PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

# Folding and Association Versus Misfolding and Aggregation of Proteins

Rainer Jaenicke

*Phil. Trans. R. Soc. Lond. B* 1995 **348**, 97-105
doi: 10.1098/rstb.1995.0050

# Folding and association versus misfolding and aggregation of proteins

RAINER JAENICKE

*Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, D-93040 Regensburg, Germany*

## SUMMARY

The acquisition of spatial structure in proteins may be described in terms of hierarchical condensation, with contributions of local interactions between next neighbours and the interactions between domains and subunits accumulating to create the marginal free energy of stabilization characteristic of the functional state of globular proteins. Domains represent independent folding units such that the overall kinetics divide into the sequential collapse of subdomains and domains and their merging to form the compact tertiary structure. In proceeding to oligomeric proteins, docking of subunits follows the formation of structured monomers. Thus, the overall mechanism of folding and association obeys consecutive uni–bimolecular kinetics. Beyond a limiting protein concentration, aggregation will outrun proper domain pairing and subunit association. In the cell, accessory proteins are involved in catalysis of the rate-determining steps of folding (proline isomerization and SH–SS exchange) and in the kinetic partitioning between folding and aggregation (chaperone action). The practical aspects of accessory proteins have been investigated in detail using immunotoxins and antibody fragments as examples. Additional concepts allowing off-pathway reactions in protein reconstruction to be kept to a minimum refer to pulse-dilation, reverse micelles and immobilization of polypeptide chains on matrices.

## 1. INTRODUCTION

The topic of this paper has three 'faces': one looking back into the heroic age when Anson for the first time observed 'reversible denaturation' of proteins; a second one facing industrial applications; and the third one with a vision of how kinetic partitioning between folding and misfolding and their consecutive reactions in the cell might work.

With respect to the first issue, the correlation between hierarchies of protein structure, stability and folding, and recent experimental developments allowing access to fast and/or local events in the course of protein self-organization, must be considered. Regarding the second issue, what started 15 years ago in basic research has developed to a huge extent: manual mixing of microlitre aliquots has changed into cubic metres of 6 M guanidine solutions, pumped in automated processes at industrial plants to yield several grammes of pharmaceutical product at a time. To the *in vitro* versus *in vivo* controversy the physical biochemist cannot contribute much, except the reductionist's scepticism against cartoons which make readers (and sometimes authors) believe that the cell is full of circles and triangles instead of molecules. What is required, and becomes increasingly accessible due to molecular biological techniques, is the structural and functional analysis of components involved in chaperoning and targeting the nascent polypeptide chain. In addition, experimental approaches have to be devised which allow cellular conditions to be mimicked. Significant progress has been made toward the goal of reproducing *in vivo* protein folding in the test tube. Examples include studies of folding in the presence and absence of accessory proteins, complementation experiments using transcription–translation systems and *in vitro* translation of mRNAs encoding truncated or otherwise mutated mRNAs.

## 2. HIERARCHIES OF FOLDING AND STABILITY

In the structural hierarchy of proteins, the different levels refer to folding as well as stability. Increasing packing density and release of water from hydrophobic residues provide the enthalpic and entropic increments of the free energy of stabilization which finally yield the marginal difference of the attractive and repulsive forces characteristic for the native-state stability of biological macromolecules. Typical $\Delta G_{stab}$ values are approximately 50 kJ $mol^{-1}$; this is equivalent to just a few weak interactions in a protein molecule with *ca.* 5000 atoms. The biological significance of this observation is three-fold: (i) optimization of the structure–function relation in the course of evolution is aimed at flexibility (catalysis, regulation, turnover) rather than stability; (ii) under physiological conditions, native globular proteins are generally on the borderline of denaturation; and (iii) because the native state is a state of minimum potential energy, folding intermediates must exhibit even lower stability than their native counterparts, so misfolding and subsequent kinetic competition of reshuffling and off-pathway reactions are expected to occur (Jaenicke 1991*a*, *b*, 1993*a*; Jaenicke & Buchner 1993). Evidence that protein biosynthesis and folding in the cell are not 100 % efficient came from turnover experiments. For

example, even under optimum growth conditions within the host, the tail spike protein of phage P22 has a yield of less than 50%. Under unbalanced physiological conditions, only proteins with incorrect conformations are produced, then continuously removed by proteolysis.

In correlating the intrinsic stability of proteins with the above hierarchy, thermodynamic measurements on point mutants, protein fragments and homologues differing in their state of association have clearly shown that each structural level makes its own contribution. As shown by NMR and other spectroscopic techniques, oligopeptides may form stable non-random conformations; at a minimum length of 15 residues, they have been shown to sustain native-like structure (Baldwin 1991). Their thermal unfolding–refolding behaviour can be quantitatively described by the standard helix–coil theories, even for short peptides. This is because both the helix nucleation constant and the enthalpy change per mole residue for helix formation are insensitive to the length of the polypeptide chain. Regarding larger fragments, it has long been known that both subdomains and domains exhibit high intrinsic stabilities which are not too different from the free energies observed for the uncleaved parent molecule. In numerous cases significant mutual stabilization has been observed both at the subdomain level and at the level of domains and subunits. A striking example is the mutual stabilization of the N- and C-terminal domains of γB-crystallin (Mayr *et al.* 1994), where the complete molecule shows the typical bimodal equilibrium transition, with the second phase superimposable over the unfolding transition of the isolated N-terminal domain fragment, whereas the isolated C-terminal domain shows surprisingly low intrinsic stability.

The mutual stabilization of subunits is illustrated by lactate dehydrogenase where the stability decreases steadily from the highly stable (native) tetramer down to domain fragments: the 'proteolytic dimer' requires structure-making salts to exhibit activity, whereas the 'structured monomer' is inactive and can only be detected as a short-lived folding intermediate on the pathway of reconstitution. The separate NAD- and substrate-binding domains are unstable but still sufficiently structured to recognize each other, exhibiting a mutual 'chaperone effect' during joint renaturation (Opitz *et al.* 1987). Extrinsic factors such as ions, cofactors and non-proteinaceous components (e.g. nucleic acids or carbohydrates) may contribute significantly to protein stability; they are also important in determining the mechanism of folding and the state of association (Jaenicke 1987).

