The EMBO Journal Vol.17 No.18 pp.5251–5254, 1998

EMBO WORKSHOP REPORT

Protein folding and misfolding inside and outside the cell

Christopher M.Dobson and R.John Ellis

Oxford Centre for Molecular Sciences, University of Oxford, South Parks Road, Oxford OX1 3QT, UK

e-mail: chris.dobson@chem.ox.ac.uk or je@dna.bio.warwick.ac.uk

The workshop was held at St Catherine's College, Oxford, from March 25–28, 1998, and attracted participants from 32 nations. Protein folding is one of the most important processes in biology since it adds functional flesh to the bare bones of genes, but it has traditionally been studied by people separated both intellectually and physically because they are training in different disciplines. The aim of the meeting was to bring together chemists and structural biologists studying how pure, denatured proteins refold spontaneously in the test tube, with biochemists and cell biologists who are concerned with how proteins fold inside living cells and medical scientists interested in the diseases that result when this process goes wrong. In this report we concentrate on general concepts and themes rather than on detailing every contribution.

Many studies have established that the vast majority of denatured protein chains are capable of refolding spontaneously to the correctly folded conformation in the absence of either other macromolecules or energy expenditure. Chris Dobson (Oxford, UK) summarized the increasingly sophisticated physical techniques used to study protein refolding, and stressed the 'new view' of this process as a three-dimensional, downhill energy search by a vast array of different initial conformations that converge by different routes on the unique functional structure. He introduced the fact that such techniques are now becoming applicable to study folding in cell-free translation extracts which are much closer to the intracellular environment than are pure proteins refolding from the denatured state. For example, it is now possible to obtain mass spectra from intact ribosomes and to characterize particular protein components from the spectrum.

There are several possible fates of newly synthesized protein chains inside cells. The major distinction between these fates is whether the chains succeed in folding correctly, or whether the chains aggregate. Aggregation has commonly been regarded as a nuisance which affects *in vitro* protein refolding studies; it is now apparent that aggregation is also a problem for cells. In the intracellular environment, the competition between folding, aggregation and degradation determines whether a polypeptide chain can achieve its functional state with the efficiency required for successful cell growth, or whether it aggregates into a state that causes cellular damage and even death.

John Ellis (Warwick, UK) reviewed evidence that aggregation is a specific process, which may be amplified by the high concentration of identical nascent chains emerging from polysomes, and by the very large increases in association constants produced by the crowding effect of the high concentration of macromolecules in the cytoplasm (340 mg/ml in *Escherichia coli*), an effect yet to be extensively studied on protein refolding *in vitro*. Combating aggregation is one of the major roles of molecular chaperones, of which there are at least 20 structurally distinct families. It is important to appreciate that protein folding occurs in several different intracellular compartments, especially in eukaryotic cells, and that the chaperone complement differs between these compartments. Thus, proteins coevolve with particular chaperones, and for meaningful *in vitro* experiments it is advisable to choose naturally occurring protein–chaperone combinations. A major theme of this workshop was the discussion of the best conceptual and methodological approaches for determining the precise basis of how cells contrive to optimize correct protein folding and reduce aggregation.

How do denatured proteins refold in the test tube?

It has been evident for many years that the sequence of a protein defines its three-dimensional fold. The question of how an unstructured (random coil) polypeptide can rapidly and efficiently find its appropriate fold from the countless alternatives is, however, a problem that has perplexed the scientific community for many years. Considerable progress in understanding this remarkable process has been made recently through a combination of theoretical and experimental advances.

A particularly important theoretical strategy has been to simulate refolding by using 'lattice models' for proteins in which residues are represented as points on a three-dimensional lattice that interact with one another according to defined potential functions. The idea is to devise models simple enough for extensive calculations to be carried out to simulate refolding, yet sufficiently complex to encapsulate key features of real proteins. Martin Karplus (Cambridge, MA) described the results of simulations using such models.

In order to fold successfully, a polypeptide chain must collapse, a process favoured by the burial of hydrophobic sidechains, while forming key contacts between residues which ensure that the native fold is formed efficiently. The simulations suggest that this process can occur for sequences where the formation of native-like interactions stabilizes the folding chain, and where stable misfolded states, which can act as kinetic traps, are avoided. As well as providing insights into the folding process, such simulations allow the results of experimental studies to

© Oxford University Press

be interpreted more fully and permit the rational design of new experimental strategies; the experiments in turn can be used to test and improve the simulations.

A variety of experimental techniques has been developed to study refolding. As well as kinetic experiments, reviewed by Christopher Dobson, a complementary approach is to adjust the solution conditions to generate stable analogues of species likely to be important in the kinetic refolding process. NMR spectroscopy is particularly important in these studies because of its ability to provide structural and dynamical information at the level of individual residues. Peter Wright (La Jolla, CA) described studies of myoglobin where a variety of partially folded states can be stabilized and have been characterized in detail. These experiments have been able to map the development of stable native-like secondary structure, and the reduction in conformational flexibility as the compactness of the protein increases. Further insights into these issues come from molecular dynamics simulations of partially folded states of proteins; Lorna Smith (Oxford, UK) described the results of such approaches with the proteins lysozyme and α-lactalbumin. An important conclusion of this work that correlates well with the experimental data is that the overall fold of a protein can form prior to the generation of specific close-packing of residues that is characteristic of the native protein.

Determining the relationship between sequence and structure is a key aspect of understanding folding, but it is also crucial for the design of novel sequences with specific properties. Luis Serrano (Heidelberg, Germany) described an example of a *de novo*-designed triple-stranded β-sheet, composed of 16 residues. Sheena Radford (Leeds, UK) described studies using biophysical techniques of the refolding of proteins with different folds, e.g. proteins that are either largely helical or largely sheet. It is now possible to begin to address the issue of the way that evolution has selected a limited, although still large, number of possible folds for polypeptide chains; whether this selection is for ease of folding rather than for stability or functional value is not clear. Oleg Ptitsyn (Bethesda, MD) suggested that key residues in the globin and cytochrome sequences may be conserved in evolution to ensure that rapid folding occurs to a specific structure. These ideas relate closely to those discussed above in terms of simulations, and have also emerged from the elegant protein engineering experiments of Alan Fersht.

In studies of protein refolding, a number of characteristics that result in slow steps and potential misfolding have been recognized. Prominent among these is the need for some proteins to form the correct isomer of peptide bonds involving proline, and the need to form the correct disulfide bonds between cysteine residues. Such problems arise for protein folding in cells, and enzymes exist that catalyse such steps. Robert Freedman (Canterbury, UK) discussed one of the best known such enzymes, protein disulfide isomerase; this enzyme allows the exchange of disulfide bonds among folding chains until the lowest energy state is reached, a feature reminiscent of the earlier steps in folding described above. This topic formed the introduction to sessions concerned with protein folding inside cells.

How do proteins fold inside cells?

As well as enzymes that isomerize covalent bonds in protein chains, cells contain a variety of molecular chaperones that control and assist the folding process.

Previous work suggests that two types of chaperone act sequentially on newly synthesized polypeptides in both the cytoplasm of prokaryotic cells and in the cytosol and mitochondria of eukaryotic cells. Small chaperones ($<$100 kDa), such as hsp70 (DnaK) and hsp40 (DnaJ), bind to hydrophobic regions on nascent chains to prevent aggregation and premature folding as elongation continues, while large chaperones ($>$800 kDa), such as GroEL, bind complete, partially folded chains individually in a central cage, where folding proceeds further until the danger of aggregation with similar chains has passed. Some aspects of these views were confirmed and extended, while others were challenged.

Elizabeth Craig (Madison, WI) reported that of the 14 hsp70-like proteins in yeast, it is the two cytosolic SSB proteins, but not the four cytosolic SSA proteins, that bind to both nascent chains and ribosomes. Binding is independent of ATP and occurs even to chains as short as 70 residues, 30–40 of which are buried in the ribosome. There are least twice as many SSB protein molecules as ribosomes in yeast, so in principle every nascent chain could have one SSB attached, but this has not yet been established. On the other hand, Bernd Bukau (Freiburg, Germany) was unable to demonstrate the binding of DnaK or DnaJ to nascent chains in *E.coli*, but presented genetic evidence that a major role of these chaperones in this organism is to assist the refolding of proteins unfolded by heat stress. Other chaperones such as hsp90 may bind to at least some types of newly synthesized chain. Johannes Buchner (Regensburg, Germany) reported that hsp90 has two distinct chaperone sites; binding by the N-terminal fragment is ATP-dependent and prefers unstructured peptides, while binding by the C-terminal fragment is ATP-independent and prefers partially folded polypeptides.

There is evidence that the GroEL of *E.coli* binds *in vivo* to only ~10–15% of all the newly synthesized cytoplasmic chains under normal growth conditions. Arthur Horwich (New Haven, CT) and Ulrich Hartl (Martinsried, Germany) independently reported the identification of some of the natural substrates for this chaperone under such conditions; these include GroEL itself, the three elongation factors, the α chain of RNA polymerase, E3 from pyruvate dehydrogenase and the β subunit of the F_1 ATPase. Hartl reviewed his folding shift hypothesis that proposes an evolutionary shift from a predominantly post-translational type of protein folding in prokaryotes to a predominantly co-translational type of protein folding in eukaryotes. Such a shift could provide the basis for the appearance in eukaryotes of large modular proteins via gene fusion events. In support of this view, Hartl reported that the expression in yeast of a mutant form of GroEL that binds partially folded chains but cannot release them does not affect the growth rate. This observation is consistent with the rapid co-translational folding of most newly synthesized chains that are released from the eukaryotic ribosome in a state not recognized by GroEL. In contrast, the expression of the mutant GroEL in *E.coli* reduces the growth rate, presumably because ~50% of the newly

synthesized but only partially folded chains bind irreversibly to the mutant GroEL. Boyd Hardesty (Austin, TX) used cell-free translation extracts to show that for newly synthesized rhodanese to fold correctly DnaJ, DnaK + GrpE and then GroEL/ES, must be added in that order, confirming the view that the small chaperones act before the large chaperones. He also presented interesting new data suggesting that the elongation factors Tu and Ts may also act as chaperones since they assist the refolding of denatured rhodanese, provided that GTP is present.

Both Horwich and Hartl support the folding cage model for GroEL action, and Hays Rye (New Haven, CT) presented additional elegant fluorescence energy transfer data in favour of this model, but this view was challenged by Alan Fersht (Cambridge, UK) on the grounds that an apical fragment of GroEL (residues 193–335) complements a temperature-sensitive mutant of *E.coli* at 43°C and enhances the activity of co-expressed GroEL in lethal GroEL knockouts. Whether this complementation requires GroES, the other component required for the cage to function, is not known, and the apical fragment is unable to act like the wild-type GroEL as the only source of GroEL in the cell at permissive temperature. Cell viability rather than growth rates were measured in this study, so the cage might be an efficiency-enhancing device essential for cells to compete in nature rather than an absolute requirement under all conditions. However, it should be noted that in this study the cells always contain at least some GroEL cages.

Helen Saibil (London, UK) presented her latest 9 Å cryoEM pictures of GroEL in action, and stated that she is now certain that the C-terminus forms a barrier between the two rings; this was complemented by a report by Keith Willison (London, UK) of similar large ATP-induced domain movements in the GroEL equivalent in the eukaryotic cytosol, CCT. Willison also suggested that CCT functions as an ATP-loading machine for its main substrate, actin, rather than as a folding cage, since nucleotide-free actin denatures irreversibly at a high rate while actin peptides that bind to CCT are mostly from the surface of the folded molecule.

Instead of folding in the cytosol, an important subset of proteins fold and are glycosylated after transport into the endoplasmic reticulum; this compartment lacks large chaperones of the GroEL type but contains calnexin and calreticulin that chaperone the folding of glycosylated proteins. Ari Helenius (Zürich, Switzerland) studied the effect of glucosidase inhibitors on the *in vivo* folding of a temperature-sensitive mutant of the VSV G protein and concluded that it is the glucose residues that determine its binding to calnexin rather than protein–protein interactions. Ineke Braakman (Amsterdam, The Netherlands) reported that the gp160 envelope protein of HIV undergoes extensive but slow post-translational folding in the ER, the signal peptide being removed only after synthesis is complete.

Walter Neupert (Munich, Germany) reviewed the evidence for the view that unidirectional protein transport into yeast mitochondria requires mitochondrial hsp70 acting as a molecular ratchet; this mechanism is proposed to result in the unfolding of the translocating protein on the cytosolic side of the outer mitochondrial envelope. He also reported the identification of two chaperones called TIM 10 and TIM 12 in the intermitochondrial membrane space that aid translocation, perhaps by functionally replacing hsp70 which is absent from this compartment. It seems that the list of proteins acting as molecular chaperones is destined to grow still further.

How is protein misfolding linked to disease?

It is increasingly clear that protein folding is not only an essential feature of the conversion of genetic information into biological activity, but is also a key feature in the control and localization of this activity. This conclusion leads naturally to the idea that the failure of proteins to fold, or to fold into an incorrect structure, can be a cause of disease. Cystic fibrosis is an example of a genetic disease where a variant protein (CFTR) is unable to fold correctly to a stable state in the endoplasmic reticulum and fails to reach the plasma membrane, eventually being degraded. Philip Thomas (Dallas, TX) and John Riordan (Scottsdale, AZ) discussed the problem of the misfolding and incorrect trafficking of CFTR mutant proteins and their links with the molecular pathology of the disease. Interestingly, even the wild-type chains do not fold with high efficiency; only ~30% of wild-type chains survive the quality-control mechanisms of the endoplasmic reticulum. It is perhaps not surprising that a wide range of mutations in different regions of the CFTR protein result in significantly reduced levels of activity in sufferers of this condition.

We have noted above that one of the roles of molecular chaperones is to prevent the aggregation of partially folded proteins. The remaining talks at the Workshop focussed on issues of aggregation which, once ignored as a topic of serious study, is now elevated to almost cult status. Jonathan King (Cambridge, MA) described elegant studies showing that misfolded intermediates result in the formation of inclusion bodies in the case of the trimeric phage P22 tail spike protein, a trimeric protein. Interestingly, there is evidence that completion of folding of the protrimer, both *in vivo* and *in vitro*, requires interchain disulfide bond formation, even although the native trimer has no such bonds. Anthony Fink (Santa Cruz, CA) showed how biophysical studies can characterize the structural properties of aggregated proteins, while Jean Baum (Piscataway, NJ) discussed remarkable real-time NMR experiments probing the molecular basis of misfolding of collagen mutants that cause *osteogenesis imperfecta*. Subsequent speakers concentrated on the aggregation of proteins to form amyloid fibrils and plaques.

Amyloid formation is associated with some 20 sporadic, genetic or infectious diseases; remarkably these fibrils have similar morphologies, despite their origin from unrelated polypeptides. Mark Pepys (London, UK) reviewed this topic, pointing out that one of these diseases, Alzheimer's, was estimated to be the most expensive medical problem in the Western world. Byron Caughey (Hamilton, MT) discussed the spongiform encephalopathies, including BSE, scrapie, and CJD, which, of course, are currently of great concern. Max Perutz (Cambridge, UK) described fibrils in Huntington's disease which have many of the characteristics of amyloid. A key issue in such diseases is the mechanism of conversion of soluble proteins into insoluble aggregates. The meeting started with the idea

that protein folding mechanisms are amenable to both theoretical analysis and investigation by biophysical methods; it concluded by discussing how aggregation could also be studied at the molecular level by similar approaches.

A variety of approaches are being applied to this problem. Carol Robinson (Oxford, UK) described incisive experiments using time-of-flight mass spectrometry to probe the nature of amyloidogenic folding intermediates, while Valerie Daggett (Seattle, WA) described molecular dynamics simulations to probe the early steps in the structural conversion associated with these proteins. David Eisenberg (Los Angeles, CA) described crystallographic studies of oligomeric proteins generated by domain swapping that could represent at least the initial events in the structural conversions of some proteins, while Perutz described a possible zipper mechanism. The possibility of a detailed molecular view of amyloid structure was raised by dramatic pictures of fibrils produced from an SH3 domain. These pictures were generated using cryoelectron microscopic image reconstruction techniques and resulted from a collaboration between the groups of Saibil and Dobson.

Will all of this effort give rise to practical benefits in terms of therapeutic treatment? There is justification for some optimism in this area. Pepys discussed several different approaches, including suppression of the production of amyloid precursor, prevention of amyloid formation and stimulation of amyloid degradation. Jeffery Kelly (La Jolla, CA) described strategies to develop drugs to treat amyloidosis resulting from mutations in the transthyretin gene. The idea here is to stabilize the tetrameric form of the protein using analogues of its natural ligand, thyroxine. Caughey described peptides of the prion protein that inhibit the conversion of the full length protein to its amyloidogenic form, while Pepys outlined approaches to tackling amyloid diseases in general by inhibiting SAP, a protein which appears to stabilize the fibrils against degradation. Most encouragingly, Pepys reported that SAP-minus mice show reduced amyloid fibril formation, and has identified a compound ‘R’ that can strip SAP from the fibrils. These are indications that the long haul from ‘theory to therapy’ has begun.