What's new in protein folding? EMBO Workshop: Protein folding and misfolding inside and outside the cell

Sheena E Radford

Address: School of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK.
E-mail: s.e.radford@leeds.ac.uk

Folding & Design 01 June 1998, **3**:R59–R63
http://biomednet.com/elecref/13590278003R0059

An EMBO workshop entitled 'Protein folding and misfolding inside and outside the cell' was held at St Catherine's College, Oxford, UK, from 24th to 28th March this year. The meeting, organised by Chris Dobson (University of Oxford), Chris Leaver (University of Oxford) and John Ellis (University of Warwick), was attended by 202 scientists and involved more than 30 lectures, as well as talks chosen from posters and the usual array of poster sessions. The aim of the workshop was to embrace the field of protein folding in its widest perspective, with talks ranging from simulations of protein folding, to potential therapies for folding diseases. The aim of this report is to summarise for readers of *Folding & Design* some of the highlights of the meeting and, in particular, to comment on the common themes that emerged from the diverse range of topics embraced. It is not my aim to summarise individual presentations in detail and I apologise to those whose work is not discussed here or is mentioned only briefly. Instead, I shall try to highlight common concepts and exciting breakthroughs, divided into the three themes of the workshop: spontaneous refolding; folding in intact cells; and folding and disease.

Spontaneous refolding

The latest view of protein folding at the molecular level was the first theme of the meeting. In his introductory overview of this area, Chris Dobson summarised the state-of-the-art in experimental methods that can be used to delineate folding mechanisms; readers of *Folding & Design* will already be familiar with many of these. The advantages of a multifaceted approach to understanding folding mechanisms was emphasised, using lysozyme and acyl CoA binding proteins as examples, including the use of stopped flow techniques linked to optical spectroscopy, such as fluorescence and circular dichroism, as well as the powerful combination of hydrogen exchange labelling detected by ^{1}H NMR or electrospray ionisation mass spectrometry (ESI MS). It is important, however, to bear in mind that folding occurs rapidly (for most proteins), heterogeneously, and from a denatured state that is by no means random. Thus, the latest important concept for the experimentalist is the 'new view' of folding, in which we visualise the folding process as a three-dimensional downhill energy search for the native state. Chris Dobson pointed out that in reality this means that potentially every molecule in a 'normal' refolding experiment (e.g. using μM protein concentrations) could fold in a unique way from a different conformation in the denatured state, but over the total ~10^{18} molecules in the refolding reaction almost the entire energy landscape is sampled. The impact and the power of the new view of folding is in our understanding of species such as the 'molten globule' and early folding intermediates in which the average properties of the ensemble are measured experimentally. The new view also highlights the need for experimentalists to design new experiments to determine the nature of the energy surface for folding, to bridge the gap between rather ill-defined views (in experimental terms) that arise from simple physical-chemical calculations of folding funnels and the ensemble average view of folding that is the necessary outcome of current methods of studying folding reactions.

The language barrier between experimentalists and theoreticians is now beginning to be bridged and, as stated by Martin Karplus (Harvard University), "the most exciting aspect in folding currently is the collaboration between theory and experiments that leads to stimulation between them". One clear example of this is in the particular views of energy landscapes described by Chris Dobson, which have emerged from a collaboration with Martin Karplus; such energy surfaces are defined in terms of experimentally understandable parameters (average effective energy, number of native contacts, and a parameter related to the configurational entropy). A second important contribution by Martin Karplus has come from simulations of the folding pathway of chymotrypsin inhibitor II (CI2), the well studied work-horse from Alan Fersht's group and the paradigm for a protein with a simple nucleation condensation mechanism of two-state folding. By recalculating the unfolding trajectory many times, Martin Karplus was able to address the question of whether there is a multiplicity of folding pathways, even for a protein with a such a simple folding mechanism. The results show, as perhaps expected, that multiple pathways exist, but importantly the average properties of molecules populating different pathways agree closely with the experimentally measured averaged ensemble pathway. Using molecular dynamics simulations, Lorna Smith (University of Oxford) also presented work that bound theory and experiment. Regions

from lysozyme that are known to be partially folded in solution were excised from the protein and unfolded using molecular dynamics. Interestingly, not all regions in the peptide sequence unfolded completely, but significant helical structure was retained in two sequences, which correlate with regions shown by NMR to be helical in a synthetic peptide corresponding to the same sequence. Long-range interactions, which could not be pinpointed by NMR as a result of degeneracy in the spectrum, were shown in the simulation to stabilise this partially folded state. This points the experimentalist towards new strategies to test the model. Both of these talks, therefore, highlight the convergence of experiment and theory and offer the exciting possibility that theoreticians may now be able to predict realistic folding mechanisms that can be tested by experimentalists.

A major breakthrough in our understanding of the behaviour of proteins in solution has arisen from the ability to study them using high resolution NMR methods. Most recently, the advent of isotope labelling and very high frequency instruments (600 MHz and higher) have allowed us to start to contemplate the precise nature of the unfolded states of proteins in solution as well as that of intermediate partially folded states. Presentations by Lorna Smith and Peter Wright (Scripps Research Institute) addressed these issues. Peter Wright described the latest view of the apo-myoglobin folding pathway gleaned from triple resonance NMR experiments. Here, the $^{13}C\alpha$ chemical shift and the ^{15}N NOE intensity were used as probes of the conformational preferences and dynamic properties of the peptide backbone in denatured apo-myoglobin, in an equilibrium partially folded state at pH 4 and in the native state. Further insights into the conformation of the denatured state were presented by Lorna Smith. She described how a model for the random coil of proteins and peptides can be derived from a knowledge of the distribution of short-range NOEs and used with high accuracy to predict $^3J_{H\alpha}$ and $^3J_{\alpha\beta}$ coupling constants of these states. Not only is this algorithm important for understanding the conformation of denatured and partially folded states, but it can be used to tease out weak, but important, configurational preferences in proteins that are functional in 'so-called' unfolded states. Lorna Smith used this approach to show that three regions in fibronectin binding protein (which belongs to this set of bizarre proteins) have a high propensity for occupying the extended region of conformational space (over a running average of five residues). Fascinatingly, all three regions (and no others) have been shown to be important for activity, demonstrating that even the unfolded state can harbour properties that are intimately involved in activity.

Although we have made great strides in our understanding of how proteins fold, it should be borne in mind that our current database of knowledge about folding has arisen from detailed work on only a handful of proteins, that may not be representative of the thousands of structures in the protein databank. Thus, we must be careful not to draw too many general conclusions about folding mechanisms at this stage. One recurring point that arose at the Workshop was the issue of whether intermediates detected experimentally at equilibrium or kinetically teach us about mechanistic issues of folding, or whether they represent trapped states or off-pathway species. In the new view of folding, the issue of on-pathway or off-pathway intermediates could be considered irrelevant because all species populated define the energy surface for folding. Nevertheless, as stated at the Workshop by Martin Karplus, an off-pathway intermediate can be defined as one whose structure does not help in the search for the native state. The issue then boils down to whether a particular species is obligate for folding or not. In the case of lysozyme, the α-domain intermediate is bypassed by about one-quarter of molecules that fold on a fast track to the native state, indicating that this species is not required for the ultimate lysozyme fold. In contrast, Peter Wright presented hydrogen exchange labelling and ESI MS data to show that the apo-myoglobin intermediate is obligate for folding, and data from my own laboratory shows that for the Greek key protein apo-pseudoazurin a late and unusually stable intermediate (relative to the native state) is undoubtedly on the folding pathway. This emphasises that new information can be learned from proteins with different folds and a wide database of folding information will be necessary before we can answer the simple, but essential, question of whether there is a universal folding code, or whether each protein has solved the protein folding problem individually.

The ultimate goal of the protein folder is to be able to design a sequence *de novo* that will fold to a unique, stable and active conformation. Luis Serrano (European Molecular Biology Laboratory, Heidelberg) presented work at the Workshop that shows we are now well on the way to achieving this goal. Thus, a novel mini β-sheet protein called Betanova, only 16 amino acids in length, was designed using an algorithm that optimises propensities for β-sheet structure, stabilising non-local interactions, rotamer libraries and solubility of amino acid sidechains. From the 10^{30} possible combinations, 100 amino acid sequences emerged and the sequence with the lowest energy was selected and made as a synthetic peptide. Astonishingly, the peptide folded into the desired monomeric triple stranded anti-parallel β-sheet structure, which, as judged by NMR parameters, has a unique structure. Although the peptide showed only weak cooperativity in thermal denaturation, this might be expected for such a small structure. This *de novo* designed triple-stranded β-sheet structure will certainly be useful for delineation of the interactions that stabilise β-sheets. Perhaps even more excitingly, however, it might form a platform on to which functionality, the ultimate challenge in folding and design, can be built.

Folding in intact cells

Though wondrous and exciting, it is important to realise that protein refolding experiments *in vitro* are only mimics of the real situation of protein folding in the cell. As we were reminded in the first lecture of the Workshop by John Ellis, folding *in vivo* occurs in a highly crowded environment (the concentration of macromolecules in the *Escherichia coli* cytoplasm is ~340 mg/ml), and may even occur co-translationally on polysomes. As a result, the possibility for misfolding and aggregation are much higher *in vivo* than *in vitro* and the cell has evolved a complex system involving chaperone proteins, which function to ensure the success of folding in the otherwise unfavourable cellular environment. As the protein sequence database increases in size, so does the number of chaperones. Thus, for example, 14 hsp70-like proteins are found in yeast alone and, as John Ellis pointed out, the list of different chaperone families now numbers > 20. Nevertheless, only 10–15% of newly synthesised proteins in *E. coli* are thought to fold using the famous GroEL machinery, the remainder presumably folding spontaneously in a manner akin to that elucidated *in vitro*, or perhaps using other chaperone systems.

Although we now have a detailed knowledge of the structure of GroEL and its co-chaperone GroES, and even of the 870 kDa GroEL–GroES complex, the identity of the proteins that require this chaperonin for folding has remained elusive. The promiscuous nature of GroEL for partially folded substrates is well known, but the fact that so few proteins appear to fold using GroEL leads to the possibility that the real substrates for the chaperonin *in vivo* might be rather limited. Two different lectures at the Workshop, by Ulrich Hartl (Max Planck Institute, Martinsreid) and Art Horwich (Yale University), presented the first clear cut data on this issue. Ulrich Hartl and coworkers have used two-dimensional gel electrophoresis to resolve newly synthesised chains that had been immunoprecipitated with anti-GroEL antibodies. Using modern tandem mass spectrometry methods the spots on the gel could then be identified. Art Horwich and coworkers used a similar analytical approach, but instead of immunoprecipitation a temperature sensitive lethal mutation in GroEL was used to precipitate, in inclusion bodies, proteins whose folding is absolutely dependent on a functional GroEL. Although far from complete (only a few of the spots have so far been identified), both sets of experiments identified elongation factor Tu as a substrate for GroEL and other proteins, including GroEL itself; other elongation factors, E3 from pyruvate dehydrogenase and components of the F1 subunit of the F0F1 ATPase were identified by one or other of the groups. This now opens the door to the long awaited biophysical experiments of the GroEL reaction cycle using a real substrate protein. Moreover, the realisation that GroEL is essential for the folding of proteins involved in protein synthesis itself offers new and exciting possibilities for communication between chaperonins and protein production.

The apical domain of GroEL and its counterpart in the thermosome and in the eukaryotic complex, CCT, also received much attention at the Workshop. In his lecture, Alan Fersht (University of Cambridge) reviewed published data from his laboratory that demonstrated the potential substrate-binding site on the surface of GroEL (located by the fortuitous binding of a histidine tag in the domain construct 191–376 to a neighbouring apical domain in the crystal) and data suggesting that this domain is able to act as a mini-chaperone, assisting the refolding of several proteins *in vitro*. He emphasised the previously published work from his laboratory that shorter constructs are more active than the 191–376 construct, and gave new examples in which immobilised mini-chaperones are being used preparatively in the laboratory. New work from Jean Chatellier in Alan Fersht's group was also presented that showed that the shortest apical domain made, encompassing residues 193–335 of GroEL, complements a temperature-sensitive mutant of GroEL and supplements the activity of co-expressed GroEL in lethal GroEL knockouts of *E. coli*. The less active full-length domain (191–376), as found and also presented in the Workshop by Ulrich Hartl, does not complement the temperature-sensitive mutants. The reasons for the increased activity of the 'mini-mini-chaperone' remain elusive. Nevertheless, the mini-chaperones have a potential use for refolding overexpressed proteins and offers exciting potentials for the refolding of proteins in the biotechnology industry.

Awe-inspiring images of GroEL obtained using cryo-electron microscopy (cryo-EM) from Helen Saibil's group (Birkbeck College) continue to impress us with the magnificent beauty of the chaperonin system. Although at low resolution, these pictures showed us the first images of the enormous domain movements that occur when GroES binds to the GroEL tetradecamer and the method, so far, is the only way that differences between the nucleotide-bound and free forms of the chaperonin have been resolved. In his lecture, Keith Willison (Institute of Cancer Research, London) showed the first images of the CCT complex in the presence of ATP, also obtained using EM. As in its prokaryotic counterpart, huge domain movements are seen when the CCT rings bind nucleotide. Because no crystal structure of this complex is yet available, and biophysical data on the conformation of the bound substrate are still lacking, the EM images provide the first exciting clues that the mechanism of action of CCT and GroEL might be similar. In her presentation, Helen Saibil also presented a 9 Å map of GroEL obtained from high resolution cryo-EM experiments using angular reconstitution methods. Comfortingly, the image bears remarkable similarity to the gross features extracted from the crystal structure of the chaperonin. The impact of this

work does not lie in the structure of GroEL itself, however, but, combined with freeze-quench methods, the technique offers exciting possibilities to see almost molecular details of this powerhouse in action.

Watching the folding of the nascent chain as it emerges from the ribosome is perhaps the ultimate goal for protein folders in all areas of the field. Just imagine the magnificence of a molecular graphics picture containing a translating ribosome, an emerging, possibly partially folded, nascent chain, which may also be decorated with various molecular chaperones. Before such a picture could be drawn, however, the chaperones that associate either with the ribosome or with the nascent chain, or both, need to be mapped out. The presentation by Betty Craig addressed this point, at least for yeast. In elegant work, Betty Craig (University of Wisconsin) was able to identify that one of the 14 hsp70 molecules in yeast, named ssb, is a ribosome-associated chaperone that also binds to the emerging nascent chain. Bernd Bukau (University of Freiburg) reminded us, however, that the situation in translating ribosomes in prokaryotes might be quite different because there is no direct equivalent to ssb in *E. coli*. The ribosomes themselves might also be important for folding the nascent chain. Indeed, as pointed out by Boyd Hardesty (University of Texas at Austin), the 50S subunit of *E. coli* ribosomes, and specifically, the 23S RNA, is capable of assisting the folding of rhodanese, at least *in vitro*. Boyd Hardesty's experiments also suggest that the elongation factors EF Tu and EF Ts may be involved in folding. Again, the situation *in vivo* might be quite different because (as pointed out at the meeting by Ulrich Hartl) newly synthesised rhodanese is critically dependent upon GroEL and GroES. So, how far off is the ultimate goal? In exciting experiments using ESI MS, Chris Dobson showed that is now possible to obtain mass spectra from intact ribosomes and that the individual components can be identified in the forest of peaks. It seems that we now have the tools to tackle the real issue of co-translational protein folding, and the ultimate goal of determining the structure of the nascent chain as it emerges from the ribosome may be within our grasp.

Folding and disease

Protein folding diseases brought the Workshop to a dramatic conclusion with a day-long session bringing together medics, biophysicists and biochemists in the common quest to find treatments for these currently incurable illnesses. Protein aggregation was a central theme in the session, which, rather than being considered merely a nuisance to *in vitro* folders, has now caught on as one of the hottest topics in folding! The role of the molecular biophysicist in folding diseases is an important one because it is only once we can understand the molecular events that occur, for example early in amyloidosis, that new ideas for the generation of novel therapies will emerge. Using purified proteins and a battery of biophysical methods that *in vitro* folders have been using for years, we are now beginning to gain insights into the properties of proteins that make them misfold. Jean Baum (Rutgers University) used state-of-the-art NMR methods to study the effect of mutations in collagen that cause *osteogenesis imperfecta*. In a feat using real-time NMR, the process of collagen triple-helix assembly could be followed at atomic resolution in real time and the effect of mutations in perturbing this process at the molecular level was revealed. Biophysical methods were also used by Phillip Thomas (University of Texas Southwestern) to analyse the ΔF508 mutation in CTFR that causes cystic fibrosis. Again, the mutation was also shown to result in a destabilised protein with perturbed folding kinetics.

A similar theme also emerged from biophysical studies of proteins involved in amyloidosis. Mark Pepys demonstrated (through collaborative work with Chris Dobson and colleagues) that mutations in human lysozyme that cause amyloidosis are destabilising and decrease the cooperativity of the native fold. Jeff Kelly (Scripps Research Institute) presented data to show that amyloid-causing mutations in transthyretin (TTR) result in an increase in the concentration of the amyloidogenic TTR monomer relative to the normal tetramer, coupled with a massive (1000-fold) increase in the rate of unfolding of the monomer. In addition, using hydrogen exchange monitored by ESI MS Carol Robinson (University of Oxford) was able to study the monomeric form of wild-type and the Val30→Met amyloidogenic mutant of TTR, revealing a decrease in protection of the monomeric state of the amyloidogenic variant, consistent with the hypothesis that edge-strand peeling could initiate aggregation. Using the latest QTOF mass spectrometer Carol Robinson was also able to show fascinating data about the early stages of aggregation of insulin. As she pointed out, the power of the QTOF lies in its ability to detect and resolve high molecular weight species, and using this instrument a range of aggregated states up to 14 subunits in size could be seen in the insulin solution immediately after the pH was reduced to 2.0. The precise nature of the species and its involvement in amyloid formation has yet to be determined; nevertheless, this work demonstrates the potential of ESI MS for the delineation of the important early stages in the aggregation process. Other highlights in this session came from Valerie Daggett (University of Washington) who described how molecular dynamics simulations have now reached the stage at which useful information about the early stages of aggregation might be gleaned, David Eisenberg (University of California, Los Angeles), who described how a model involving domain swapping might provide a route to the production of ordered fibrils, and Max Perutz (University of Cambridge) who described how polar zippers might be responsible for fibril formation in Huntingdon's disease. Finally, Chris Dobson showed

a wonderful three dimensional image of a protein fibril, obtained by Helen Saibil using cryo-EM. The fibril was made from an SH3 domain at acidic pH. Although not associated with any disease, this is the first detailed cryo-EM image of a fibril and gives hope for a fuller picture of the structure of amyloid in the future.

What are the possibilities for potential therapies for amyloidosis? As detailed by Mark Pepys, several strategies can be envisaged, including suppressing the production of the amyloid precursor, preventing amyloid formation and increasing amyloid degradation. Suppression of the production of the amyloid precursor has been taken up by Jeff Kelly. In his lecture he described how stabilising native TTR by binding an analogue of thyroxine (the natural ligand) prevents amyloidosis *in vitro*, and the results in a mouse model are also very encouraging. There was also encouraging news from Byron Caughey (NIAID Rocky Mountain Laboratories) who presented data that a short peptide from the prion protein (PrP) is able to inhibit the conversion of soluble human prion protein to the insoluble form. Finally, using the third strategy mentioned above, Mark Pepys has developed a potential universal treatment for amyloidosis that is based on the observation that serum amyloid P component (SAP) binds to most amyloid deposits *in vivo* and protects the fibril from degradation. Thus, by stripping SAP off the fibrils the aggregates are degraded naturally, presumably by phagocytosis. Drug screening has produced a compound with the desired properties and the work has progressed as far as mouse models, in which the compound was also effective as an anti-amyloid. We now wait with bated breath to discover the results of trials with the compound in man.

Summary

So what did we learn from the meeting in Oxford? For me, the most exciting aspect was the power of an interdisciplinary approach that is now able to link theoretical simulations of folding to experiments *in vitro* and *in vivo*, and ultimately to the solution of a medical problem in man. The prospects for the future are exciting. New protein sequences are emerging almost daily through the various genome projects and the crystallographers continue to excite us with new and ever more complex structures at atomic resolution. The ability to tackle large protein assemblies by combining different structural approaches and to understand the conformational preferences of molecules that even lack a unique structure using NMR also bode well for the future. It is perhaps now only a matter of time before we will start to see predictions of folding mechanisms *ab initio*, images of nascent chains emerging from ribosomes, and new therapies for folding diseases. What is clear is that the rapid innovations in this dynamic field will make for many exciting workshops to come!

Acknowledgements

I would like to thank my colleagues for critically reading this report prior to its publication. I also thank the organisers for putting together such an exciting workshop.