PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

What does Protein Refolding in vitro Tell us about Protein Folding in the Cell? [and Discussion]

R. Jaenicke, A. R. Clarke, F.-U. Hartl, G. H. Lorimer, R. J. Ellis and P. Viitanen

Phil. Trans. R. Soc. Lond. B 1993 **339**, 287-295
doi: 10.1098/rstb.1993.0027

What does protein refolding *in vitro* tell us about protein folding in the cell?

R. JAENICKE

Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, W-8400 Regensburg, F.R.G.

SUMMARY

The classical *in vitro* denaturation–renaturation studies by Anson, Anfinsen, Neurath, Pauling and others clearly suggested that the primary structure of proteins determines all higher levels of protein structure. Protein folding in the cell is inaccessible to a detailed analysis of its kinetic mechanism. There are obvious differences: nascent proteins acquire their native structure co- and post-translationally, with half-times in the minutes range, whereas refolding starts from the complete polypeptide chain, with rates varying from seconds to days. In the cell, accessory proteins are involved in regulating the rate of folding and association. Their role can be analysed both *in vivo*, by mutant studies, or by coexpression together with recombinant model proteins, and *in vitro*, by folding experiments in the absence and in the presence of 'foldases' and molecular chaperones, with the following general results: (i) folding is a sequential process involving native-like structural elements and a 'collapsed state' as early intermediates; (ii) the major side-reaction is caused by 'kinetic partitioning' between correct folding and wrong aggregation; (iii) rate-determining steps may be assisted by protein disulphide isomerase, peptidyl prolyl-*cis-trans*-isomerase, and molecular chaperones; and (iv) extrinsic factors, not encoded in the amino acid sequence, may be of crucial importance.

1. INTRODUCTION

The statement that proteins acquire their spatial structure autonomously in a spontaneous reaction has been an accepted dogma for more than half a century (Anson 1945). Experimental proof came from the observation that both folding *in vivo* and refolding *in vitro* lead to the 'native' state. Considering the different starting conditions of the two processes, this is not a trivial coincidence, as folding of the nascent polypeptide chain occurs in a 'vectorial' fashion (from the N- to the C-terminal end), whereas refolding starts from the complete molecule. Experience from a great variety of water-soluble 'globular' proteins has recently been supplemented by successful attempts to renature integral membrane proteins (Popot *et al.* 1987; Surrey & Jähnig 1992). Thus, the one-to-one relationship of the amino acid sequence and its corresponding three-dimensional structure can be generalized for all proteins, independent of their compartmentation. In this context, the two categories, soluble proteins and membrane proteins, clearly indicate that the environment needs consideration as the second necessary requirement in defining a specific three-dimensional structure on the basis of a given amino acid sequence. Recombinant protein techniques (varying the guest–host relation) and investigations of protein stability within the range of biologically relevant solvent parameters may be applied to determine the constraints of external conditions on protein structure, function and energetics (Jaenicke 1991*a*).

2. FOLDING MECHANISMS DEDUCED FROM DENATURATION–RENATURATION EXPERIMENTS

There have been occasional reports of proteins which (for unknown reasons) are inaccessible to renaturation. For most proteins, optimization of *in vitro* reconstitution has been shown to reach high yields (Jaenicke 1987). In comparing these results, it is commonly assumed that *in vivo* structure formation yields 100% native protein. However, a number of observations clearly indicate that there is misfolding and misassembly in the cell (Hurtley & Helenius 1989; Pelham 1989). Obviously, protein secretion or translocation through the endoplasmic reticulum (ER) and subsequent specific degradation provide an inherent quality control, leading to the apparent yield of 100%. Under unbalanced physiological conditions, *in vivo* folding (like *in vitro* reconstitution) gives rise to wrong conformers which may reduce the yield considerably. The fact that proteins, due to their low free energies of stabilization, are basically close to their denaturation transition implies that the native state *in vivo* may occupy a whole set of substates where 'misfits' are continuously removed by proteolysis (Jaenicke 1991*b*). Thus, the yield of (re-)folding does not represent a significant difference between structure formation *in vivo* and *in vitro*.

The fact that directionality of protein synthesis obviously does not determine the native three-dimensional structure of a given protein molecule may be caused by the general mechanism of protein folding,

Phil. Trans. R. Soc. Lond. B (1993) **339,** 287–295
Printed in Great Britain

which has been described as 'folding by parts' (Wetlaufer 1981). Merging and docking of local structures contribute increments of stability, 'guiding' the folding polypeptide chain to its native state. Evidence supporting this 'hierarchy of self-organization' first came from systematic studies of the stability of protein fragments, e.g. from thermolysin (Jaenicke 1991*c*). Domain recognition and cumulative domain stabilization provide a second argument supporting hierarchical condensation (Jaenicke 1991*a*). Further evidence in favour of local structures as 'nucleation sites' for protein folding came from studies of the stability of non-random conformations of polypeptides with chain lengths below 20 residues (Scholtz & Baldwin 1992).

Such local structures (as transients in rapid equilibrium with the fully randomized chain) may be assumed in both the nascent as in the refolding polypeptide chain. For small proteins such as ribonuclease (RNase), the 'unfolded state' seems still to retain residual structure. It is tempting to assume that these locked structures could be initiating steps in protein folding. Although such structures would be only marginally stable, they would efficiently reduce the conformational space, thus directing (and enhancing) the folding reaction. Clearly, the role of such 'initiation sites' must be restricted to early folding steps, prior to the formation of well-populated intermediates and much before the rate-limiting step(s). A number of experimental observations seems to support the idea that α-helices, β-turns, and hydrophobic clusters are 'seeds' of protein self-organization (Wright *et al.* 1988; Yu & King 1988). However, in the case of bovine pancreatic trypsin inhibitor (BPTI) and RNase, Creighton (1988) has pointed out that non-random conformations in the unfolded or nascent proteins are insignificant for the kinetics of folding. The intrinsically stable α-helix at the N-terminus of RNase A cannot serve as an initiation site, because it is incorporated into the folded conformation only as the last detectable step in the overall folding process (Brems & Baldwin 1984). Thus, folding models involving initiation by local structural elements and subsequent modular assembly, are still hypothetical and must be taken with care.

