159.

REZNIKOFF, W. AND GOLD, L. (1986). *Maximising gene expression*. Butterworth, Stoneham, MA, USA.

RODER, H., ELÖVE, G.A. AND ENGLANDER, S.W. (1988). Structural characterisation of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature* **335**, 700–704.

ROTH, R.A. AND PIERCE, S.B. (1987). *In vivo* cross-linking of protein disulphide isomerase to immunoglobulins. *Biochemistry* **26**, 4179–4182.

RUDOLPH, R. (1985). The use of limited proteolysis in studies on protein folding and association. *Biochemical Society Transactions* **13**, 308–311.

RUDOLPH, R., FUCHS, I. AND JAENICKE, R. (1986). Reassociation of dimeric cytoplasmic malate dehydrogenase is determined by slow and very slow folding reactions. *Biochemistry* **25**, 1662–1669.

SASSENFELD, H.M. AND BREWER, S.J. (1984). A polypeptide fusion designed for the purification of recombinant proteins. *Bio/Technology* **2**, 76–81.

SCHEIN, C.H. (1989). Production of soluble recombinant proteins in bacteria. *Bio/Technology* **7**, 1141–1149.

SCHEIN, C.H. (1990). Solubility as a function of protein structure and solvent components. *Bio/Technology* **8**, 308–317.

SCHEIN, C.H. AND NOTEBORN, M.H.M. (1988). Formation of soluble recombinant proteins in *Escherichia coli* is favoured by lower growth temperature. *Bio/Technology* **6**, 291–294.

SCHOEMAKER, J.M., BRASNETT, A.H. AND MARSTON, F.A.O. (1985). Examination of calf prochymosin accumulation in *Escherichia coli*: disulphide linkages are a structural component of prochymosin containing inclusion bodies. *EMBO Journal* **4**, 775–780.

SCHONER, R.G., ELLIS, L.F. AND SCHONER, B.E. (1985). Isolation and purification of protein granules from *Escherichia coli* cells overproducing bovine growth hormone. *Bio/Technology* **3**, 151–154.

SCHULKE, N. AND SCHMID, F.X. (1988). Effect of glycosylation on the mechanism of renaturation of invertase from yeast. *Journal of Biological Chemistry* **263**, 8832–8837.

SEELEY, J.E. AND YANG, R. (1988). *Method for promoting intramolecular disulphide bond formation in recombinant proteins contained in a denaturant solution*. European Patent 0,312,358.

SEMISOTNOV, G.V., RODIONOVA, N.A., KUTYSHENKO, V.P., EBERT, B., BLANK, J. AND PTITSYN, O.B. (1987). Sequential mechanism of folding of carbonic anhydrase B. *FEBS Letters* **224**, 9–13.

SHARMA, S.K. (1986). Recovery of soluble human renin from inclusion bodies produced in recombinant *Escherichia coli*. *Journal of Biotechnology* **4**, 119–124.

SHARMA, S.K., EVANS, D.B., TOMICH, C.S.-C., CORNETTE, J.C. AND ULRICH, R.G. (1987). Folding and activation of recombinant human prorenin. *Biotechnological Applications in Biochemistry* **9**, 181–193.

SHIRE, S.J., BÖCK, L., OGEZ, J., BUILDER, S., KLEID, D. AND MOORE, D.M. (1984). Purification and immunogenicity of fusion VP1 protein of foot and mouth disease virus. *Biochemistry* **23**, 6474–6480.

SHOEMAKER, K.R., KIM, P.S., YORK, E.J., STEWART, J.M. AND BALDWIN, R.L. (1987). Tests of the helix dipole model for stabilisation of α-helices. *Nature* **326**, 536–567.

SHOEMAKER, K.R., FAIRMAN, R., SCHULTZ, R.A., ROBERTSON, A.D., YORK, E.J., STEWART, J.M. AND BALDWIN, R.L. (1990). Side-chain interactions in the C-peptide helix phe8 ...his^{+}12. *Biopolymers* **29**, 1–11.