## 3. MECHANISM OF FOLDING AND ASSOCIATION

The fact that protein folding must occur within biologically feasible time (i.e. a timespan which is much shorter than an organism's lifetime) excludes the possible existence of a random-search mechanism over all conformational space, forcing the assumption that there must be kinetic pathways for folding.

Advances in spectroscopic methodology laid a foundation for the understanding of ordered pathways and well defined intermediates. Starting from local, next-neighbour interactions, 'seeds' in the folding proceed to rate-determining late events such as proline isomerization and disulphide formation. Small proteins or constituent domains of large proteins commonly collapse, in a highly cooperative manner, into a compact structure. This structure is usually native-like, although there have been reports showing that non-native intermediates may occur. As a general principle, fast secondary structure formation precedes slow multi-step rearrangements, observable on the seconds to minutes timescale. However, due to limited time resolution and problems associated with multi-step and multiple pathways, it has not yet been possible to elucidate the complete timecourse of folding. Currently, the most detailed mechanisms available refer to small single-chain one-domain proteins: basic pancreatic trypsin inhibitor (BPTI), ribonuclease, hen egg-white lysozyme, and barnase (cf. Jaenicke & Buchner 1993).

The following general conclusions have been derived from these model systems: (i) in agreement with the above two-phase mechanism, there are compulsory pathways of folding which are, at least in part, sequential; (ii) secondary structure formation is driven by local hydrophobic surface minimization and precedes tertiary structure formation; and (iii) tertiary interactions become increasingly defined as water release consolidates the hydrophobic core.

In proceeding to domain proteins and protein assemblies, the previous conclusions remain widely unchanged. This is because proteins 'fold by parts' i.e. the domains fold and unfold independently according to

$$N_i - N_j \rightleftharpoons N_i - U_j \rightleftharpoons U_i - U_j, \qquad (1)$$

where $i$ and $j$ refer to different domains in their native (N) and unfolded (U) states. (This observation is true both *in vitro* and in the cell; Jaenicke 1993*a*.) The mechanism is significant with respect not only to protection of the nascent polypeptide chain from proteases, and the evolution of multifunctional enzymes, but also as a contribution to the rate enhancement of protein folding.

There are examples where biological function requires the cooperation of domains; for example, two domains forming one active centre. In such cases, domain pairing may occur as an additional step in the overall folding reaction, whereas domain folding represents a precursor reaction. In such cases, the above folding reaction (see equation 1) may contain domain pairing as an additional step. For a detailed analysis of sequential folding over the time range from milliseconds to seconds, see Blond-Elguindi & Goldberg (1990).

In oligomeric proteins, subunit assembly corresponds to domain pairing. The preceding steps on the pathway are again: formation of elements of supersecondary structure, collapse to subdomains and domains with the final formation of structured monomers, which then associate to yield the correct stoichiometry of the native quaternary structure. It




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

is evident that the collision complex may undergo intramolecular rearrangement to reach the state of maximum packing density and minimum hydrophobic surface area. Thus, folding and association may be followed by further first-order steps so that, in the simplest case, the overall reaction of a dimer may be written as a sequential uni–bi–unimolecular reaction

$$2\mathscr{M} \longrightarrow 2\mathrm{M}' \xrightarrow{k_1} 2\mathrm{M} \xrightarrow{k_2} \mathrm{M}_2 \longrightarrow \mathrm{N}, \qquad (2)$$

with $\mathscr{M}$,M′,M as unfolded, intermediate and structured monomer, N as native dimer and $k_1$, $k_2$ as first- and second-order rate constants (Jaenicke 1987).

Monitoring of single steps along the pathway of folding and association is dependent on the specific structure–function relation for a given quaternary structure. In most cases, biological function requires the native state of association so that the final rate-determining step can be monitored by measuring activity. Preceding steps may be accessible to spectral analysis, crosslinking, HPLC and a wealth of other methods (Jaenicke & Rudolph 1988).

The classical repertoire of methods does not allow analysis of the initial fast phase of protein folding although, recently, folding reactions in the sub-milliseconds time range became accessible, using time-resolved absorption spectroscopy (Jones *et al.* 1993).

## 4. OFF-PATHWAY REACTIONS

There are three stages where side reactions on the folding pathway may compete with proper folding and association: hydrophobic collapse; merging and 'swapping' of domains; and docking of subunits. In all cases recognition is involved, in the sense that specific substructures or surfaces must be preformed so that folding can proceed towards maximum packing density and minimum hydrophobic surface area.

Both hydrophobic collapse and domain merging involve intramolecular rearrangements. Owing to the high local concentration of the reacting groups or surfaces within one and the same polypeptide chain, they are not significantly affected by neighbouring molecules: i.e. they obey first-order kinetics, the slowest isomerization reaction determines the overall rate. In the case of domain proteins, the relative stabilities of the domains and the contributions of the domain interactions to the overall stability are crucial. The significance of the linker peptide connecting two well-defined domains has been studied by using grafting experiments; for example, mutual exchange of the linkers of β- and γ-crystallin (Jaenicke 1994). Domain contacts dominate over subunit contacts in both transplants. The recombinant (separate) domains do not interact with each other, stressing the above-mentioned argument for local concentration. The suggestion that separate domains or other substructures may recognize each other and form complexes has been known ever since the discovery of RNaseS. True domain recognition, and the folding and association of nicked polypeptide chains in lactate dehydrogenase, have been studied in detail by Opitz *et al.* (1987). The fact that, in this case, the yield does not exceed 15% illustrates the importance of side reactions (see below).

Kinetic partitioning upon reactivation of the related monomeric octopine dehydrogenase (ODH), yields 70% intact ODH and 30% inactive protein with native-like secondary structure, but increased hydrophobicity (Teschner *et al.* 1987):

$$\begin{array}{l} \mathscr{M} \longrightarrow \mathrm{U} \longrightarrow \mathrm{I_N} \longrightarrow \mathrm{N} \quad (70\,\%) \\ \qquad\qquad\quad \downarrow \\ \qquad\qquad\quad \mathrm{I_N^*} \qquad\quad (30\,\%). \end{array} \qquad (3)$$

N represents the native state characterized by correct domain pairing, whereas $\mathrm{I_N^*}$ is trapped, probably as a consequence of wrong domain interactions and subsequent aggregation.

In going from of single-chain domain proteins to protein assemblies (with $n$ subunits), kinetic competition of folding and association comes into play. This is because subunit association requires that the monomers are close to their proper conformation before they coalesce to form the native quaternary structure. If folding intermediates expose wrong hydrophobic contact sites, aggregation rather than association will result, and because now bimolecular steps are involved in the process, the second-order kinetics of association are enhanced at high protein concentration. The occurrence of inclusion bodies as a common feature of recombinant protein technology illustrates the consequences. The underlying kinetic mechanism

$$\begin{array}{ccccccc} n\mathscr{M} & \longrightarrow & n\mathrm{M}' & \xrightarrow{k_1} & n\mathrm{M} & \xrightarrow{k_2} & \mathrm{M}_n \\ \downarrow & & \downarrow & & & & \\ \mathscr{M}_n & & \mathrm{M}'_n & & & & \end{array} \qquad (4)$$

(with $\mathscr{M}$,M′,M as unfolded, collapsed and structured monomers, and $k_1$, $k_2$ as first- and second-order rate constants) corresponds to equation (3) which represents the limiting case at high dilution; it is sufficient to quantitate the observed inverse concentration dependence of reactivation and aggregation (Kiefhaber *et al.* 1991).

Taking equation (4) into consideration, three questions need to be answered: (i) what is the committed step in aggregate formation; (ii) when is the structured monomer committed to end up as the native protein; and (iii) what is known about the structure of aggregates and their constituent polypeptide chains? Regarding the first two points, it was shown that commitment to aggregation was a fast reaction, whereas the kinetics of the 'commitment to renaturation' followed precisely the slow kinetics of overall reactivation. This means that there are fast precursor reactions on the folding path (collapsed states) which still permit aggregation, whereas, after a certain intermediate has been formed, slow shuffling leads 'one way' to the native state (Goldberg *et al.* 1991). Concerning the structure of aggregates, electron microscopy and circular dichroism show that wrong subunit interactions give rise to irregular networks with a broad distribution of highly structured particles at least ten times the size of the native proteins. They resemble the native protein in its spectral properties,




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *Characterization of heart-muscle lactate dehydrogenase in its native, denatured and aggregated states*

| | native state | denatured state | aggregated state |
|---|---|---|---|
| | pH 7.6 | pH 2.3 | pH 7.6 |
| specific activity/(U mg$^{-1}$) | 639 ± 40 | 0 | 0 |
| maximum fluorescence/nm | 299 ± 1 | 333 | 331 |
| $-[\theta]222$ nm (deg cm$^2$ dmol$^{-1}$)[a] | 15300 ± 1000 | 10800 | 14700 |
| sedimentation coefficient/S | 7.6 ± 0.2 | 2.7 | ⩾ 20 |
| molecular mass/Da[b] | 140000 ± 1500 | 35000 | ~ 1000000 |
| $k_{2,\mathrm{react}} \times 10^{-3}$/(M$^{-1}$ c$^{-1}$)[c] | | | |
| 10 mM GdmCl | — | 11.4 | 11.6 |
| 30 mM GdmCl | — | 4.3 | 4.7 |
| 120 mM GdmCl | — | 0.92 | 0.81 |

[a] Ellipticity at 222 nm.
[b] From sedimentation equilibrium.
[c] Second-order rate constant of reactivation at 20 °C after acid denaturation, resolubilization in 6 M guanidinium-chloride (GdmCl) at pH 2 and subsequent dilution in 0.1 M phosphate buffer pH 7.6 (residual GdmCl concentration).

as far as turbidity allows this conclusion (see table 1). For details of the structural characteristics of inclusion bodies, refer to Bowden *et al.* (1991).

## 5. KINETIC PARTITIONING: BIOTECHNOLOGICAL ASPECTS

### (*a*) *Aggregation and inclusion body formation*

As discussed earlier, aggregation *in vitro* and formation of inclusion bodies within the cell correspond to each other; overexpression leads to high local concentrations of folding intermediates and yields precipitates instead of native protein. There are a variety of strategies to circumvent this problem.

First, one might use weaker promoters, and thus reduce the concentration. Because the aggregation reaction is above second order, this will drastically reduce the local concentration of folding intermediates, thus favouring correct folding and association. However, there are two reasons why this approach is of little practical use: first, the overall yield of the recombinant protein per gram cell mass is decreased; and second, its purification requires a full-scale separation of the guest molecule from the bulk of the host proteins, whereas purification of inclusion bodies (with their characteristic low heterogeneity) is highly simplified. Thus it is often the case that experimentation starts with the inclusion bodies; optimization focuses on *in vitro* reconstitution of the mixture after solubilization (and denaturation) in, for example, guanidinium chloride or urea. A discontinuous pulse-dilution technique has been devised to perform the dilution–reconcentration cycle in an economic way (Rudolph 1990). A specific quantity of protein is diluted and then reactivated at a concentration less than 1 μM; upon approaching the final value of reconstitution, a new aliquot of the concentrated denatured protein solution is added, and so on until the whole batch is transferred. The method has two advantages: (i) the actual concentration of the folding intermediate never exceeds the critical concentration of aggregation; and (ii) the increasing concentration of renatured protein exerts a stabilizing effect on the folding intermediate, comparable to serum albumin which is commonly used. Additives such as arginine may strongly increase the yield by shuffling aggregates back on the productive folding path.