In the given context, one may ask whether the N- and C-terminal stretches of the polypeptide chain are of special significance: are both indispensable for proper folding and stability, or does the vectorial nature of cotranslational folding put higher weight on the N-terminus? Taking again RNase and LDH as examples, it has been shown that the N-termini of both polypeptide chains can be cleaved without altering the gross structure of the proteins; however, their stability is drastically reduced. On the other hand, after chopping off the C-terminus, reactivation-reoxidation of RNase is blocked, whereas LDH can still be reconstituted (Jaenicke 1988, 1991*a*).

Which part of the peptide backbone is indispensable for correct folding cannot be predicted. There are numerous examples, from small proteins to large multimeric systems, where circular permutation of parts of the sequence, fragmentation, chain extension, derivatization, joining of subunits, or hybrid formation by peptide interchange, have shown that tertiary structure formation and subunit assembly may tolerate a wide range of variations in either sequence or chain connectivities (Jaenicke 1987; Urfer & Kirschner 1992). Obviously, certain core regions determine the overall topology, while 'peripheral parts' of the polypeptide chain may be altered or even lacking. Where modified proteins have been produced by recombinant DNA techniques, activity has often been found in both fractions, i.e. supernatant and inclusion bodies. Subsequent 'renativation' to the authentic protein clearly suggests that the process of structure formation *in vitro* can differ only quantitatively from the process in the cell.

The *in vivo* versus *in vitro* issue refers mainly to the kinetics of protein folding. It does not affect the generally accepted view that the structure of proteins is fully determined by their amino acid sequence. Thus, the question arises, why does folding in the cell take something of the order of minutes, whereas *in vitro* refolding varies from seconds to days or more, and what are the implications of these kinetic differences with respect to the mechanisms of protein folding within and without the cell?

3. CATALYSIS

The hypothesis that folding could be catalysed goes back to Anfinsen's early redox experiments. After being ignored for almost a generation, the idea resurfaced again after a number of 'accessory proteins' were found to catalyse known rate-limiting steps of protein folding. Because such accessory proteins are ubiquitous and abundant in all cells, it was suggested that protein self-organization in general is either catalysed or mediated by 'foldases' or 'molecular chaperones'. Evidence supporting this view came from the observation that in the case of the cell inventory of *Escherichia coli* more than half of the cellular proteins form binary complexes with GroEL (chaperonin 60) during refolding (Viitanen *et al.* 1992). Regarding possible evolutionary advantages of assisted instead of spontaneous folding, kinetic partitioning between correct structure formation on one hand, and misfolding or misassembly on the other might be a reasonable hypothesis. As in the case of proof-reading or repair processes at the level of DNA, mechanisms to correct errors in protein folding would make the cell less susceptible to folding defects attributable, for example, to the marginal intrinsic stability of proteins.

(*a*) *Protein disulphide isomerase*

In connection with the catalysis of the disulphide exchange reaction, the corresponding 'shuffling enzyme', protein disulphide isomerase (PDI), has been characterized in detail. As one would predict for a catalyst involved in protein folding and protein secretion, the enzyme is localized in the lumen of the ER where it may approach concentrations in the millimolar range. Its biological significance is difficult to prove, since depletion experiments in dog pancreas

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

cells may be perturbed by the redox composition of the glutathione buffer in the presence of PDI (Bulleid & Freedman 1988; Noiva & Lennarz 1992). However, there is evidence proving a clear correlation between PDI activity and the synthesis of cystine-containing proteins, thus strongly suggesting the biological importance of the enzyme.

The catalytic mechanism of PDI and the three-dimensional structure of its homologue, thioredoxin, have been determined. Making up ca. 10% of the total protein within the lumen of the ER, its poor catalytic efficiency is compensated by its abundance (Hawkins & Freedman 1991; Hawkins *et al.* 1991*a*,*b*). Apart from its SH–SS shuffling activity, PDI represents part of a multifunctional 'machinery' involved in proline hydroxylation, as well as several transfer reactions (Noiva & Lennarz 1992). PDI contains thioredoxin-like sequences (WCGPCK) and additional cysteine residues; correspondingly, it is inactivated by chemical modification of its thiol groups. Obviously the reduced form of the enzyme is involved in the catalytic function (Lyles & Gilbert 1991; Hawkins & Freedman 1991). Interestingly, the enzyme shares the KDEL C-terminal sequence with molecular chaperone proteins such as BiP and Grp 94, supporting the view that it may be involved in selectively targeting or retaining proteins in the ER. This conclusion may be correlated with the fact that disulphide-containing proteins are commonly extracellular; they are synthesized on membrane-bound ribosomes and concurrently translocated into the lumen of the ER, to be finally secreted. Most extracellular proteins contain disulphide bonds. Evidence from the thiol–disulphide status of nascent polypeptide chains in cells and cell-free systems has shown that disulphide bonds are formed within domains as soon as the completed domain enters the lumen of the ER. Translocation and crossbridge formation run parallel, with protein synthesis as rate-determining step (Freedman 1991; Noiva & Lennarz 1992). In the case of prokaryotes, oxidation of thiol groups occurs in the periplasm, so that high yields of cystine-containing recombinant proteins may become accessible by targeting the guest protein into the periplasmic space; coexpression of PDI together with extracellular proteins may enhance the yield of recombinant proteins considerably (R. Glockshuber, unpublished results).