SILEN, J.L. AND AGARD, D.A. (1989). The α-lytic protease pro-region does not require a physical linkage to activate the protease domain *in vivo*. *Nature* **341**, 462–464.

SIZMANN, D., KEILMANN, C. AND BÖCK, A. (1990). Primary structure requirements for the *in vivo* maturation of penicillin acylase from *E. coli* ATCC 11105. *European Journal of Biochemistry* **192**, 143–151.

SMITH, A.T., SANTANA, N., DACEY, S., EDWARDS, M., BRAY, R.C., THORNLEY, R.N.F. AND BURKE, J.F. (1990). Expression of a synthetic gene for horseradish peroxidase *c* in *Escherichia coli* and folding and activation of the recombinant enzyme with Ca^{2+} and haem. *Journal of Biological Chemistry* **265**, 13335–13343.

SMITH, M.C., COOK, J.A., FURMAN, T.C. AND OCCOLOWITZ, J.L. (1989). Structure and activity dependence of recombinant human insulin like growth factor II on disulphide bond pairing. *Journal of Biological Chemistry* **264**, 9314–9321.

STATES, D.J., CREIGHTON, T.E., DOBSON, C.M. AND KARPLUS, M. (1987). Conformations of intermediates in the folding of pancreatic trypsin inhibitor. *Journal of Molecular Biology* **195**, 731–739.

STRANSBERG, L., VEIDE, A. AND ENFORS, S.-O. (1987). Production of a hybrid protein Staphylococcal protein A/*E. coli* β-galactosidase with *Escherichia coli*. *Journal of Biotechnology* **6**, 225–238.

TAKAGI, H., MORINAGI, Y., TSUCHIYA, M., IKEMURA, H. AND INOUYE, M. (1988). Control of folding of proteins secreted by a high expression secretion vector, pIN-III-*omp*A: 16-fold increase in production of active subtilisin E in *Escherichia coli*. *Bio/Technology* **6**, 948–950.

TAKAHASHI, N., HAYANO, T. AND SUZUKI, M. (1989). Peptidyl-prolyl *cis–trans* isomerase is the cyclosporin A-binding cyclophilin. *Nature* **337**, 473–475.

TANDON, S. AND HOROWITZ, P. (1986). Detergent-assisted refolding of guanidinium chloride-denatured rhodanese. *Journal of Biological Chemistry* **261**, 15615–15618.

TANDON, S. AND HOROWITZ, P. (1989). Reversible folding of rhodanese. *Journal of Biological Chemistry* **264**, 9859–9866.

TANG, J.-G. AND TSOU, C.-L. (1990). The insulin A and B chains contain structural information for the formation of the native molecule. *Biochemical Journal* **268**, 429–435.

TIMASHEFF, S.N. AND ARAKAWA, T. (1989). Stabilisation of protein structure by solvents. In *Protein Structure; A Practical Approach* (T.E. Creighton, Ed.), pp.331–345. IRL Press, Oxford.

TSUNENAGA, M., GOTO, Y., KAWATA, Y. AND HAMAGUCHI, K. (1987). Unfolding and refolding of a type κ immunoglobulin light chain and its variable and constant fragments. *Biochemistry* **26**, 6044–6051.

UDGAONKAR, J. AND BALDWIN, R.L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335**, 694–699.

UDGAONKAR, J. AND BALDWIN, R.L. (1990). Early folding intermediate of ribonuclease A. *Proceedings of the National Academy of Sciences USA* **87**, 8197–8201.

WEIR, M.P. AND SPARKS, J. (1987). Purification and renaturation of interleukin-2. *Biochemical Journal* **245**, 85–91.

WEISS, J.B., RAY, P.H. AND BASSFORD, P.J., JR (1988). Purified SecB protein of *Escherichia coli* retards folding and promotes membrane translocation of the maltose-binding protein *in vitro*. *Proceedings of the National Academy of Sciences USA* **85**, 8978–8982.

WEISSMAN, J.S. AND KIM, P.S. (1991). Reexamination of the folding of BPTI:

predominance of native intermediates. *Science* **253**, 1386–1393.

WHITE, T.B., BERGET, P.B. AND NALL, B.T. (1987). Changes in conformation and slow refolding kinetics in mutant *iso*-2-cytochrome *c* with replacement of a conserved proline residue. *Biochemistry* **26**, 4358–4366.

WIECH, H., STUART, R. AND ZIMMERMAN, R. (1990). Role of cytosolic factors in the transport of proteins across membranes. *Seminars in Cell Biology* **1**, 55–63.

WINGFIELD, P.T., PAIN, R.H. AND CRAIG, S. (1987). Tumour necrosis factor is a compact trimer. *FEBS Letters* **211**, 179–184.

WINGFIELD, P.T., GRABER, P., ROSE, K., SIMONA, M.G. AND HUGHES, G.J. (1987a). Chromatofocussing of N-terminally processed forms of proteins. Isolation and characterisation of two forms of interleukin-1β and bovine growth hormone. *Journal of Chromatography* **387**, 291–300.

WINGFIELD, P.T., MATTALIANO, R.J., MACDONALD, H.R., CRAIG, S., CLORE, G.M., GRONENBORN, A.M. AND SCHMEISSNER, U. (1987b). Recombinant derived interleukin-1α stabilised against specific deamidation. *Protein Engineering* **1**, 413–417.

WINGFIELD, P.T., GRABER, P., MOONEN, P., CRAIG, S. AND PAIN, R.H. (1988). The conformation and stability of recombinant-derived granulocyte-macrophage colony stimulating factors. *European Journal of Biochemistry* **173**, 65–72.

WINGFIELD, P.T., GRABER, P., TURCATTI, G., MOVVA, N.R., PELLETIER, M., CRAIG, S., ROSE, K. AND MILLER, C.G. (1989). Purification and characterisation of a methionine specific aminopeptidase from *Salmonella typhimurium*. *European Journal of Biochemistry* **180**, 23–32.

WITTRUP, K.D., MANN, M.B., FENTON, D.M., TSAI, L.B. AND BAILEY, J.E. (1988). Single-cell light scatter as a probe of refractile body formation in recombinant *Escherichia coli*. *Bio/Technology* **6**, 423–426.

WOOD, L.C., WHITE, T.B., RAMDAS, L. AND NALL, B.T. (1988). Replacement of a conserved proline eliminates the absorbance-detected slow folding phase of *iso*-2-cytochrome *c*. *Biochemistry* **27**, 8562–8568.

WRIGHT, P.E., DYSON, H.J. AND LERNER, R.A. (1988). Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. *Biochemistry* **27**, 7167–7175.

YAMADA, T., KATO, K., KAWAHARA, K. AND NISHIMURA, O. (1986). Separation of recombinant human IL-2 and methionyl IL-2 produced in *Escherichia coli*. *Biochemical and Biophysical Research Communications* **135**, 837–843.

ZETTLMEISSL, G., RUDOLPH, R. AND JAENICKE, R. (1979). Reconstitution of lactic dehydrogenase. Noncovalent aggregation *vs*. reactivation. 1. Physical properties and kinetics of aggregation. *Biochemistry* **18**, 5567–5571.

ZETTLMEISSL, G., RUDOLPH, R. AND JAENICKE, R. (1982). Rate-determining folding and association reactions of the reconstitution pathway of porcine skeletal muscle lactic dehydrogenase after denaturation by guanidine hydrochloride. *Biochemistry* **21**, 3946–3950.

ZHU, X., OHTA, Y., JORDAN, F. AND INOUYE, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* **339**, 483–484.