Little is known about specific groups involved in the aggregation reaction, but early systematic experiments suggested that, in addition to covalent disulphide linkages (Jaenicke 1967), hydrophobic interactions were of major importance. Recent *in vitro* and *in vivo* studies confirm this result (Hurtley & Helenius 1989; Mitraki & King 1989; Rudolph 1990; Helenius *et al.* 1992). The monomeric two-domain enzyme rhodanese became a test case; it is inaccessible to reactivation because of its high tendency to form aggregates, but competition by detergents for the hydrophobic aggregation sites resulted in successful renaturation (Tandon & Horowitz 1986). More detailed insights came with the study of mutants; in the case of bovine growth hormone, an extension of the hydrophobic surface was shown to result in enhanced aggregation (Brems *et al.* 1988). The partitioning between folding and aggregation has been most intensively studied in the case of the tailspike endorhamnosidase from *Salmonella* phage P22 and the numerous mutants of this protein, which have been shown to either increase or suppress aggregation (Mitraki & King 1992).

The wild-type trimer is highly stable; the *in vivo* and *in vitro* folding behaviour is very similar (Fuchs *et al.* 1991). On release from the ribosome or upon dilution from denaturant solutions, the polypeptides fold into a conformation sufficiently structured for proper assembly; most of the β-sheet secondary structure along with their aromatic amino acid side-chains are close to the native state, yet even as protrimers they are still highly unstable. The anomalous stability is only acquired in a slow rearrangement reaction when the intertwined parallel β-helices merge to form the native trimer and, correspondingly, a significant part of the tailspike folding reaction occurs after subunit association (R. Seckler, personal communication).




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

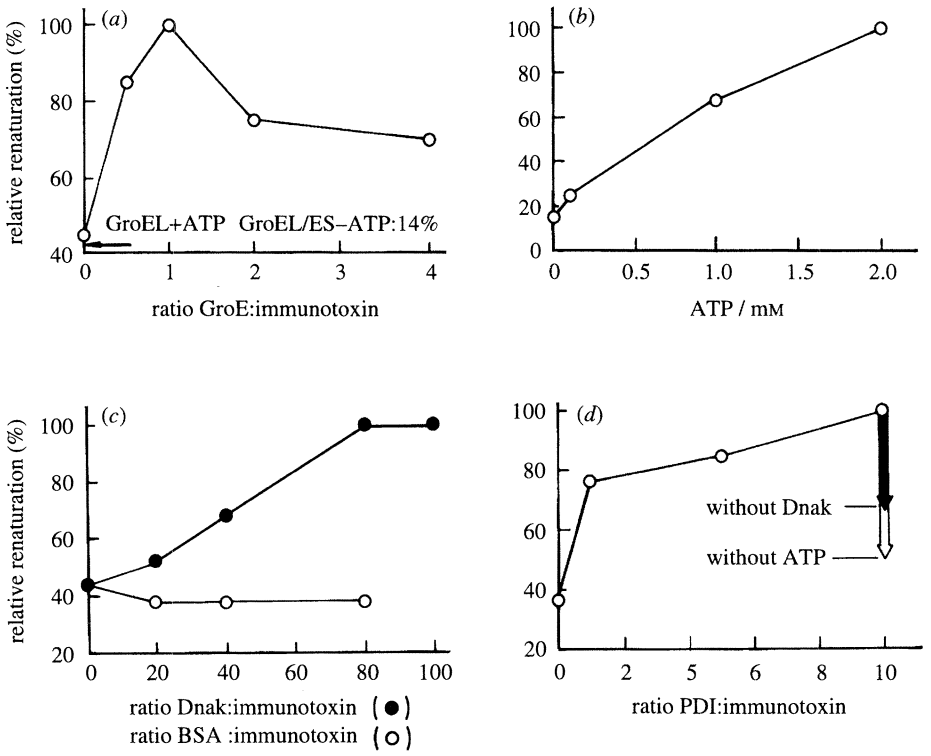


Figure 1. Renaturation of a single-chain immunotoxin facilitated by chaperones and protein disulphide isomerase. (*a*) GroE-facilitated renaturation of the GdmCl-denatured reduced protein. Optimum renaturation at equimolar GroE:immunotoxin ratio. GroEL alone has no effect; GroEL–GroES in the absence of ATP traps folding intermediates. (*b*) Effect of Mg-ATP on the yield of renaturation. (*c*) DnaK-facilitated renaturation after GdmCl-denaturation. BSA (or denatured DnaK), heated to 100 °C have no effect. (*d*) PDI-mediated oxidation of denatured and reduced immunotoxin in the presence of 60-fold excess DnaK. As indicated by the arrows, PDI and DnaK show synergistic effects (Buchner *et al.* 1992).

Both during structure formation *in vivo* and during refolding *in vitro*, the fraction of chains capable of maturing to the native form decreases with increasing temperature, with the remaining polypeptides accumulating as aggregates (Mitraki *et al.* 1993). These are formed from partly-folded intermediates that can either form native tailspike protein or form aggregates. Temperature-sensitive folding (*tsf*) point mutations reduce the folding yield at elevated temperatures, whereas second-site suppressor mutations (*su*) improve folding under such conditions (Mitraki & King 1992). Both types of mutations act by altering the stability of folding intermediates, *tsf*-substitutions destabilizing and *su*-substitutions stabilizing. In the native structure, denaturation is kinetically controlled and the effects of mutations are masked by the complicated unfolding pathway (Danner & Seckler 1993).