In vitro, the disulphide exchange reaction has been widely used to unravel the folding pathway of small monomeric single-domain proteins such as BPTI and RNase A (Creighton *et al.* 1992). Compared with the PDI-catalysed reaction, even under optimum conditions, thiolate reoxidation is found to be slower by about one order of magnitude. In the present context, the controversial sequence of cross-bridges formed along the folding pathway is not helpful, as there is no way to follow the reaction *in vivo* at a structural or time resolution of comparable accuracy. What is important is that the enzyme does not determine either the nature of the intermediates, or the sequence in which they appear or disappear. Starting from the partially oxidized protein with non-native cystine bonds, PDI catalyses the reshuffling as it does with the reduced enzyme in the presence of oxidized glutathione; the folding mechanism remains unaltered.

There seem to be distinct variations in the redox properties of cysteine, depending on the local environment. Considering the high excess of reduced glutathione (GSH) in blood plasma and in the cytosol of hepatic cells, one might wonder whether disulphide bonds are formed at all under the reducing conditions *in vivo*. Comparative reactivation experiments under physiological GSH-GSSG and pH conditions have shown that typical SH-proteins and proteins containing disulphide bridges exhibit characteristic differences in their renaturation behaviour: at cellular levels of GSH and GSSG, RNase is fully oxidized, whereas the SH groups in lactate and malate dehydrogenase remain in their reduced state (Rudolph & Fuchs 1983). Whether the local environment of SH groups affects their redox properties, or whether steric hindrance keeps thiol groups apart from one other, is unclear. The folding rates are only insignificantly influenced by the GSH-GSSG ratio; they differ from the rates of *in vivo* folding during translation by at least two orders of magnitude so that one may conclude that the *in vivo* oxidation proceeds via a different pathway (perhaps with the help of a smaller oxidant). In the case of cystine-containing proteins, disulphide bonds are commonly formed during translation.

In summarizing the PDI results, it is obvious that the *in vitro* folding experiments are in accordance with *in vivo* observations. The unique contribution of *in vitro* studies is to the structure–function relationship and the mechanism of the shuffling enzyme and its complex 'micro-compartmentation'. The corresponding machinery marks the limit of molecular analysis and new techniques have to be developed to supplement the molecular analysis at the cellular level. From the methodological point of view, the application of the SH-SS exchange reaction served to unravel the concept of domain folding and the detailed folding pathway of small single-domain model proteins. Whether the sequential BPTI shuffling mechanism reflects the folding process in the cell remains to be shown.

(*b*) *Peptidyl prolyl cis-trans isomerase*

Rotamases as catalysts involved in the isomerization of the peptide bond are the second class of 'foldases'. The best-known example, peptidyl prolyl *cis-trans* isomerase (PPIase), has clearly been shown to enhance the folding of a variety of proteins with 'essential *cis*-proline residues' (Jaenicke 1987; Schmid *et al.* 1992). Translation yields the all-*trans* configuration of the peptide chain. In cases where the final structure of a protein contains *cis*- residues, the high activation energy of *trans-cis* isomerization will give rise to slow steps on the folding pathway which may be rate-limiting for the overall reaction.

The mechanism and sequence specificity of PPIases are still unresolved. This situation will change in the near future as the three-dimensional structure has recently been solved to high resolution (cf. Lorimer 1992). The fact that proline isomerases are ubiquitous

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

from bacteria to mammals, and that many proteins contain *cis* X-Pro bonds in their native state, suggests that the enzyme plays a biologically important role. Whether the enzyme catalyses protein folding by kinetic competition with off-pathway side-reactions cannot be stated with certainty because PPIases also serve other biological functions, e.g. as immunophilins (Rosen & Schreiber 1992).

As a result of the physicochemical and enzymological characterization of PPIase, it has become clear that proline *cis-trans* isomerization is responsible for the slow folding steps observed in many proteins containing *cis* X-Pro bonds (Schmid *et al.* 1992). Comparing the refolding kinetics of proteins with their proline content, it becomes clear that extra proline residues may or may not contribute to slow kinetic phases. Depending on the local environment, the activation energy of *cis-trans* isomerization for a specific sequence may be anomalously low, or local constraints may trap certain portions of the polypeptide chain in their native format, even in the denatured state (cf. Jaenicke 1987).

Obviously, none of the previous conclusions could be drawn from *in vivo* experiments. However, it is still unresolved whether proline isomerization is relevant to *in vivo* protein folding (Noiva & Lennarz 1992). If so, the following mechanism may be assumed: the nascent all-*trans* polypeptide chain collapses rapidly to form a compact state with elements of secondary structure and sufficient stability to expel water from its hydrophobic interior; subsequent slow and very slow steps lead to the native-like state which finally undergoes *trans*→*cis* isomerization of those prolines that are *cis* in the native state, thus forming the most stable conformation. In considering this mechanism, extrapolation from model peptides suggests that protein folding should be much slower than estimates deduced, e.g. from enzyme induction *in vivo*. There are three possible explanations as to why proline isomerization does not limit *in vivo* protein folding: (i) *trans* prolines may be trapped in their initial configuration; (ii) the occurrence of proline residues in solvent-exposed turns causes their state of isomerization to be of minor functional significance; and (iii) constraints of the chain conformation may greatly decrease the energy barrier of proline isomerization, thus speeding up the reaction. These arguments do not consider the idea that PPIases may operate at the site of protein translation or translocation, thus enhancing the *trans*→*cis* isomerization at well-defined sites depending on the sequence specificity of the enzyme. This hypothesis illustrates the way in which *in vivo* and *in vitro* folding studies might be tied together in order to solve relevant questions of protein self-organization.

4. MOLECULAR CHAPERONES

Molecular chaperones are ubiquitous in all cells and cell compartments. They include all four highly conserved families of soluble proteins which are generally expressed constitutively, but may become abundant after environmental stress. Molecular chaperones 'mediate the correct assembly of other polypeptides but are not components of the functional assembled structures; their proposed function is to assist polypeptides to self-assemble by inhibiting alternative assembly pathways that produce nonfunctional structures' (Ellis & van der Vies 1991). By implying multiple turnovers, chaperone action contains one element of catalysis, but there has been no clear-cut evidence so far proving chaperone-induced rate enhancement through stabilization of the transition state of any step on the pathway of folding and association. In many cases, chaperones are found to decelerate protein folding *in vitro* rather than accelerating it. The question how interactions between such chaperones and their target polypeptide chains simultaneously inhibit incorrect but promote correct protein interactions, is difficult to rationalize, especially faced with the 'promiscuity' of some molecular chaperones (Viitanen *et al.* 1992). For cartoonists the problem is easy to solve, because spheres and squares have different interacting surfaces that clearly suggest varying association constants and transconformation reactions ranging from 'unfolded to approximately linear conformation' to the 'molten globule state'.

There is ample evidence that targeting of nascent polypeptide chains from the ribosome to their final destination, as well as protein assembly or organelle biogenesis and other cellular processes, involve the transient exposure of interactive protein surfaces to the intracellular environment. At the same time, it is well-established that in the process of folding and association of oligomeric proteins wrong aggregation represents a major side reaction (Jaenicke 1987). In both cases, molecular chaperones may provide unspecific complementary binding sites, with affinities proportional to the structural difference between the transient and the native state of the protein. Comparing this model with the action of PDI and PPIases, mechanisms different from conventional enzyme catalysis must be involved. The fact that in some, but not all, cases ATP and additional accessory proteins are required to release the substrate protein proves that ATPase activity is not always involved in chaperone action. The evidence from reconstitution experiments using various model substrates refers mainly to the structure–function relationship and to the specificity of chaperone–target interactions. The following paragraphs give only a selection of *in vitro* studies which seem promising in answering central biological questions.

(i) *Kinetic partitioning*

Upon folding of large polypeptide chains, molecular chaperones keep misfolded or partially folded protein molecules from aggregation, allowing kinetic partitioning between folding, association and wrong aggregation. In the case of the GroE system, with citrate synthase as target protein, the chaperone mediates correct folding rather than converting incorrect structures or aggregates back to the native state. In the presence of ATP and GroES, the target protein is released (Buchner *et al.* 1991). With rubisco as a model, the ATPase reaction is found to be K^+

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

dependent, with GroES as the mediator (Viitanen *et al.* 1990).

(ii) *GroEL as 'heat shock protein' suppresses thermal aggregation*

The *in vivo* 'heat shock' effect of molecular chaperones can be mimicked *in vitro* by suppressing the aggregation of thermally denatured proteins in the presence of the chaperone. In the case of α-glucosidase, GroEL has been shown to bind the unfolded polypeptide chain. At room temperature, this chain is released again by MgATP (plus GroES), leading back to the native state; however, at high temperature, the chain aggregates much more rapidly than native α-glucosidase, proving that the substrate protein is released in a conformation different from the native functional state (Höll-Neugebauer *et al.* 1991). In similar experiments, Martin *et al.* (1992) were able to show that mitochondrial heat shock protein 60 (hsp 60) forms complexes with a variety of proteins of the mitochondrial matrix, and in this way prevents irreversible aggregation. The 'unscrambling' function of molecular chaperones in terms of the capacity to promote the disassembly of aggregated proteins is still controversial. Hypothetically, cycles of binding of a heat shock protein to interactive surfaces exposed by the stress could be followed by ATP-mediated release, in this way triggering a conformational drift toward the native state of the substrate protein. From *in vitro* experiments, the GroE system has clearly been shown to be unable to dissolve aggregates. Successful 'resurrection' of heat-denatured RNA polymerase has been reported for DnaK from *Escherichia coli*, while a clathrin-uncoating activity has been found for hsc 70. In the latter case, 'stressed protein' is renatured, whereas 'irreversibly denatured polypeptide chains' are subjected to degradation. Apparently, heat shock proteins of the hsp 70 family form larger complexes connected with general 'editing functions' (Hartl *et al.* 1992; Langer *et al.* 1992).

(iii) *Molecular chaperone-target interaction*

The mode of binding of polypeptides to molecular chaperones is still unknown, since the three-dimensional structure of binary chaperone-protein complexes has yet to be solved. PapD (a protein that mediates the assembly of pili in *Escherichia coli*) is the only protein for which a crystal structure is presently available; the ATPase-containing proteolytic fragment of the 70 kDa heat shock cognate protein (hsc 70) lacks the domain involved in target recognition (cf. Lorimer 1992). PapD seems to expose an anomalous hydrophobic surface; however, high resolution X-ray analysis of the chaperone–substrate interface would be required to propose a clear-cut mechanism with any degree of confidence. For the time being, model studies are the only way to gain insight into the type of interactions. Rhodanese has been shown to be inaccessible to reactivation at room temperature except in the presence of detergents or urea (Mendoza *et al.* 1991*a*). GroEL (chaperonin 60) can replace the denaturants, slowing down the rate of reconstitution, but at the same time drastically increasing the yield. Obviously, the chaperone does not catalyse the renaturation reaction; rather it prevents aggregation of intermediates with exposed hydrophobic surfaces (Mendoza *et al.* 1991*b*). That folding may occur while the target protein is bound to the surface of the chaperone has its parallel in the reactivation of enzymes bound to solid matrices (Gottschalk & Jaenicke 1991). Mechanistically, the doughnut structure of GroEL might suggest threading or channeling of the polypeptide chain through a pore, as in the process of translation on the ribosome; however, there is no sound basis for any kind of model.

(iv) *Conformation of the target protein*

Attempts to characterize the conformational state of the substrate protein bound to the 'active site' of its chaperone have made use of spectroscopic methods, limited proteolysis and functional assays (e.g. enzymic activity, ELISA), mainly with GroEL as molecular chaperone, and model proteins with a-helical, all-β, 'molten globule' or random coil conformations as targets. Landry & Gierasch (1991), using NMR and a peptide corresponding to the N-terminal α-helix of rhodanese, were able to show that the target peptide adopts the structure of an amphipathic α-helix in the process of binding to the chaperone, whereas free in aqueous solution, its conformation is random. This result supports the idea that the hydrophobic surface of the helix serves as an anchor in the 'binding pocket' of the chaperone. However, this result cannot be generalized since all-β structures, such as antibody fragments, also interact with GroEL (Schmidt & Buchner 1992). van der Vies *et al.* (1992) studied rubisco from *Rhodospirillum rubrum*; the enzyme forms a dimer with amino acid residues of both subunits contributing to the active sites. Denaturation–renaturation at temperatures above 10°C, and in the absence of GroEL, leads to irreversible aggregation, caused by the kinetic competition of folding and association frequently observed for oligomeric systems (Jaenicke 1987). The commitment to aggregation is rapid and proceeds until the concentration of intermediates has fallen to a 'critical concentration'. In the presence of GroEL, interaction with the chaperone suppresses aggregation. The arrest of the spontaneous folding reaction involves the binary complex between GroEL and the monomeric intermediate, with its secondary structure partially restored, but its tertiary structure still strongly distorted. Neither the dimeric precursor, nor the native enzyme form stable GroE complexes, in accordance with the idea that chaperone interaction involves hydrophobic parts of the target protein which are buried in the native state.

(v) *GroEL and the role of GroES and ATP*

The assumption that folding takes place while the target protein is bound to the surface of GroEL is confirmed by competition experiments using rhodanese as substrate and casein as competitive ligand (Martin *et al.* 1991). Under renaturation conditions where rhodanese undergoes aggregation rather than reactivation, the function of the chaperone depends on the sequence in which the components of the GroE system are added. Recovery of rhodanese activity is

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

observed only if GroEL is preincubated with the unfolded rhodanese, and subsequently released, following the sequence GroES + MgATP. The yield is drastically reduced if GroES is added after MgATP, and vanishes completely if casein is used in the preincubation step. Thus folding on the surface of GroEL requires the presence of GroES as a coupling factor; only the complete complex, with ATP hydrolysis as the final step, releases the target protein in a form which is fully committed to the native state, thus blocking 'wrong aggregation'. The role of ATP is still obscure; its stoichiometry (130 ± 20 moles MgATP hydrolysed per mole of native rhodanese; Martin *et al.* 1991) may be overestimated as a consequence of the high background ATPase activity. The ratio of target molecules per GroE complex is also still unresolved, different approaches suggesting either one (Martin *et al.* 1991) or two molecules per GroEL double ring (M. Schmidt & J. Buchner, unpublished results).

(vi) *Specificity*

Molecular chaperones are assumed to exhibit sequence specificity in polypeptide chain binding and release. Flynn *et al.* (1989) were able to show that BiP and cytoplasmic hsc 70 are able to bind short peptides in an ATP-dependent manner, albeit with high K_m values. Whether there is a unique consensus sequence or a structural motif or any colligative property of the peptide involved in the interaction is still unresolved. In the case of the SecB protein from *Escherichia coli*, the chaperone arrests folding by binding unfolded precursor proteins with high affinity, in this way keeping them in their translocation-competent form. There are no specific interactions with the leader sequence, which functions to retard folding to optimize complexation with the chaperone (Randall *et al.* 1990). Selectivity of SecB has been attributed to kinetic partitioning, in the sense that rapidly folding cytoplasmic proteins escape chaperone binding whereas proteins with leader peptides, which fold more slowly, are trapped. Binding occurs rather indiscriminately to proteins in a non-native conformation, which may still contain elements of secondary and tertiary structure (Hardy & Randall 1991).

In the case of GroEL, the lack of specificity is striking; as shown by *in vitro* binding experiments, more than 50% of the soluble proteins of *E. coli*, in their unfolded or partially folded states, form stable binary complexes with the chaperone. These complexes are long-lived enough to be isolated by gel-filtration, and are efficiently discharged on adding MgATP (Viitanen *et al.* 1992). Evidently, folding of the major part of the proteins in *Escherichia coli* takes place on the GroEL molecular chaperone; this is highly promiscuous exhibiting exceedingly low sequence specificity.

(vii) *Heat shock in thermophiles and hyperthermophiles*

Intermediates on the folding pathway of proteins show only marginal stability. Considering proteins from thermophilic organisms, kinetic partitioning under physiological conditions may yield non-native protein if molecular chaperones do not mediate proper folding. In the case of the thermophilic bacterium *Thermus thermophilus*, Taguchi *et al.* (1991) have shown that a GroEL analogue is involved in the folding of a variety of dehydrogenases at temperatures where in the absence of the chaperone, thermal denaturation outruns structure formation. Even in hyperthermophiles a heat shock response is observed (Trent *et al.* 1991; Phipps *et al.* 1991). Taking *Pyrodictium occultum* PL19 ($T_{opt} = 105°C$, $T_{max} = 110°C$) as an example, the level of ATPase activity is enhanced markedly upon shifting the temperature from 102°C to 108°C, the enzyme accumulating to almost 1% of the total soluble protein. Structure analysis reveals a quaternary structure composed of stacked discs with eightfold symmetry (Phipps *et al.* 1991). Functional analysis proves the enzyme to exhibit characteristics of a hyperthermophilic molecular chaperone (J. Buchner, R. Jaenicke, H. Sparrer, unpublished results).

5. CONCLUDING REMARKS

The previous examples illustrate the wealth of methods that have been used in recent *in vitro* studies on molecular chaperones in the attempt to understand the mechanisms involved in 'assisted structure formation', and to elucidate the traffic between cellular compartments at the molecular level. Questions such as: how do chaperones arrest nascent polypeptide chains in a translocation-competent form, or how do they protect folding intermediates from premature association and turnover, as catalysts or mediators, obviously require additional approaches at the cellular level. Presently, these approaches do not allow the various stages of protein self-organization to be analysed at a resolution comparable to the *in vitro* analysis. Whether this resolution reflects the real processes in the cell, or whether molecular biophysicists have been focusing their attention on metabiology instead of biology, remains to be seen. The fact that extrinsic factors and cell-biological effects may be involved in protein self-organization has been appreciated from the very beginning of *in vitro* folding studies. Molecular chaperones are just one class of these effectors, others being ions, cofactors, and conjugation components such as carbohydrates, lipids, nucleic acids, etc. Their effects on protein folding and protein association has been previously discussed (Jaenicke 1991*b*). In the present context, only three points need to be added: (i) in contrast to the above mentioned 'promiscuity' of some chaperone action, protein self-assembly in terms of domain and subunit recognition is highly specific. This holds for both the stable end product, and for inactive kinetic intermediates on the assembly pathway, so that in heterogeneous systems (such as the cell) no 'compartmentation' is required to guarantee correct tertiary or quaternary contacts (Jaenicke 1987). (ii) In spite of the general experience that optimum *in vitro* 'renativation' of proteins is an art closer to alchemy than chemistry, additives devised to mimic the physical characteristics of the cytosol (viscosity, pH, local concentrations of carbohydrates, etc.) do not seem to exhibit drastic effects on the folding mechanism (R. Jaenicke,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

G. Kern, V. Rehaber, unpublished results). (iii) In connection with conjugation, mutant studies on yeast invertase have shown that glycosylation enhances thermal stability, simultaneously protecting the glycoprotein from aggregation. In this respect, the carbohydrate moiety mimics the effect of a molecular chaperone; correspondingly, glycosylation inhibits the interaction with GroEL (Kern *et al.* 1992*a*,*b*). The fact that polyols enhance protein stability is an old experience: compatible solvent components have long been in use in the laboratory, and even longer in nature, where polyols as extrinsic protectants may help organisms to cope with extreme physical conditions (Jaenicke 1991*a*).

This review was written during a stay at the Physiological Institute, University of Zürich and the Fogarty International Center, NIH, Bethesda, Maryland. I thank Christian Bauer for his kind hospitality. Fruitful discussions with Dr G. Böhm, Dr J. Buchner, Dr V. Rehaber, Dr R. Rudolph, Dr F. X. Schmid, Dr M. Schmidt, Dr J. Schumann and Dr R. Seckler are gratefully acknowledged. Work done in the author's laboratory was generously supported by Grants of the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie and the BAP Program of the European Community.

This paper is dedicated to Professor John T. Edsall on the occasion of his 90th birthday

REFERENCES

Anson, M.L. 1945 Protein denaturation and the properties of protein groups. *Adv. Prot. Chem.* **2**, 361–386.

Brems, D.N. & Baldwin, R.L. 1984 Amide proton exchange used to monitor the formation of a stable α-helix by residues 3 to 13 during folding of RNase S. *J. molec. Biol.* **180**, 1141–1156.

Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F.X. & Kiefhaber, T. 1991 GroE facilitates refolding of citrate synthase by suppressing aggregation. *Biochemistry* **30**, 1586–1591.

Bulleid, N.J. & Freedman, R.B. 1988 Defective co-translational formation of disulphide bonds in PDI-deficient microsomes. *Nature, Lond.* **335**, 649–651.

Creighton, T.E. 1988 On the relevance of non-random polypeptide conformations for protein folding. *Biophys. Chem* **31**, 155–162.

Creighton, T.E., Weissman, J.S. & Kim, P.S. 1992 The disulphide folding pathway of BPTI. *Science, Wash.* **256**, 111–114.

Ellis, R.J. & van der Vies, S.M. 1991 Molecular chaperones. *A. Rev. Biochem.* **60**, 321–347.

Flynn, G.C., Chappell, T.G. & Rothman, J.E. 1989 Peptide binding and release by proteins implicated as catalysts of protein assembly. *Science, Wash.* **245**, 385–390.

Freedman, R.B. 1991 Protein disulphide isomerase: an enzyme that catalyzes protein folding in the test tube and in the cell. In *Conformations and forces in protein folding* (ed. B. R. Nall & K. A. Dill), pp. 204–214. AAAS Washington.

Gottschalk, N. & Jaenicke, R. 1991 Authenticity and reconstitution of immobilized enzymes. *Biotech. Appl. Biochem.* **14**, 324–335.

Hardy, S.J.S. & Randall, L.L. 1991 A kinetic partitioning model of selective binding of non-native proteins by SecB. *Science, Wash.* **251**, 439–443.

Hartl, F.U., Martin, J. & Neupert, W. 1992 Protein folding in the cell: The role of molecular chaperones hsp 70 and hsp 60. *A. Rev. Biophys. Biomol. Struct.* **21**, 293–322.

Hawkins, H.C. & Freedman, R.B. 1991 The reactivities and ionization properties of the active site dithiol groups of mammalian PDI. *Biochem. J.* **275**, 335–339.

Hawkins, H.C., de Nardi, M. & Freedman, R.B. 1991*a* Redox properties and cross linking of the dithiol/disulphide active sites of mammalian PDI. *Biochem. J.* **275**, 341–348.

Hawkins, H.C., Blackburn, E.C. & Freedman, R.B. 1991*b* Comparison of the activities of PDI in catalysing disulphide isomerisation in a protein substrate. *Biochem. J.* **275**, 349–353.

Höll-Neugebauer, B., Rudolph, R., Schmidt, M. & Buchner, J. 1991 Reconstitution of a heat shock effect in vitro: influence of GroE on the thermal aggregation of α-glucosidase from yeast. *Biochemistry* **30**, 11609–11614.

Hurtley, S.M. & Helenius, A. 1989 Protein oligomerization in the ER. *A. Rev. Cell. Biol.* **5**, 277–307.

Jaenicke, R. 1987 Folding and association of proteins. *Prog. Biophys. molec. Biol.* **49**, 117–237.

Jaenicke, R. 1988 Is there a code of protein folding? In *Protein structure and protein engineering*, vol. 39 (*Colloquium Mosbach*) (ed. E.-L. Winnacker & R. Huber), pp. 16–36. Berlin, Heidelberg & New York: Springer-Verlag.

Jaenicke, R. 1991*a* Protein stability and molecular adaptation to extreme conditions. *Eur. J. Biochem.* **202**, 715–728.

Jaenicke, R. 1991*b* Protein folding: local structures, domains, subunits and assemblies. *Biochemistry* **30**, 3147–3161.

Jaenicke, R. 1991*c* Protein folding. In *Applications of enzyme biotechnology* (ed. J. W. Kelly & T. O. Baldwin), pp.137–152. New York & London: Plenum Press.

Kern, G., Schülke, N., Schmid, F.X. & Jaenicke, R. 1992*a* Quaternary structure and stability of internal, external and core-glycosylated invertase from yeast. *Protein Sci.* **1**, 120–131.

Kern, G., Schmidt, M., Buchner, J. & Jaenicke, R. 1992*b* Glycosylation inhibits the interaction of invertase with the chaperone GroEL. *FEBS Lett.* **305**, 203–205.

Landry, S.J. & Gierasch, L.M. 1991 The chaperonin GroEL binds a polypeptide in an α-helical conformation. *Biochemistry* **30**, 7359–7362.

Langer, T., Lu Chi, Echols, H., Flanagan, J., Hayer, M.K. & Hartl, F.-U. 1992 Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature, Lond.* **356**, 683–689.

Lorimer, G.H. 1992 Role of assessory proteins in protein folding. *Curr. Opin. Struct. Biol.* **2**, 26–34.

Lyles, M.M. & Gilbert, H.F. 1991 Catalysis of the oxidative folding of RNaseA by PDI: Dependence of the rate on the composition of the redox buffer. *Biochemistry* **30**, 613–619.

Martin, J., Horvich, A.L. & Hartl, F.U. 1992 Role of chaperonin hsp60 in preventing protein denaturation under heat-stress. *Nature, Lond.* (In the press.)

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

Mendoza, J.A., Rogers, E., Lorimer, G.H. & Horowitz, P.M. 1991*a* Chaperonins facilitate the in vitro folding of monomeric mitochondrial rhodanese. *J. biol. Chem.* **266**, 13044–13049.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Mendoza, J.A., Rogers, E., Lorimer, G.H. & Horowitz, P.M. 1991*b* Unassisted refolding of urea-unfolded rhodanese. *Nature, Lond.* **266**, 13587–13591.
Noiva, R. & Lennarz, W.J. 1992 Protein disulphide isomerase. *J. biol. Chem.* **267**, 3553–3556.
Pelham, H.R.B. 1989 Control of protein exit from the ER. *A. Rev. Cell. Biol.* **5**, 1–23.
Phipps, B.M., Hoffmann, A., Stetter, K.O. & Baumeister, W. 1991 A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria. *EMBO J.* **10**, 1711–1722.
Popot, J.-L., Gerchman, S.-E. & Engelman, D.M. 1987 Refolding of bacteriorhodopsin in lipid bilayers. *J. molec. Biol.* **198**, 655–676.
Randall, L.L., Topping, T.B. & Hardy, S.J.S. 1990 No specific recognition of leader peptide by SecB, a chaperone involved in protein export. *Science, Wash.* **248**, 860–863.
Rehaber, V. & Jaenicke, R. 1992 Stability and reconstitution of GAPDH from the hyperthermophilic eubacterium *Thermotoga maritima*. *J. biol. Chem.* **267**, 10999–11006.
Rosen, M.K. & Schreiber, S.L. 1992 Natural products as probes of cellular function: Studies on immunophilins. *Angew. Chem. Int. Ed. (Engl.)* **31**, 384–400.
Rothman, J.E. & Schmid, S.L. 1986 Enzymatic recycling of clathrin from coated vesicles. *Cell* **46**, 5–9.
Rudolph, R. & Fuchs, I. 1983 Influence of glutathione on the reactivation of enzymes containing cysteine or cystine. *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 813–820.
Schmid, F.X., Mayr, L., Mücke, M. & Schönbrunner, E.R. 1993 Prolyl isomerases: the role in protein folding. *Adv. Protein Chem.* (In the press.)
Schmidt, M. & Buchner, J. 1992 Interaction of GroE with an all-β protein. *J. biol. Chem.* **267**, 16829–16833.
Scholtz, J.M. & Baldwin, R.L. 1992 The mechanism of α-helix formation by peptides. *A. Rev. Biophys. biomolec. Struct.* **21**, 95–118.
Surrey, T. & Jähnig, F. 1992 Refolding and oriented insertion of a membrane protein into a lipid bilayer. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7457–7461.
Taguchi, H., Konishi, J., Ishii, N. & Yoshida, M. 1991 A chaperonin from a thermophilic bacterium, *Thermus thermophilus*, that controls refolding of several thermophilic enzymes. *J. biol. Chem.* **266**, 22411–22418.
Trent, J.D., Nimmesgern, E., Wall, J.S., Hartl, F.-U. & Horwich, A.L. 1991 A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein TCP1. *Nature, Lond.* **354**, 490–493.
Urfer, R. & Kirschner, K. 1992 The importance of surface loops for stabilizing an eightfold βα barrel protein. *Protein Sci.* **1**, 31–45.
van der Vies, S.M., Viitanen, P.V., Gatenby, A.A., Lorimer, G.H. & Jaenicke, R. 1992 Conformational states of Rubisco and their interaction with chaperonin 60. *Biochemistry* **31**, 3635–3644.
Viitanen, P.V., Lubben, T.H., Reed, J., Goloubinoff, P., O'Keefe, D.P. & Lorimer, G.H. 1990 Chaperonin-facilitated refolding of Rubisco and ATP hydrolysis by chaperonin 60 are K^+ dependent. *Biochemistry* **29**, 5665–5671.
Viitanen, P.V., Gatenby, A.A. & Lorimer, G.H. 1992 Purified GroEL interacts with the nonnative states of a multitude of *E. coli* proteins. *Protein Sci.* **1**, 361–369.
Wetlaufer, D.B. 1981 Folding of protein fragments. *Adv. Protein Chem.* **34**, 91–69.
Wright, P.E., Dyson, H.J. & Lerner, R.A. 1988 Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. *Biochemistry* **27**, 7167–7175.
Yu, M.-H. & King, J. 1988 Surface amino acids as sites of ts folding mutations in the P22 tailspike protein. *J. biol. Chem.* **263**, 1424–1431.

Discussion

A. R. Clarke (*Molecular Recognition Centre, University of Bristol, U.K.*). Is Professor Jaenicke convinced that the increase in yield of a protein folding reaction observed in the presence of the chaperonins and ATP is due merely to a suppression of aggregation of sticky chains?

R. Jaenicke. The only reason for my claim is that the percentage of light scattering, which is a measure of aggregation, is an exact mirror-image of the percentage of enzymic activity.

A. R. Clarke. I work with lactate dehydrogenase from a bacillus, and find that quite large effects of the chaperonins on the yield of active enzyme can be seen at concentrations where aggregation does not occur. Therefore I am wondering whether the chaperonins have another level of activity as well as the suppression of aggregation; this may be the ATP-requiring level where energy is used to unfold misfolded chains. One would expect that the purely passive suppression of aggregation would slow down the rate of refolding, and yet when chaperonins and ATP are added the folding rate is only slightly reduced, but the yield of active enzyme is increased.

R. Jaenicke. We observe that the rate of enzyme reactivation is reduced in the presence of the chaperonins and believe that this is due to the reduction in concentration of the folding molecules since some of them are bound to the chaperonin. If ATP is now added, reactivation of the enzyme proceeds with more or less identical kinetics to that of the protein folding in the absence of the chaperonins.

F.-U. Hartl (*Memorial Sloan-Kettering Cancer Center, New York, U.S.A.*). I would like to comment on Professor Jaenicke's remark about whether a chaperone itself requires a chaperone to assist its folding. It has been shown that if the bacterial chaperonin 10 is synthesized chemically it can function in a protein refolding assay, and it has also been reported that the bacterial chaperonin 60 can spontaneously reassemble after denaturation. But it has also been shown that the mitochondrial chaperonin 60, like other proteins, requires pre-existing mitochondrial chaperonin 60 to assemble correctly *in vivo*.

R. Jaenicke. I believe that there are some differences between organisms as to whether the folding of some chaperones requires pre-existing chaperones or not.

F.-U. Hartl. I am not aware of any observation that demonstrates this.

G. H. Lorimer (*Du Pont de Nemours, Wilmington, U.S.A.*). During *in vitro* protein refolding experiments

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the protein can be trapped by adsorbing onto the glass and plastic vessels used, especially at very low protein concentrations. So it does not follow that protein that does not fold correctly is necessarily in an aggregated form.

R. Jaenicke. I agree with Dr Lorimer: we observed this phenomenon in our early experiments. One can overcome this problem to some extent by adding bovine serum albumin. Another problem when working at very low concentrations of protein is the need to wait for a long enough time before estimating the final yield of active protein. This is indispensable in case a kinetic analysis is intended.

R. J. Ellis (*Department of Biological Sciences, University of Warwick, U.K.*). What is the evidence that *in vivo* proteins fold by parts as they are being synthesized? Professor Jaenicke mentioned the example of IgG, but this may be a special case as it is cotranslationally transported and folds in the lumen of the endoplasmic reticulum where the chaperonin group of molecular chaperones have not been found.

R. Jaenicke. Examples always quoted are serum-albumin and IgG.

G. Lorimer. The *in vivo* experiments with circularly permutated proteins show that the vectorial synthesis of a polypeptide chain from the aminoterminus is not essential for the chain to fold correctly.

P. Viitanen (*Du Pont de Nemours, Wilmington, U.S.A.*). Are there any examples where the folding protein spends so much time leaving a metastable state that is on an incorrect assembly pathway that this becomes rate-determining? If chaperonins can suppress intermolecular aggregations that produce incorrect structures, can they also suppress intramolecular interactions that produce incorrect structures and thus change the rate-determining step?

R. Jaenicke. The step that commits the chain to aggregation occurs very quickly as it requires only the initial collapse of the chain to a compact non-native structure. With respect to the second part of your question, it has been observed, e.g. for Fab fragments, that the overall folding kinetics can be slowed down significantly in the presence of chaperone proteins.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES