ANTHONY ARTALE
MED LIB NATHAN CUMMINGS CTR (S-46)
MEMORIAL SLOAN KETTERING CANCER CTR
1275 YORK AVENUE
NEW YORK, NY 10021
UNITED STATES

Telephone: 212/639-7441
Fax: 646/422-2316

Title:	**PROTEIN ENGINEERING**
DB Ref. No.:	IRN10409592
ISSN:	ISSN02692139
Vol./Issue:	N/A
Date:	1991
Pages:	711-717
Article Title:	THE FUNCTIONAL MECHANISM OF HEAT SHOCK
Article Author:	HUBBARD, J
Report Number:	IRN10409592
Publisher:	IRL PRESS,
Client Number:	DDS36735/GANGI-DINO, RITA

Transmission problem
Please redo
Manon

Protein Engineering vol.4 no.7 pp.711–717, 1991

REVIEW

The role of heat-shock and chaperone proteins in protein folding: possible molecular mechanisms

T.J.P.Hubbard[1] and C.Sander[2]

Protein Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka 565, Japan and [2]European Molecular Biology Laboratory, Heidelberg, Germany

[1]Present address: Cambridge Centre for Protein Engineering, MRC, Hills Road, Cambridge, UK

Recently some heat-shock proteins have been linked to functions of 'chaperoning' protein folding *in vivo*. Here current experimental evidence is reviewed and possible requirements for such an activity are discussed. It is proposed that one mode of chaperone action is to actively unfold misfolded or badly aggregated proteins to a conformation from which they could refold spontaneously; that improperly folded proteins are recognized by excessive stretches of solvent-exposed backbone, rather than by exposed hydrophobic patches; and that the molecular mechanism for unfolding is either repeated binding and dissociation ('plucking') or translocation of the protein backbone through a binding cleft ('threading'), allowing the threaded chain to refold spontaneously. The observed hydrolysis of ATP would provide the energy for active unfolding. These hypotheses can be applied to both monomeric folding and oligomeric assembly and are sufficiently detailed to be open to directed experimental verification.

Key words: chaperone/heat-shock proteins/protein engineering/protein folding

Introduction

Heat-shock proteins (hsps) are a small family of proteins, rapidly synthesized in large quantities by cells when heat shocked or otherwise stressed (Neidhardt *et al.*, 1984). Hsps are found in all prokaryotic and eukaryotic cells yet studied and form several families, with sequences that are closely homologous between species. Hsps and constitutively expressed homologues (hscs) have been shown to be essential for cell viability in several systems (Craig *et al.*, 1987; Kusukawa and Yura, 1988; Normington *et al.*, 1989); however until recently, their function was poorly understood.

Hsp70 and groEL families

Two universal hsp families are those of hsp70 and groEL. Pelham (1986) proposed that the hsp70 family might be involved in the assembly and disassembly of proteins and protein-containing structures, based on the binding of insoluble nucleolar proteins by some hsps during heat shock, the localization in the endoplasmic reticulum (ER) of other hsps, and their ATPase activity. With the discovery of the role of rubisco-binding protein (RBP) in the assembly of rubisco holoenzyme (ribulose biphosphate carboxylase) and its sequence homology to GroEL (Hemmingsen *et al.*, 1988), it was proposed that hsps form part of a larger family of 'molecular chaperones', a term first used by Laskey and Earnshaw (1980). Ellis (1987) suggested that the chaperone protein function is to ensure correct folding of polypeptides and their assembly into oligomers.

GroEL family members are required for oligomeric assembly

GroEL family members are GroEL in *Escherichia coli*, hsp60 in mitochondria (Reading *et al.*, 1989) and RBP in chloroplasts. They have now all been shown to be essential for the assembly of some oligomeric proteins *in vivo* (Hemmingsen *et al.*, 1988; Cheng *et al.*, 1989; Goloubinoff *et al.*, 1989b), *in vitro* (Goloubinoff *et al.*, 1989a), self assembly (Cheng *et al.*, 1990; Lissin *et al.*, 1990) and even the correct folding of some monomeric proteins, after their import into mitochondria (Ostermann *et al.*, 1989). Overexpression of GroEL has also been shown to suppress the effects of mutations in a wide variety of genes, implying an ability to generally rescue folding defects (Van Dyck *et al.*, 1989).

Role of hsp70 family members in membrane translocation

The hsp70 family has been strongly linked to membrane translocation processes on both sides of the membrane (Neupert *et al.*, 1990). In eukaryotes the cystolic Hsc70 facilitates post-translational translocation of proteins into mitochondria and the ER, by interacting with proteins before they are translocated (Chirico *et al.*, 1988; Deshaies *et al.*, 1988). Hsp70 family members located within the ER, e.g. immunoglobulin heavy chain binding protein BiP (Kassenbrock *et al.*, 1988), also referred to as glucose-regulated protein GRP-78, and within mitochondria (Craig *et al.*, 1989) have also been implicated in membrane translocation (Kang *et al.*, 1990; Scherer *et al.*, 1990; Vogel *et al.*, 1990). In *E.coli* the DnaK gene product (an hsp70 homologue, herein referred to simply as DnaK) appears to facilitate protein export (Phillips and Silhavy, 1990). It was suggested early on that an enzyme unfolds proteins that are to be translocated into mitochondria (Rothman and Kornberg, 1986), after it had been shown that translocating proteins were in an incompletely folded state (Eilers and Schatz, 1986). Hsp70 is suggested to act as an unfoldase in this process since pre-incubation of proteins with urea can mimic Hsc70's stimulatory effect on translocation (Eilers *et al.*, 1988). Proteins have also been shown to translocate across the membrane in an extended conformation (Ostermann *et al.*, 1989). The presence of GroEL can also enhance membrane translocation (Phillips and Silhavy, 1990), with mutations in GroEL blocking the export of some proteins (Kusukawa *et al.*, 1989). GroEL appears to enhance translocation by forming complexes with precursor proteins, maintaining them in a translation competent state (Lecker *et al.*, 1989). Like Hsp70, this does not seem to be an entirely passive process since *in vitro* GroEL/ES inactivates refolded β-lactamase precursor (a monomeric protein destined for translocation) when it is added in the absence of ATP (Laminet *et al.*, 1990). The implication is that GroEL also unfolds proteins destined for translocation.

Other functions of hsp70

Hsp70 family members have a variety of functions other than aiding membrane translocation. DnaK has been shown *in vitro* to restore the activity of heat-inactivated RNA polymerase

(Skowyra *et al.*, 1990) and dissociate and activate aggregated DnaA protein (Hwang *et al.*, 1990), both in ATP-dependent processes. The *in vivo* renaturation of the temperature-sensitive bacteriophage λcI*857* repressor protein, after thermal inactivation, also appears to require DnaK (Gaitanaris *et al.*, 1990). In eukaryotic cells, mitochondrial Hsp70 mutants block the correct refolding of imported proteins, as well as translocation (Kang *et al.*, 1990). Although they have not been shown directly to participate in protein folding or assembly, cytoplasmic Hsp70s appear to associate co-translationally with most nascent polypeptide chains (Beckmann *et al.*, 1990) and BiP is known to bind to aberrant polypeptides in the ER (Kassenbrock *et al.*, 1988). Hsc70 also disrupts tight clathrin–protein complexes without chemical degradation (Rothman and Schmid, 1986), perhaps in a similar fashion to the way DnaK removes the λP protein from the tight replication complex bound to λDNA (Zylicz *et al.*, 1989). This mode of DnaK action in *E.coli* involves an accessory protein coded by the gene DnaJ (Wickner *et al.*, 1991). Recently SCJ1, a yeast homologue of the DnaJ gene, was identified (Blumberg and Silver, 1991). Hybridization experiments indicate the existence of a family of related genes in yeast, suggesting an interaction of the type DnaK–DnaJ or Hsp70–SCJ1 may be a general requirement for Hsp70 function.

DnaK functions as a chaperone in two ways

Experiments with the *E.coli* Hsp70 homologue DnaK suggest that the same chaperone protein may affect the amount of aggregated protein in the cell in two ways. Without ATP present DnaK protects host RNA polymerase (RNAP) from heat inactivation (Skowyra *et al.*, 1990). This would be consistent with a protective binding activity that stabilizes the folded state in adverse environmental conditions. Only with ATP present can DnaK actively dissolve existing RNAP aggregates and restore RNAP activity (Skowyra *et al.*, 1990). This restoration activity is a state changing (unfoldase) function. The DnaK756 mutant of DnaK, which lacks ATPase activity, is able to protect RNAP from heat shock but not to restore activity to heat-inactivated samples. Since both protection and restoration are carried out by one protein it is likely that mechanisms of action are closely linked.

Crystal structure of a domain of Hsc70

The ATPase activity of the N-terminal domains of hsp70 family proteins can now be understood in molecular detail as the crystal structure of the N-terminal 44 kDa fragment of Hsc70 has been determined (Flaherty *et al.*, 1990) and found to be quite similar to that of actin and also to hexokinase (Kabsch *et al.*, 1990; Flaherty *et al.*, 1991). Actin is known to form filaments and some hexokinases are active as dimers. For comparison, hsp70 family members have been identified as monomers, dimers and higher order aggregates under different conditions (Guidon and Hightower, 1986; Heuser and Steer, 1989). Interestingly, a mechanism for the clathrin-uncoating activity of Hsc70, analogous to muscle contraction, was proposed before the similarity to actin was discovered (Heuser and Steer, 1989). The structure of these ATPase domains consists of two lobes of roughly equal size with a deep cleft between them at the bottom of which ATP/ADP binds. Comparison of the actin and hexokinase structures suggests that without ATP the left is open and that binding of ATP leads to a large subdomain movement, around a putative hinge, closing the cleft. Hydrolysis of ATP is likely to weaken the affinity for bound nucleotide, facilitating opening of the cleft and ADP/ATP exchange (hypothesis, not yet proven experimentally).

Structure of the Hsp70 C-terminal domain

Much less is known about the structure or function of the C-terminal 26 kDa of Hsc70. This domain may contain the substrate recognition domain (Chappell *et al.*, 1986), and model-building studies have suggested a structure similar to that of the HLA peptide-binding cleft (Rippmann *et al.*, 1991). However, one may argue that the peptide-binding site is in the N-terminal, ATPase domain. This is where glucose binds to hexokinase and a number of different proteins to actin and where ATP hydrolysis could most directly affect protein conformation. Also, the C-terminal domain is less conserved in sequence between species and may therefore only be responsible for the functional variations within the hsp70 family.

GroEL structure

Neither a crystal structure nor details of the catalytic mechanism is known for any groEL family member. It is known, however, that proteins in this class form 14mers (a double ring each of seven subunits) and, at least in the case of GroEL, associate with a small accessory protein (GroES) which exists as a single ring of approximately seven subunits (Chandrasekhar *et al.*, 1986).

Common function of chaperone proteins

What all members of both chaperone families appear to have in common is that they are weak ATPases and have activities involving a transient association with proteins in a way that affects their folded state without chemical modification. Members of both families also seem to interact preferentially with proteins in non-native conformations. What is the relation of heat-shock proteins to protein folding? On the one hand *in vitro* refolding experiments have led to the belief that no catalysts are required for either monomeric folding (Anfinsen, 1973) or oligomeric assembly (Jaenicke, 1987). On the other hand, the experimental evidence reviewed here indicates not only that chaperones transiently associate with proteins, but that both *in vivo* and *in vitro* they are necessary for correct oligomeric assembly and even monomeric folding for some proteins under some conditions. The different ways in which chaperones might influence folding are reviewed here and possible molecular mechanisms proposed.

Requirements for preventing misfolding and promoting unfolding

Definition of states A/M, U and F

Figure 1 shows three theoretically possible classes of conformational states for a protein chain: unfolded (**U**), correctly folded (**F**), and aggregated/misfolded (**A/M**). The (**A/M**) class contain all molecules in non-native, but compact conformations. The classes of compact states (**F** and **A/M**) are thermodynamically at least marginally more stable than **U**, so at equilibrium they will be favoured. **U** can be considered as a collection of transition states that a protein passes through after synthesis or after active unfolding. Very few molecules will be in state **U** at any one time.

The need for chaperones

Given that many protein chains can fold spontaneously *in vitro*, **U–F** is probably favoured kinetically over **U–A/M** *in vivo* for most proteins. However, in the complex and dynamic cellular environment, random events may eventually lead a significant proportion of molecules to a misfolded state. For some proteins and/or under some conditions (such as heat shock) the **U–A/M** transition may even be the favoured one. The principal function

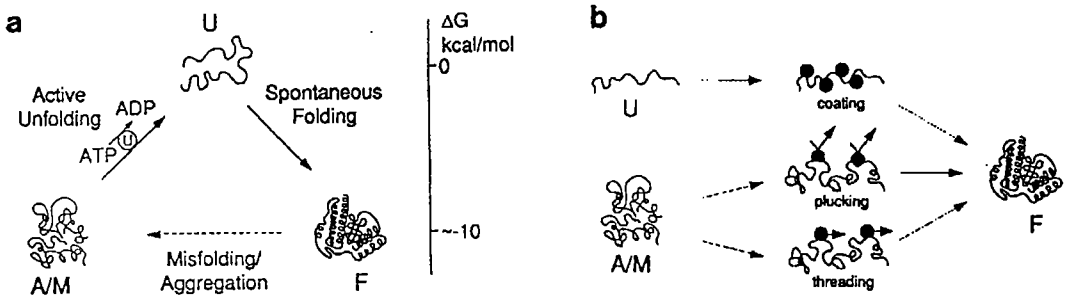

Fig. 1. (a) Hypothetical active unfolding cycle between folded (**F**), unfolded (**U**) and aggregated/misfolded (**A/M**) states of proteins. The existence of a kinetically favoured correct folding pathway will favour the process **U→F** over **U→A/M**. Random collisions and atomic fluctuations, especially in conditions of heat stress, however, will overcome the energy barrier between **F** and **A/M**, leading over time to a significant population of the improperly folded state **A/M**. A postulated unfoldase activity (**u**) can drive the thermodynamically unfavourable processes of **A/M→U** or **F→U** at the expense of ATP breakdown. If the chaperone protein **u** has a higher affinity for **A/M** than **F**, and/or if there is a favoured folding pathway **U→F**, a cycle of active unfolding and refolding will keep the highest proportion of molecules in the properly folded state **F**. Note that state (**U**) is unstable and only sparsely populated. (b) Three different mechanisms for aiding correct folding. (i) *Coating* an unfolded or partly folded chain protects it from improper aggregation in unfavourable environmental conditions and later release allows it to fold properly. This process had no requirement for ATP. (ii) ***Plucking*** or repeated binding is an active unfolding process that requires ATP; the chaperone protein binds to and dissociates from the target protein in rapid succession. At each encounter, a section of the polypeptide chain is straightened out. (iii) In the active and energy-requiring unfolding process of ***threading***, the chaperone protein moves along the chain, untangling it in the process, in analogy to threading on the ribosome during protein synthesis. In each process, after release of the chain from the chaperone protein, it preferentially folds to the native state **F**.

of chaperones may therefore be to reduce the loss of 'expensive' protein chains into functionally useless aggregates.

Previous models: binding to incompletely folded chains

Previous models of chaperone function have concentrated on chaperones binding to nascent, partly unfolded proteins (or subunits of oligomers) transiently, discouraging the formation of aggregates and/or enhancing correct folding by, for example, stabilizing folding intermediates (Pelham, 1986; Rothman, 1989). A chaperone working this way would either make **U→F** kinetically more accessible or **U→A/M** less accessible or both. In an extension to this first model, it has been argued that protein chains that start to fold during synthesis on the ribosome may be more likely to fold incorrectly and that chaperones could encourage the correct **U→F** pathway by preventing folding until synthesis is complete (Rothman, 1989). Such transient binding might favour folding over aggregation if during the time of binding the environment changed: for example, if chain synthesis was completed; if missing subunits became available; if a destabilizing heat shock ended; or if a slow-folding transition in an unbound part of the molecule was completed, e.g. proline *cis–trans* isomerization or disulphide-bond exchange.

Protection by binding may not be sufficient

Protection of a partially folded chain through binding may, however, be insufficient. If, for example, the state **A/M** is thermodynamically favoured (e.g. due to the much larger number of possible conformations in state **A/M** compared with state **F**) then an activity affecting the kinetic preference for **U→F** over **U→A/M** will not prevent the transition of molecules from **F** to **A/M** states as a result of intramolecular atomic motion and collisions between molecules. Nor would such an activity be able to restore the activity of already misfolded proteins, as some chaperone proteins have been shown to do (e.g. Skowyra *et al.*, 1990).

Unfoldase action

Reducing the amount of undesired aggregated protein accumulating in a cell can be achieved by either removing molecules in that state from the system or by restoring the molecules to another state. The former can be achieved by selectively degrading chains that have become misfolded and resynthesizing molecules of that protein. However, much less energy would be expended if molecules could be actively 'pumped' from the conformation **A/M** to **F** (Figure 1). In our view, chaperones are 'unfoldases' which actively unfold aggregated proteins to state **U**, from where state **F** is accessible through spontaneous refolding. Such activity would require energy: chaperones are observed to turn over one ATP molecule per protein chain bound (Rothman and Schmid, 1986). The free energy gain of ATP→ADP hydrolysis is ~10 kcal/mol, roughly equal to the energy of stabilization of state **A/M** compared with state **U** (Figure 1). Using this energy, chaperones could dynamically oppose the accumulation of aggregates by catalysing the removal of molecules from state **A/M**.

Possible molecular mechanisms for chaperone action

Proposals for a mechanism of the above chaperone activities have to answer at least two questions: (i) by which molecular event does the chaperone recognize and bind to incompletely folded proteins or misfolded proteins, and (ii) in what way is a target protein kept from misfolding or brought out of its misfolded state?

Recognition of non-native protein folds

During destabilizing events, such as heat shock, less stable parts of a protein (e.g. loops) may become locally unfolfed/misfolded. Chaperone protection could be by associating with these regions, so reducing the chance of further unfolding/misfolding/aggregation. In this case the molecular recognition event would be solely that of associating with any region of misfolded protein chain.

How might partially folded or misfolded chains be recognized? It has been suggested that chaperones function through binding to hydrophobic patches (e.g. Pelham, 1986). However, in recent experiments selective binding of hydrophilic peptides to Hsc70 and BiP was observed (Flynn *et al.*, 1989; DeLuca-Flaherty *et al.*, 1990). Instead of hydrophobic patches, main-chain polar groups could be the target for recognition. The idea is based in part on the observation that in correctly folded proteins most backbone NH and CO groups are involved in hydrogen bonds

in segments of secondary structure, like α-helices and β-sheets, away from the protein surface. The protein solvent accessible surface is thus depleted in exposed polar backbone groups, relative to an unfolded chain (Baumann *et al.*, 1989). Misfolded proteins very likely have incompletely formed secondary structure, especially proteins with a significant portion of β-sheets, and so expose extensive regions of backbone to solvent. In molecular detail, exposed backbone segments could be recognized by an extended piece of chain that forms a β-sheet-like association with the protein chain to be bound. Side chains would be involved in this association, modulating peptide specificity. The recognition of exposed backbone may be analogous to the recognition of locally unfolded regions of globular proteins by proteases.

Protection against misfolding by coating

Selective binding to partially or fully unfolded chains is sufficient to explain protection against misfolding or aggregation, while adverse environmental conditions persist. No ATP hydrolysis is required (Skowyra *et al.*, 1990). Protection by binding would be more efficient if several or many chaperone proteins bind to a single chain. Hsp70 family members have been identified as various multimers experimentally (Guidon and Hightower, 1986; Heuser and Steer, 1989). Interestingly, the structural similarity between the ATPase domain of Hsp70 and actin suggests that Hsp70 may function through oligomerization. Protection by coating is reminiscent of the protection of single strands of nucleic acid by special single-strand-binding proteins and might explain the ability of DnaK to protect a RNA polymerase from heat-induced aggregation best when there is a large excess of chaperone over target protein (Skowyra *et al.*, 1990).

Active unfolding by 'threading'

A possible mechanism for active unfolding is one in which the protein chain is unravelled by a threading process: the chaperone complex bound to a section of protein backbone slides along the chain, 'unravelling' its folded conformation (Figure 2). The actual sliding process does not necessarily need energy input (similar to the mechanism postulated for the sliding of repressors on DNA (Berg *et al.*, 1983)). The ATP→ADP breakdown observed (Rothman and Schmid, 1986) could provide energy for threading, for straightening out a section of polypeptide chain and/or for peptide release. This process might unravel the chain partially, or completely from N- to C-terminus. If the latter, protein chains emerging sequentially out of the chaperone's binding cleft would resemble those emerging from the ribosome, during protein synthesis.

Sequential threading could be particularly well suited for the purpose of bringing proteins to their properly folded state, if there are special folding pathways for protein chains emerging from a cleft. Note that it is likely that many proteins are partially folded before ribosomal synthesis is complete, as synthesis is slower than folding. Also, in the case of subtilisin, the N-terminus has been shown to 'guide' folding in an intermolecular process (Zhu *et al.*, 1989). Therefore the folded state (F) of some proteins may be kinetically more easily accessible if folding is allowed to occur in a sequential way. A threading unfoldase would thus enhance the rate of spontaneous refolding to F by bringing the target protein from the A/M state to a U state from which sequential folding is favoured.

A threading mechanism is consistent with the role of Hsc70 in membrane translocation. Hsc70 may 'thread' a protein chain into the membrane, 'unravelling' its folded conformation in the process and organelle Hsp70s may receive it on the inside. It is only a short conceptual leap from threading through a membrane to threading within the cytosol.

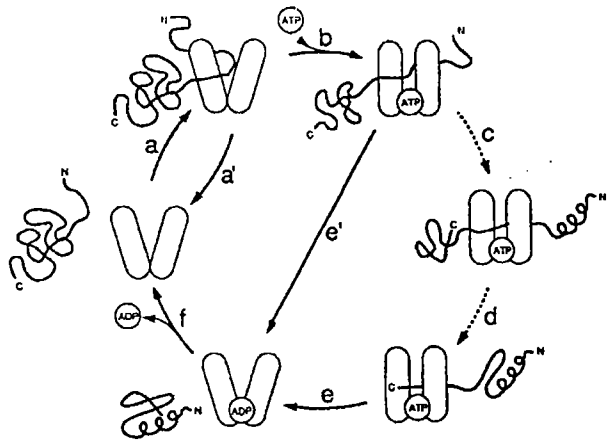

Fig. 2. A schematic model for hsp70 family members (i) binding to protect unfolded regions and (ii) catalysing protein unfolding by plucking or threading. Two ovals represent the two lobes seen in the structure of the ATP binding domain of Hsc70. a, Exposed chain of misfolded protein bound, recognized by backbone hydrogen-bonding groups; a', release of bound chain; b. ATP binds, causing formation of tighter complex with misfolded chain, resulting in unfolding of local surrounding chain (plucking); c and d, chain threads, perhaps completely from N- to C-terminus; e, ATP hydrolysed after threading, allowing chain release; e', ATP hydrolysed after binding without threading, allowing chain release (repeated cycles of binding and dissociation ('plucking') can have an unfolding effect very similar to that of threading); f, ADP release to return to initial state. Continuous arrows indicate fast transitions, dashed indicate slow ones. The selection of path (b) instead of (a') may be as a result of interactions between Hsp70 molecules, bound to neighbouring stretches of protein chain. Hsp70 may form higher-order structures perhaps in a way analogous to actin filaments. There is indirect evidence that Hsp70 may function at least as a dimer, since *in vivo* some Hsp70 mutants are less deleterious when co-expressed than when individually expressed. This could be explained if the mutations were on interacting molecular surfaces of a protein complex and complemented each other (C.Nicolet and E.Craig, personal communication). After b, in the case of 'threading', path c would be followed, whereas for 'plucking' e' would be followed.

Active unfolding by 'plucking'

An alternative to threading is a process in which the ATP form of the chaperone binds a polypeptide and releases it after hydrolysis (without threading), perhaps going through many cycles of encounter, binding and release (repeated binding). After each encounter, the probability for being nearer the unfolded state would be enhanced, e.g. by 'straightening out' a backbone segment of several residues induced by a conformational change of the chaperone resulting from ATP hydrolysis (Pelham, 1986). The conformational change may be aided by interaction between two or more bound chaperones. The principal difference between the plucking and the threading processes would be in the details of on−off rates and in the sequential nature of threading.

An unfoldase activity, by either mechanism, can also help explain the observed function of hsp70 family members of disrupting tight protein complexes. Proteins with regions of exposed backbone in their natural folded state would be recognized and unfolded as if they were misfolded proteins. This would fit with the specific recognition of the hinge region peptide of clathrin (presumably exposed) by Hsc70 (DeLuca-Flaherty *et al.*, 1990).

Backbone recognition and threading in oligomeric assembly

Can chaperone action by GroEL be explained in terms of backone recognition and threading? GroEL has been suggested to provide surfaces similar to the faces of the missing subunits. However, while contacts between subunits of oligomers are fairly

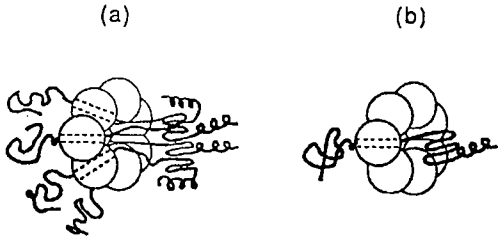

Fig. 3. A simplified diagram of two alternative models for groEL family members catalysing unfolding by threading, providing conditions for several chains to fold and oligomerize. (a) The symmetry of the 14mer suggests that up to seven chains could be actively refolded or threaded by GroE at the same time, facilitating oligomer assembly by simple spatial proximity of the emerging chains. (b) The central hole of the 14mer is suggestive of a channel for the accommodation and threading of a single chain. Oligomeric assembly could still be facilitated indirectly through the presentation of prefolded monomers. (GroEL molecules are made up of 14 subunits, but only seven units are shown for clarity.)

hydrophobic, interactions with GroEL may not be predominantly so (Ostermann *et al.*, 1989). Backbone recognition of incompletely folded chains is therefore plausible.

The oligomeric structure of GroEL suggests a possible mechanism for enhancing oligomeric assembly through threading. The co-operative binding of several unfolded chains might create the effective local concentration and conformation suitable for oligomerization (Figure 3). Monomers that aggregated incorrectly would remain substrates for repeated unfolding.

Unlike DnaK of the hsp70 family, GroEL does not seem able to facilitate the *in vitro* refolding of proteins directly from aggregates (Buchner *et al.*, 1991). In mitochondria it now seems clear that hsp70 associates with translocated chains before they are passed on to the groEL family member Hsp60, suggesting that hsp70 performs more of an unfoldase function while groEL directs proper oligomerization (Ostermann *et al.*, 1989).

It is not clear how far the analogy between the groEL and hsp70 families can be extended. In spite of similarities in function, we have not been able to identify significant sequence similarities between them (T.J.P.Hubbard and S.Sander, unpublished results). An important part of the ATP-binding site common to actin, hexokinase and Hsp70 is a pair of pseudo-symmetrical β-hairpins (residues 6–20 and 195–209 in the hsc70 ATPase domain X-ray structure, code 1HSC in the Protein Bank Data Base of the Brookhaven National Laboratory, New York). Their sequences are highly conserved within each family, contain glycine-rich turns and at least the first hairpin in each sequence has a conserved aspartic acid residue. We find two conserved regions in GroEL sequences containing similar features (residues 79–93 and 407–421 in *E.coli* GroEL sequence, code GROL$ECOLI in the protein sequence database SWISSPROT of A.Bairoch, University of Geneva and EMBL Data Library, Heidelberg), but without any homology to the Hsp70 in the surrounding sequence. Whether this indicates any structural or functional homology is unclear. Interestingly, a single conserved region of this type is also found in another family of heat-shock proteins, Hsp90 (residues 76–90 in *E.coli* Hsp90, code HTPG$ECOLI in SWISSPROT). Hsp90 is known to be a dimer and has been shown to be part of the signal transduction pathway for steroid hormones (Picard *et al.*, 1990).

Discussion

Why does the cell need help in folding? The experimental evidence for a role for chaperones in protein unfolding and refolding in the cell is now overwhelming. However, we see no challenge to the basic belief that the sequence of a protein directly determines its specific folded conformation. Models in which chaperones influence the protein-folding pathway through different binding specificities to different parts of a sequence (Rothman, 1989) appear too complicated to us in that they would require very many different types of chaperone proteins. Our simpler mechanisms are consistent with the view that folding aids are needed because correct folding, although reprogrammed in the sequence, is difficult to achieve or maintain in the crowded cellular interior. Natural sequences, optimized in evolution for activity and physiological environment as well as folding, may not provide an unambiguous pathway to the correct folded state. The cellular compromise, as in many other processes, may be to recognize the errors and correct them, rather than to perfect the folding pathway of each protein. Mechanisms to correct errors in protein folding would also make cells much less susceptible to genetic folding defects (Van Dyk *et al.*, 1989).

Experimental tests

What can be done to experimentally prove or disprove the proposed mechanism? Protein chains which apparently cannot fold correctly have been shown to remain associated with chaperone molecules (Beckmann *et al.*, 1990; Kassenbrock *et al.*, 1988). It would be interesting to determine if this association is static or dynamic. If chaperones are unfoldases, their activity should be independent of the ability of a protein substrate to fold correctly. Models involving either threading or repeated binding and release of misfolded chains imply that the association is dynamic, with free chaperone exchanging with bound chaperone continuously. This might be tested by determining the rate of exchange of chaperone between two similar systems of chaperones with bound substrates. If each system used a different protein substrate and started with either labelled or unlabelled chaperone, the occurrence of the different possible complexes after mixing might be monitored. The rates of exchange and of ATP consumption could be compared for substrates able or unable to fold correctly.

With a threading activity, the proximity of parts of the chain to a chaperone molecule might be expected to vary with time. Cross-linking experiments, used to show the interaction of the N-terminus of unfolded proteins with GroEL (Bochkareva *et al.*, 1988), might be extended to identify any time dependence of association with the chaperone. If threading occurs it might also be blocked by a protein chain with a large chemical group bound to it, as was observed for the mitochondrial import of dihydrofolate reductase with bound methotrexate (Eilers and Schatz, 1986). If such an inactivated complex can be generated, electron microscopy might reveal a complex of a globular protein with two tails of partially unfolded protein. If inactivated complexes of oligomeric GroEL chaperones could be generated in this way, the number of chains simultaneously bound might be determined by subsequent cross-linking. Crystallography on complexes between chaperones and bound peptides would, of course, provided a detailed view of the modes of binding.

SecB, trigger factor and SRP

If active unfolding, backbone recognition and threading do exist, how widespread are these processes? It is interesting to note the similarities between the post-translocation pathways into the mitochondria and ER involving Hsc70; the pathway into the outer cell membrane of *E.coli* involving SecB and trigger factor (Lecker *et al.*, 1989); and the co-translational pathway involving the signal recognition particle (SRP) into the ER. Each pathway requires a mechanism for recognizing proteins destined for translocation. It has been shown that recognition can be independent of the

leader peptide, since SecB recognizes proteins destined for export when unfolded, even when the leader peptide is absent (Randall *et al.*, 1990). Although SecB (an oligomer) does not appear to bind ATP, unlike the hsp70 and groEL families, it may bind exposed backbone. Tight binding to unfolded regions would temporarily prevent refolding (Hardy and Randall, 1991). Rather than act as recognition targets, leader peptides may modulate protein folding, by making it more difficult to fold correctly or by causing an alternative folding pathway leading to a misfolded state. This would be a counterexample to the requirement of a pro-sequence for the correct folding of subtilisin (Zhu *et al.*, 1989), but together these examples imply that the folding pathway can be directed by the N-terminal (leader) sequence and that folding may need to proceed in a sequential way.

How did chaperones evolve? Probably they first evolved as an aid to membrane translocation in primitive cells and were later used as an aid in protein refolding in difficult circumstances. The systemic costs of having a limited number of chaperones is presumably lower than that of optimizing the stability of many proteins against a variety of environmental 'hazards'. The existence of chaperones has important practical consequences. In protein engineering, where overexpressed recombinant proteins frequently do not fold correctly but form inclusion bodies (e.g. Mitraki and King, 1989), several authors have suggested that simultaneous overexpression of a chaperone protein may be helpful (e.g. Ellis, 1990). Protein folders may draw the conclusion that folding pathways are less stable with respect to perturbation of external conditions than is the native protein structure and that therefore prediction of just the native structure may be easier than simulation of correct folding pathways. In cell biology, many more fascinating details of the action of chaperone proteins and their interaction with other proteins remain to be discovered. The mechanisms proposed here, while fitting current experimental evidence, are unproven. The detailed mechanism of chaperone action will have to be determined by experiment. It is hoped that this discussion will stimulate specific experiments to test for active unfolding, backbone recognition and threading.

Recent experiments: refolding by GroEL/ES

As this paper was about to go to press, an elegant series of experiments by Martin *et al.* (1991) was published, shedding light on details of the refolding reaction catalysed by GroE. In these experiments the reactivation of the unfolded chain of the enzymes rhodanese and dihydrofolate reductase is measured in the presence of GroEL, GroES, MgATP and casein (a competitor of the refolded substrate), and various combinations thereof. Several remarkable facts emerge which include the following:

(i) Use of ATP is 130 ± 20 molecules per reactivated rhodanese enzyme (293 residues), much higher than previously reported for chaperones.

(ii) The enzyme stays bound to the GroEL/ES complex during the refolding reaction (no significant displacement by casein).

(iii) The refolding reaction requires GroEL, GroES, and MgATP. Without GroES, there can be repeated cycles of release and binding, without any beneficial effect on refolding.

(iv) The refolding of dihydrofolate reductase, which can take place spontaneously, i.e. without GroE, is slowed down by GroEL/ES and even more so by GroEL alone.

The following picture emerges from these observations. Partly folded protein chains bind and stick to GroEL. With MgATP present, they are released at a higher rate and can rebind, but there is no active unfolding by repeated binding and release. The GroEL/ES complex, however, not only binds but also actively refolds proteins, releasing them only in a properly folded or folding-competent state. This active refolding process may be threading, starting from the point of attachment and proceeding to the end of the chain with an energy usage of one ATP molecule per residue threaded. Threading would lead to proper folding in that it brings the chain into a locally extended conformation from which folding can proceed without aggregation or misfolding, along a folding pathway analogous to the one presumably followed at the ribosome during protein synthesis.

The observed energy usage for rhodanese presents an interesting opportunity for a quantitative test of the threading hypothesis. Suppose that the sites of attachment to GroE are approximately randomly distributed along the chain of the enzyme. Then, on average, one-half of the chain is threaded per reactivated enzyme, so that the energy usage measured by Martin *et al.* amounts to $(130 \pm 20)/(293/2) = 0.89 \pm 0.14$ or about one molecule of ATP per residue. This numerical agreement may be fortuitous, but suggests a series of experiments in which energy usage is measured for enzymes of different size. From the threading model one would predict that usage of ATP is proportional to chain length.

Acknowledgements

We thank Paul Travers, Haruki Nakamura, Tom Blundell, Hugh Pelham, Alfonso Valencia and Ken Holmes for useful discussions. T.J.P.H. is grateful to the Protein Engineering Research Institute for facilities and the European Commission (EC) for financial support and C.S. to the Human Frontiers Science Program for support.

References

Afinsen,C.B. (1973) *Science*, **181**, 223–230.
Baumann,G., Froemmel,C. and Sander,C. (1989) *Protein Engng*, **2**, 329–334.
Beckmann,R.P., Mizzen,L.E. and Welch,W.J. (1990) *Science*, **248**, 850–854.
Berg,O.G., Winter,R.B. and von-Hippel,P.H. (1983) In Hunt,T., Prentis,S. and Tooze,J. (eds), *DNA makes RNA make Proteins*. Elsevier Biomedical Press, Amsterdam, pp. 3–9.
Blumberg,H. and Silver,P.A. (1991) *Nature*, **349**, 627–630.
Bochkareva,E.S., Lissin,N.M. and A.S.,G. (1988) *Nature*, **336**, 254–257.
Buchner,J., Schmidt,M., Fuchs,M., Jaenicke,R., Rudolph,R., Schmid,F.X. and Kiefhaber,T. (1991) *Biochemistry*, **30**, 1586–1591.
Chandrasekhar,G.N., Tilly,K., Woolford,C., Hendrix,R. and Georgopoulos,C. (1986) *J. Biol. Chem.*, **261**, 12414–12419.
Chappell,T.G., Welch,W.J., Schlossman,D.M., Palter,K.B., Schlesinger,M.J. and Rothman,J.E. (1986) *Cell*, **45**, 3–13.
Cheng,M.Y., Hartl,F.U., Martin,J., Pollock,R.A., Kalousek,F., Neupert,W., Hallberg,E.M., Hallberg,R.L. and Horwich,A.L. (1989) *Nature*, **337**, 620–625.
Cheng,M.Y., Hartl,F.U. and Horwich,A.L. (1990) *Nature*, **348**, 455–458.
Chirico,W.J., Walters,M.G. and Blobel,G. (1988) *Nature*, **332**, 805–810.
Craig,E.A., Kramer,J. and Kosic-Smithers,J. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 4156–4160.
Craig,E.A., Kramer,J., Shilling,J., Werner,W.M., Holmes,S., Kosic,S.J. and Nicolet,C.M. (1989) *Mol. Cell. Biol.*, **9**, 3000–3008.
DeLuca-Flaherty,C., McKay,D.B., Parham,P. and Hill,B.L. (1990) *Cell*, **62**, 875–887.
Deshaies,R.J., Koch,B.D., Werner-Washburne,M., Craig,E.A. and Schekman,R. (1988) *Nature*, **332**, 800–805.
Eilers,M., Hwang,S. and Gottfried,S. (1988) *EMBO J.*, **7**, 1139–1145.
Eilers,M. and Schatz,G. (1986) *Nature*, **322**, 228–232.
Ellis,J. (1987) *Nature*, **328**, 378–379.
Ellis,J. (1990) *Nature*, **345**, 484.
Flaherty,K.M., DeLuca,F.C. and McKay,D.B. (1990) *Nature*, **346**, 623–628.
Flaherty,K.M., McKay,D.B., Kabsch,W. and Holmes,K.C. (1991) *Proc. Natl Acad. Sci. USA*, in press.
Flynn,G.C., Chappell,T.G. and Rothman,J.E. (1989) *Science*, **245**, 385–390.
Gaitanaris,G.A., Papavassiliou,A.G., Rubock,P., Silverstein,S.J. and Gottesman,M.E. (1990) *Cell*, **61**, 1013–1020.
Goloubinoff,P., Christeller,J.T., Gatenby,A.A. and Lorimer,G.H. (1989a) *Nature*, **342**, 884–889.

Goloubinoff,P., Gatenby,A. and Lorimer,G.H. (1989b) *Nature*, **337**, 41–47.
Guidon,P.T. and Hightower,L.E. (1986) *Biochemistry*, **25**, 3231–3239.
Hardy,S.J.S. and Randall,L.L. (1991) *Science*, **251**, 439–443.
Hemmingsen,S.M., Wolldord,C., Can der Vies,S.M., Tilly,K., Dennis,D.T., Georgopoulos,C.P., Hendrix,R.W. and Ellis,R.J. (1988) *Nature*, **333**, 330–334.
Heuser,J. and Steer,C.J. (1989) *J. Cell. Biol.*, **109**, 1457–1466.
Hwang,D.S., Crooke,E. and Kornberg,A. (1990) *J. Biol. Chem.*, **265**, 19244–19248.
Jaenicke,R. (1987) *Progr. Biophys. Mol. Biol.*, **49**, 117–237.
Kabsch,W., Mannherz,H.G., Suck,D., Pai,E.F. and Holmes,K.C. (1990) *Nature*, **347**, 37–44.
Kang,P.J., Ostermann,J., Shilling,J., Neupert,W., Craig,E.A. and Pfanner,N. (1990) *Nature*, **348**, 137–143.
Kassenbrock,C.K., Garcia,P.D., Walter,P. and Kelly,R.B. (1988) *Nature*, **333**, 90–93.
Kusukawa,N., Yura,T., Ueguchi,C., Akiyama,Y. and Ito,K. (1989) *EMBO J.*, **8**, 3517–3521.
Kusukawa,N. and Yura,T. (1988) *Genes Dev.*, **2**, 874–882.
Laminet,A.A., Ziegelhoffer,T., Georgopoulos,C. and Pluckthun,A. (1990) *EMBO J.*, **9**, 2315–2319.
Laskey,R.A. and Earnshaw,W.C. (1980) *Nature*, **286**, 763–767.
Lecker,S., Lill,R., Ziegelhoffer,T., Georgopoulos,C., Bassford,P.J., Kumamoto,C.A. and Wickner,W. (1989) *EMBO J.*, **8**, 2703–2709.
Martin,J., Langer,T., Boteva,R., Schramel,A., Horwich,A.L. and Hartl,F.U. (1991) *Nature*, **352**, 36–42.
Mitraki,A. and King,J. (1989) *Bio/Tech.*, **7**, 690–697.
Neidhart,F.C., VanBogelen,R.A. and Vaughn,V. (1984) *Annu. Rev. Genet.*, **18**, 295–329.
Neupert,W., Hartl,F.U., Craig,E.A. and Pfanner,N. (1990) *Cell*, **63**, 447–450.
Normington,K., Kohno,K., Kozutsumi,Y., Gething,M.-J. and Sandbrook,J. (1989) *Cell*, **57**, 1223–1236.
Ostermann,J., Horwich,A.L., Neupert,W. and Hartl,F.U. (1989) *Nature*, **341**, 125–130.
Pelham,H.R.B. (1986) *Cell*, **46**, 959–961.
Phillips,G.J. and Silhavy,T.J. (1990) *Nature*, **344**, 882–884.
Picard,D., Klursheed,B., Garabedian,M.J., Fortin,M.G., Lindquist,S. and Yamamoto,K.R. (1990) *Nature*, **348**, 166–168.
Randall,L.L., Topping,T.B. and Hardy,S.J. (1990) *Science*, **248**, 860–863.
Reading,D.S., Halberg,R.L. and Myers,A.M. (1989) *Nature*, **337**, 655–659.
Rippmann,F., Taylor,W.R., Rothbard,J.B. and Green,N.M. (1991) *EMBO J.*, **10**, 1053–1059.
Rothman,J.E. (1989) *Cell*, **59**, 591–601.
Rothman,J.E. and Kornberg,R.D. (1986) *Nature*, **322**, 209–210.
Rothman,J.E. and Schmid,S.L. (1986) *Cell*, **46**, 5–9.
Scherer,P.E., Krieg,U.C., Hwang,S.T., Vestweber,D. and Schatz,G. (1990) *EMBO J.*, **9**, 4315–4322.
Skowyra,D., Georgopoulos,C. and Zylicz,M. (1990) *Cell*, **62**, 939–944.
Van Dyk,T., Gatenby,A.A. and LaRossa,R.A. (1989) *Nature*, **342**, 451–453.
Vogel,J.P., Misra,L.M. and Rose,M.D. (1990) *J. Cell Biol.*, **110**, 1885–1895.
Wickner,S., Hoskins,J. and McKenney,K. (1991) *Nature*, **350**, 165–167.
Zhu,X., Ohta,Y., Jordan,F. and Inouye,M. (1989) *Nature*, **339**, 483–484.
Zylicz,M., Ang,D., Liberek,K. and Georgopoulos,C. (1989) *EMBO J.*, **8**, 1601–1609.

Received on October 27, 1989; accepted on July 31, 1991