### (*b*) *Reconstitution in the presence of accessory proteins*

Inclusion bodies are the product of intracellular aggregation. They differ from aggregates formed in the test tube by their high packing density and large size, which may sometimes span the entire diameter of the cell (Bowden *et al.* 1991). As a consequence they can be harvested easily and washed by fractionated centrifugation, favouring their use in the downstream processing of recombinant proteins. The protein of interest can then be isolated using the same method of *in vitro* denaturation–renaturation as described before; detailed guidelines are described by Rudolph (1990).

More recently, the repertoire of methods has been extended by attempts to mimic conditions *in vivo*, with respect to folding catalysts and chaperone proteins. Two examples of this are as follows. The first deals with the reactivation of a denatured and reduced immunotoxin (B3($F_o$)-PE38KDEL) composed of the $V_H$ region of a carcinoma-specific antibody, which is connected by a flexible linker to the corresponding $V_L$ chain, which is in turn fused to truncated Pseudomonas exotoxin. The chimeric protein contains three disulphide bonds, one in each antibody domain, and one in the toxin part. Upon renaturation, aggregation of non-native polypeptide chains and the formation of incorrect disulphide linkages create more than 90% inactive molecules. Attempts to improve the yield by adding the *Escherichia coli* chaperones GroE and DnaK and bovine protein disulphide isomerase (PDI) were successful. As illustrated in figure 1, it was discovered that both GroE and DnaK influence the reaction: GroEL alone inhibits reactivation, whereas the complete GroE-system significantly increases the yield of active protein. DnaK exhibits the same effect both in free and immobilized form, which allows the chaperone to be reused in the downstream processing of the protein. PDI stimulates reactivation synergistically, and under optimum conditions reactivation yields are doubled compared with non-enzymic disulphide bond formation (Buchner *et al.* 1992).




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 2. *Rate acceleration of the folding of oxidized Fab fragments by PPIs*

(Apparent rate constants at 10 °C determined by fitting the kinetic data obtained by ELISA to a first-order reaction. Long-term and short-term denaturation were carried out by incubating Fab in GdmCl for > 2 h and 20 s, respectively. CP = cyclophilin; FKBP = FK506 binding protein; CsA = cyclosporin A. Data taken from Lilie *et al.* 1991.)

| | $k_{app}/(min^{-1})$ | | |
|---|---|---|---|
| | without CP | with CP | with FKBP |
| long-term denaturation | 0.032 ± 0.003 | | |
| short-term denaturation | 0.120 ± 0.010 | | |
| ratio PPI/Fab | | | |
| 1 | | 0.050 ± 0.003 | 0.040 ± 0.003 |
| 5 | | 1.000 ± 0.005 | 0.053 ± 0.002 |
| 10 | | 0.110 ± 0.005 | 0.063 ± 0.005 |
| 20 | | 0.110 ± 0.008 | 0.069 ± 0.007 |
| 30 | | 0.110 ± 0.008 | 0.064 ± 0.003 |
| 30 + 20 μM CsA | | 0.032 ± 0.003 | |
| Maximum yield (%) | 30 | 39 | 40 |

The second example deals with the renaturation, purification and down-stream processing of antibody fragments. As with immunotoxin, cytoplasmic expression of murine antibody chains (MAK33) in *E. coli* results in the formation of inclusion bodies. Rudolph, Buchner and coworkers designed a procedure for renaturation which produces microbially expressed authentic Fab-fragments at levels of up to 40% of the total amount of recombinant protein. Faced with a system known to show 'assisted folding' in the cell, a whole set of solvent parameters (temperature, protein concentration, redox buffer, labelling components) had to be varied to mimic conditions *in vivo* (Buchner & Rudolph 1991). More recently, Lilie *et al.* (1993, 1994) included folding catalysts in the investigation. It turns out that in the case of the oxidized Fab fragment peptidyl prolyl isomerases (PPIs) not only accelerate the refolding reaction, but also increase the proportion of correctly folded molecules, thus proving that catalysis affects kinetic partitioning by enhancing a rate-limiting step on the folding path (see table 2). Obviously, proline *cis–trans* isomerization is involved in the folding reaction. However, apart from acting as a folding catalyst, PPI also stabilizes folding intermediates in a similar way to serum albumin, or to the increasing concentration of native protein in the pulse renaturation approach mentioned previously.

PDI has no chaperone effect on the renaturation of oxidized Fab. Instead, this enzyme increases the yield of reactivation, at the same time shifting the redox dependence from a $GSH^2$:GSSG ratio around 10 mM to less than 1 mM observed for the spontaneous reaction. Again, there is kinetic competition but this time it is between domain folding and the interaction of PDI with its target cysteine residues (Lilie *et al.* 1994).

### (*c*) *Reverse micelles and immobilized proteins*

Examining the previous results, it could be assumed that overexpression of a specific protein alongside chaperones and folding catalysts in the same plasmid would finally yield 100% of the desired protein in its native state. So far, active research toward this goal has shown little success. It may be worthwhile, therefore, to devise other means by which off-pathway reactions may be eliminated. There are two alternatives for consideration: (i) reconstitution in reverse micelles; and (ii) folding of polypeptide chains bound to solid matrices.

Regarding the properties of proteins in reverse micelles, it has been found that at low water content (< 6%) proteins show enhanced stability (Luisi & Magid 1986; Gómez-Puyou 1992). Mixing a denatured protein (e.g. in 6 M guanidinium chloride) with micelle-forming compounds, at a sufficiently low protein:micelle ratio, a more or less monodisperse system may be established in which at most one monomer per micelle is present (Luisi & Magid 1986, p.447 f). Because the micelles contain renaturation buffer, isolated 'caged' subunits will form structured monomers in separate micelles. Only when the micelles merge or protein transfer is complete, can assembly occur. By using this approach, preliminary reconstitution experiments with oligomeric enzymes resulted in high levels of reactivation without significant side reactions (A. Gómez-Puyou *et al.*, unpublished results; R. Jaenicke, unpublished results).

The idea of using matrix-bound polypeptides in protein folding goes back to experiments designed to determine the catalytic properties of isolated subunits of oligomeric enzymes. By using low levels of activation and low protein concentrations, Chan (1970) succeeded in fixing oligomers only at one or very few sites so that after denaturation–dissociation and subsequent washing, monomers only were covalently bound to the matrix. Their renaturation led to unexpectedly high yields which could easily be quantitated by hybridization and similar methods.

Expanding this idea, more recent folding studies on immobilized enzymes have shown that the properties of the matrix (solid or gel, cross-linking, porosity, polarity, etc.) and the linker connecting the protein to the matrix need careful consideration (Gottschalk & Jaenicke 1991). It is obvious that this approach will be adapted for technological application; using poly-




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

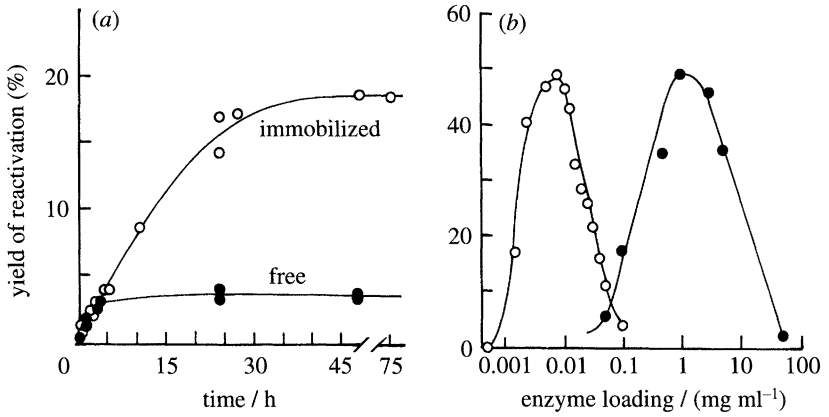


Figure 2. Renaturation–reactivation of soluble and immobilized. α-glucosidase-arg$_6$ after denaturation in 6M GdmCl. (*a*) Kinetics of reactivation at 0.1 mg ml$^{-1}$ α-glucosidase concentration. (*b*) Profiles of the kinetic competition of folding and aggregation. The immobilized enzyme allows reconstitution up to 5 mg ml$^{-1}$ (G. Stempfer & R. Rudolph, unpublished data).

ionic N- or C-terminal tails to immobilize the protein to an ion-exchange resin in a reversible fashion. The outcome in the case of α-glucosidase is most encouraging: the enzyme with its arg$_6$ tail shows long-term stability over a period of weeks; its reactivation yield after preceding GdmCl denaturation is increased at least fivefold, and the upper limit of protein concentration where maximum reactivation without aggregation can still be accomplished, is shifted from 10 μg ml$^{-1}$ to 5 mg ml$^{-1}$ (see figure 2).

## 6. KINETIC PARTITIONING: CELLULAR ASPECTS

There are three possible rate-limiting steps in the self organization of proteins: disulphide formation; proline *cis–trans* isomerization; and self assembly. On the other hand, there are two enzymes, localized in the appropriate cellular compartments, which catalyse the first two steps, and a whole range of chaperones and helper proteins to assist the third reaction.

Faced with the wealth of recent data which seem to round up in well-defined reaction cycles, one might suspect that, regarding the 'vision' of how kinetic partitioning *in vivo* might work, it is sufficient to summarize briefly the enzyme mechanisms and specificities, on one hand, and the regulation of intra- and intermolecular interactions in the cavities of chaperones, on the other. However, both folding catalysis and chaperoning are far from being solved. The current state of our knowledge is reviewed in Jaenicke & Buchner (1993); only certain new aspects will be outlined here.

### (*a*) *Folding catalysts*

The biological significance of PDI, PPI and chaperones is now well-established. The three-dimensional structure of representatives of all three types of accessory protein have been resolved. DsbA, the PDI-homologue in *E. coli*, shows a close relation to thioredoxin; its second domain may be involved in communication with a more complex mechinery, including a membrane-spanning unit (DsbB). The anomalous redox and stability properties of DsbA (with the reduced form of the enzyme showing higher stability than the oxidized one) have been resolved elegantly by a number of recent studies (Bardwell & Beckwith 1993; Wunderlich & Glockshuber 1993).

It seems that PPIs are ubiquitous enzymes, catalysing rotation around the X-pro peptide bond (at least *in vitro*) and inhibiting the signal transduction processes involved in, for example, immunosuppression. The three-dimensional structure of two different representatives of the PPI family has been elucidated, supporting a 'catalysis by distortion' mechanism. The substrate specificity is low.

Obviously, PPIs serve various functions in the cell. To date there is little direct evidence that they serve as folding catalysts, apart from the location of the enzyme in the endoplasmic reticulum, and its effect on the formation of collagen in fibroblasts.

### (*b*) *Chaperones*

With respect to the involvement of chaperones in the folding and targeting of proteins, the field is moving so fast (and producing many inconsistent results) that it seems premature to propose general mechanisms of chaperone action and interaction even for the best-known systems DnaK–DnaJ–GrpE and GroEL–GroES (Jaenicke 1993; Jaenicke & Buchner 1993). X-ray structures are now becoming available and therefore cartoons should soon become obsolete.

Open questions about the role of GroEL and GroES deal with the stoichiometries and roles of the various components in the cycle of ATP hydrolysis, the substrate specificity and the conformational state of the bound polypeptide and, finally, the functional differences of single- versus double-doughnut assemblies, as well as symmetrical versus asymmetrical quaternary structures. Some of these questions have been addressed in a series of recent papers (Azem *et al.* 1994; Schmidt *et al.* 1994; Todd *et al.* 1994) which suggest that GroEL–GroES exhibit half-of-the-sites activity, involving both asymmetrical and symmetrical forms of the complex. As a result of a single turnover, unfolded protein (originally bound to the asymmetrical complex) dissociates in a non-native state, in a way that is




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

consistent with intermolecular transfer of the substrate protein between toroids of high and low affinity. It remains to be shown how this model can be correlated or combined with other reaction schemes proposed on the basis of equilibrium experiments (cf. Hendrick & Hartl 1993).

Whether chaperones catalyse protein folding, acting as helper proteins and 'unfoldases', is still controversial. In numerous cases chaperones have been shown to slow down reactivation, probably due to their high affinity for their unfolded substrate. 'Resurrection of aggregates' is the exception rather than the rule: for steric reasons because they do not fit into the 'active sites'; or for energetic reasons because aggregates are commonly trapped in a deep energy well. Support for the steric argument comes from lag phases observed in the GroE-assisted formation of oligomeric proteins (Jaenicke & Buchner 1993).

Regarding the state of association of the released protein substrate, direct proof for the monomeric state (i.e. against oligomerization on the chaperone) came from reconstitution experiments with glutamin synthetase, where the formation of the native dodecamer has been shown to occur in solution after release of the subunits from the binary GroEL-complex (Fisher 1993). What drives the release, and what the polypepticle chain in the chaperone-substrate complex looks like, is still unresolved.

With respect to protection of the nascent polypeptide chain from misfolding, or chemical modification, the question arises 'when in the life of a protein, folding and chaperoning start'. Experimental examination of this question is difficult because of the heterogeneity and low concentration of nascent polypeptides. Previous work pointed to cotranslational folding-by-parts. Recently, monoclonal antibodies have been used to show that antigen recognition may already start at the subdomain level (Friguet *et al.* 1994).

Subdomains are approximately twice the size required by DnaJ for the recognition of its substrate. How the binding of the DnaJ–Hsp 70 chaperone assembly correlates with the *in vitro* translation studies remains to be shown. It seems there may be still earlier processes involved. It has been established for a long time that in the case of proteins targeted to the endoplasmic reticulum, complex formation with SRP starts when the polypeptide is just leaving the ribosome (Rapoport 1992). There is one more component, NAC, involved in targeting, which raises the question does the nascent chain leave the ribosome through a cleft rather than a channel? NAC is a heterodimeric protein which binds exclusively to the nascent polypeptide, shielding non-signal peptide regions from promiscuous interactions with the signal recognition particle. As shown by Wiedmann *et al.* (1994), in the absence of NAC, proteins lacking the signal peptide are recognized and mistargeted by the SRP. Adding NAC restores the specificity of SRP binding, as well as correct targeting and translocation. Obviously, NAC binds to signal peptides that are only partly exposed from the ribosome, serving as an adaptor between the translation complex and the cellular folding and transport machineries, thereby protecting nascent chains from premature and inappropriate interactions. Thus, the nascent chain is tunnelled through a groove between the ribosome and NAC. Correspondingly, the length of the amino-acid sequence accessible to the cytosol is unexpectedly short: using truncated mRNA as a means to produce fixed peptides of constant length, it turns out that no more than *ca.* 15 amino-acid residues away from the ribosomal peptidyl transferase site the growing polypeptide becomes the target of chaperone-like components regulating their interactions (M. Wiedmann, personal communication).

In summary, the picture that emerges from studies of both prokaryotes and eukaryotes clearly shows that off-pathway reactions in the cell are blocked by a whole arsenal of components from the moment the peptide emerges. The reason why, in spite of this protection, protein biosynthesis and subsequent structure formation commonly does not yield 100% (cf. Hurtley & Helenius 1989) seems obvious: interactions involved in chaperone action must not be too strong, otherwise the nascent polypeptide chains would be trapped. Thus, during the off-reaction of complex formation, molecules may escape from the folding pathway ending up in protein turnover.

This review was written during a stay at the Fogarty International Center, NIH, Bethesda, Maryland. Fruitful discussions with Dr J. Bardwell, Dr J. Buchner, Dr R. Glockshuber, Dr F.-U. Hartl, Dr G. Lorimer, Dr R. Seckler and Dr M. Wiedmann are gratefully acknowledged. Work carried out in the author's laboratory was generously supported by grants from the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, and Boehringer Mannheim GmbH.

This paper is dedicated to Professor Hans Neurath on the occasion of his 85th birthday.

## REFERENCES

Azem, A., Kessel, M. & Goloubinoff, P. 1994 Characterization of a functional $GroEL_{14}$ $(GroES_7)_2$ chaperonin hetero-oligomer. *Nature, Lond.* **265**, 653–656.

Baldwin, R.L. 1991 Experimental studies of pathways of protein folding. *CIBA Found. Symp.* **161**, 190–205.

Bardwell, J. & Beckwith, J. 1994 The bonds that tie: catalyzed disulfide bond formation. *Cell* **74**, 769–771.

Blond-Elguindi, S. & Goldberg, M.E. 1990 Kinetic characterization of early immunoreactive intermediates during refolding. *Biochemistry* **29**, 2409–2417.

Bowden, G.A., Paredes, A.M. & Georgiou, G. 1991 Structure and morphology of inclusion bodies in E. coli. *Biotechnology* **9**, 725–730.

Brems, D.N., Plaisted, S.M., Kaufmann, E.W., Lund, M. & Tomich, C.-S.C. 1988 Stabilization of an associated folding intermediate of BGH by site-directed mutagenesis. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3367–3371.

Buchner, J. & Rudolph, R. 1991 Renaturation, purification and characterization of $recF_{ab}$-fragments produced in *E. coli*. *Biotechnology* **9**, 157–162.

Buchner, J., Brinkmann, U. & Pastan, I. 1992 Renaturation of a single-chain immunotoxin facilitated by chaperones and PDI. *Biotechnology* **10**, 682–685.

Chan, W.W.-C. 1970 Matrix-bound protein subunits. *Biochem. biophys. Res. Commun.* **41**, 1198–1204.

Danner, M. & Seckler, R. 1993 Mechanism of phage P22 tailspike protein folding mutation. *Protein Sci.* **2**, 1869–1881.





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Fisher, M.T. 1993 On the assembly of dodecameric glutamine synthetase from stable chaperonin complexes. *J. biol. Chem.* **268**, 13777–13779.

Friguet, B., Djavady-Ohaniance, L., King, J. & Goldberg, M.E. 1994 *In vitro* and ribosome- bound folding intermediates of P22 tailspike protein detected with monoclonal antibodies. *J. biol. Chem.* **269**, 15945–15949.

Fuchs, A., Seiderer, C. & Seckler, R. 1991 *In vitro* folding of phage P22 tailspike protein. *Biochemistry* **30**, 6598–6604.

Goldberg, M.E., Rudolph, R. & Jaenicke, R. 1991 A kinetic study of the competition between renaturation and aggregation of lysozyme. *Biochemistry* **30**, 2790–2797.

Gómez-Puyou, A. (ed.) 1992 *Biomolacules in organic solvents.* Boca Raton: CRC Press.

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

Helenius, A., Marquardt, T. & Braakman, I. 1992 The endoplasmic reticulum as a protein-folding compartment. *Trends Cell Biol.* **2**, 227–231.

Hendrick, J.P. & Hartl, F.-U. 1993 Molecular chaperone functions of heat-shock proteins. *A. Rev. Biochem.* **62**, 349–384.

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

Jaenicke, R. 1967 Intermolecular forces in the process of heat aggregation of globular proteins. *J. Polymer Sci.* **16**, 2143–2160.

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

Jaenicke, R. & Rudolph, R. 1988 Folding proteins. In *Protein structure: a practical approach* (ed. T.E. Creighton), pp. 191–223. Oxford: IRL Press.

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 and assemblies. *Biochemistry* **30**, 3147–3161.

Jaenicke, R. 1993*a* What does protein refolding *in vitro* tell us about protein folding in the cell? *Phil. Trans. R. Soc.Lond.* B **339**, 287–295.

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

Jaenicke, R. & Buchner, J. 1993 Protein folding: From 'unboiling an egg' to 'catalysis of folding'. *Chemtracts: Biochem. molec. Biol.* **4**, 1–30.

Jaenicke, R. 1994 Eye lens proteins: Structure, superstructure, stability and genetics. *Naturwissenschaften.* **81**, 423–429.

Jones, C.M., Henry, E.R., Hu, Y., Chan, C.-K., Luck, S.D., Bhuyan, A., Roder, H., Hofrichter, J. & Eaton, W.A. 1993 Fast events in protein folding initiated by ns laser photolysis. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11860–11864.

Kiefhaber, T., Rudolph, R., Kohler, H.-H. & Buchner, J. 1991 Protein aggregation *in vitro* and *in vivo*. *Biotechnology* **9**, 825–829.

Lilie, H., Lang, K., Rudolph, R. & Buchner, J. 1993 Prolin isomerases catalyze antibody folding *in vitro*. *Protein Sci.* **2**, 1490–1496.

Lilie, H., McLaughlin, S., Freedman, R. & Buchner, J. 1994 Influence of PDI on antibody folding in vitro. *J. biol. Chem.* **269**, 14290–14296.

Luisi, P.L. & Magid, L.J. 1986 Solubilization of enzymes and nucleic acids in hydrocarbon micellar solutions. *CRC Crit. Rev. Biochem.* **20**, 409–474.

Mayr, E.-M., Jaenicke, R. & Glockshuber, R. 1994 Domain interactions and connecting peptides in lens crystallins. *J. molec. Biol.* **235**, 84–88.

Mitraki, A. & King, J. 1989 Protein folding intermediates and inclusion body formation. *Biotechnology* **7**, 690–697.

Mitraki, A. & King, J. 1992 Amino acid substitutions influencing intracellular protein pathways. *FEBS Lett.* **307**, 20–25.

Mitraki, A., Danner, M., King, J. & Seckler, R. 1993 Ts-mutations and second-site suppressor substitutions affect folding of the P22 tsp *in vitro*. *J. biol. Chem.* **268**, 20071–20075.

Opitz, U., Rudolph, R., Jaenicke, R., Ericsson, L. & Neurath, H. 1987 Proteolytic dimers of porcine LDH. *Biochemistry* **26**, 1399–1406.

Rapoport, T.A. 1992 Transport of proteins across the ER membrane. *Science, Wash.* **258**, 931–936.

Rudolph, R. 1990 Renaturation of recombinant, disulfide-bonded proteins from 'inclusion bodies'. In *Modern methods in protein and nucleic acid research* (ed. H. Tschesche), pp. 149–171. Berlin, New York: de Gruyter.

Schmidt, M., Rutkat, K., Rachel, R., Pfeifer, G., Jaenicke, R., Viitanen, P., Lorimer, G. & Buchner, J. 1994 Symmetric complexes of GroE chaperonins as part of the protein folding cycle. *Science, Wash.* **265**, 656–659

Tandon, S. & Horowitz, P. 1986 Detergent-assisted refolding of GdmCl-denatured rhodanese. *J. biol. Chem.* **261**, 15615–15618.

Teschner, W., Rudolph, R. & Garel, J.-R. 1987 Intermediates on the folding pathway of ODH from *Pecten jacobaeus*. *Biochemistry* **26**, 2791–2796.

Todd, M.J., Viitanen, P.V. & Lorimer, G.H. 1994 Dynamics of the chaperonin ATPase cycle: Implications for facilitated protein folding. *Science, Wash.* **265**, 659–666.

Wiedmann, B., Sakai, H., Davis, T.A. & Wiedmann, M. 1994 A protein complex required for signal-sequence-specific sorting and translocation. *Nature, Lond.* **370**, 434–440.

Wunderlich, M. & Glockshuber, R. 1993 Redox properties of PDI (DsbA) from *E. coli*. *Protein Sci.* **2**, 717–726.





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES