Ubiquitin: a small protein folding paradigm

Sophie E. Jackson*

Received 18th January 2006
First published as an Advance Article on the web 27th April 2006
DOI: 10.1039/b600829c

For the past twenty years, the small, 76-residue protein ubiquitin has been used as a model system to study protein structure, stability, folding and dynamics. In this time, ubiquitin has become a paradigm for both the experimental and computational folding communities. The folding energy landscape is now uniquely characterised with a plethora of information available on not only the native and denatured states, but partially structured states, alternatively folded states and locally unfolded states, in addition to the transition state ensemble. This Perspective focuses on the experimental characterisation of ubiquitin using a comprehensive range of biophysical techniques.

Introduction

It appears that the name ubiquitin is an apt one—ubiquitin not only being ubiquitous in a cellular environment in nature, but also widespread in its use in research laboratories worldwide. The extensive use of ubiquitin results from its favourable properties—it is a small protein (76 residues in length) that has a highly structured native state which is very stable. Its high stability has been known for some time and this may be linked with the function of ubiquitin, which becomes covalently attached to lysine side chains in proteins thereby targeting them for degradation by the proteasome.[1] As far back as the 1970s, NMR studies had established that ubiquitin is stable from pH 1.2 to 8.5 and from 23 to 80 °C.[2] Its stability has proved of great use particularly in the original preparations of the protein from non-recombinant sources where high temperatures (90 to 95 °C) were used to precipitate out many other cellular proteins, with ubiquitin remaining folded and soluble.[3] Perhaps more important than all of these factors, including the relative ease of preparing the pure protein, however, was the fact that for many years the protein was available straight from a bottle from Sigma. Nowadays, of course, recombinant protein made in bacterial sources is usually used.

For all these reasons, once the structure of ubiquitin had been solved in the mid 1980s, it fast became the mainstay of many studies on protein stability and folding. Over the past twenty years, this small protein has been subjected to an extensive range of biophysical techniques, denaturation methods and protein engineering studies. Together, these studies have provided one of the most detailed pictures of the complex nature of the energy landscape for folding so far obtained for a protein. In addition to the many experimental studies, which are reviewed in this Perspective, ubiquitin has also been the focus of many computational studies using a variety of approaches. Unfortunately, these are beyond the scope of this particular review, however, interested readers can find out more about computational methods in a number of recent review articles[4,5] and in the June 2002 issue of *Accounts of Chemical Research* dedicated to molecular dynamics simulations of biomolecules.[6]

This Perspective starts with the original structural studies which led onto the subsequent work on stability and folding. It ends with recent studies on the redesign of the hydrophobic core of ubiquitin, the latest studies using new biophysical techniques such as atomic force microscopy (AFM), and novel approaches to the study of protein structure and dynamics.

Structural studies

The three-dimensional structure of ubiquitin was first solved by X-ray crystallography at 2.8 Å and then 1.8 Å resolution in the mid 1980s.[7,8] The protein was found to have a tightly-packed globular structure in which a mixed parallel–anti-parallel β-sheet packs against an α-helix to form the hydrophobic core (Fig. 1). This structure, termed a β-grasp fold, is widespread and many proteins and protein families are found to adopt this topology. In contrast to the crystal structure, a solution structure of wild-type ubiquitin was not reported and coordinates deposited until

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, U.K. CB2 1EW. E-mail: sej13@cam.ac.uk; Fax: +44 (0)1223 336362; Tel: +44 (0)1223 762011

Sophie Jackson gained a degree in chemistry from Oxford University in 1987, and went on to join the laboratory of Prof. Sir Alan Fersht for her PhD, first at Imperial College, London and then in Cambridge. After a spell as a research fellow at Peterhouse, Cambridge, she was a postdoctoral fellow in Prof. Stuart Schreiber's laboratory at Harvard University. In 1995, she started her own research group as a Royal Society University Research Fellow and is currently a senior lecturer in the chemistry department.

Sophie E. Jackson

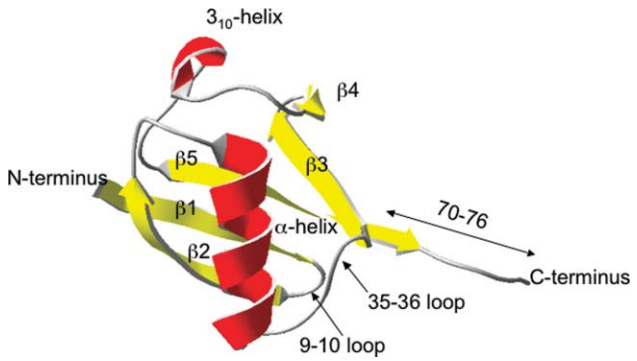

Fig. 1 Secondary and tertiary structure of ubiquitin from the 1.8 Å crystal structure (1ubq.pdb).[8] Shown in yellow are the five β-strands that make up the mixed parallel–anti-parallel β-sheet, shown in red are the major α-helix and the 3_{10}-helix. Also indicated are the loop regions which have been engineered in different studies.

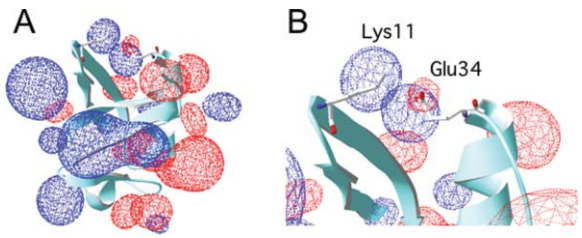

Fig. 2 **A**. Structure of ubiquitin showing the electrostatic potential generated by the acidic and basic groups (a positive potential is indicated in blue and a negative potential in red). **B**. Structure of ubiquitin showing the side chains of residues Lys11 and Glu34 which form a salt bridge on the surface of the protein. Also shown are the electrostatic potentials from these and nearby charge residues. These charges have been engineered by the Makhatadze group to form stabilised ubiquitin variants.[19]

1998. In this case, the solution structure was solved as a part of a larger study using ubiquitin as a test protein for developing new NMR methods and for the validation of new structural information obtained from a sample prepared in a dilute liquid crystal medium.[9] The fact that the solution structure was not published until 1998 is somewhat surprising given that 1D-NMR studies on ubiquitin were being performed as early as 1977,[2] some assignments and histidine titrations were reported in 1980,[10] and a full 2D assignment was published by two groups in 1987.[11–13] Ubiquitin has, however, played a very important role in the development of NMR methodology, it being used as a test case in many studies. A full review of this work is beyond the scope of this Perspective, but for interested readers I recommend work from the Bax group (http://spin.niddk.nih.gov/bax/).

Stability studies

Ubiquitin has been used extensively as a model system for investigating the factors governing the stability of proteins. Much of the work done in this area has been carried out by the Makhatadze and Privalov groups starting with differential scanning calorimetry measurements in the early 1990s.[14] The first experiments looked at the effects of salts and surface charges on the stability of the wild-type protein,[15,16] work which was quickly followed up with protein engineering studies. Through a detailed analysis of the effect of surface charge on stability, the Makhatadze group have produced a stabilised variant of ubiquitin by the optimization of charge–charge interactions on the protein surface (Fig. 2).[17] This group has also demonstrated that removal of surface charge–charge interactions leaves the protein folded and very stable,[18] the work culminating in a recently published paper giving guidelines for the protein engineering of surface charge.[19] Other groups have also studied the numerous charge–charge interactions on the surface of ubiquitin to investigate how neighbouring residues influence the strength of these interactions and pK_as.[20]

Protein engineering techniques have also been employed to study the contribution of the contact between the hydrophobic residues Ile30 and Ile36 at the C-terminus of the α-helix to protein stability.[21] In this case, 16 variants were produced with the full combination of four hydrophobic residues at the two positions, all of which were less stable than the wild-type protein. Other groups have also studied the effect of mutation of hydrophobic core groups on the stability.[22–24] In these cases, mutation of buried hydrophobic core residues destabilises the native state of ubiquitin with the exception of Val→Ala26 which is slightly more stable than the wild type.[22] The fact that almost all mutations in the core are destabilising is perhaps not that surprising given the remarkable sequence conservation of ubiquitins.[25] The effects of the burial of polar and non-polar groups in the interior of ubiquitin have also been measured and, as expected, the packing of non-polar groups in the core is found to be favourable, whilst the burial of polar groups is highly destabilising unless the destabilisation resulting from dehydration can be compensated for by the formation of hydrogen bonds.[26]

The Makhatadze group has also investigated the stability of mutants of the C-cap and the C′ position of the C-capping box of the α-helix. In the first of these studies, they established that glycine is the most favourable residue in the C′ position, the result of preferential hydration of the peptide backbone.[27] At the C-cap itself, a large variation in stability was observed for the different mutants made, the stability correlating well with hydrophobicity in this case.[28] Recently, a detailed investigation of the factors contributing to stability at solvent exposed positions in the middle (Ala28) and at the C-terminus (Asp32) of the α-helix of ubiquitin was undertaken. The results showed that, for non-charged amino acids, the helical propensity is similar at both positions and similar to the propensities measured in model peptides.[29] In contrast, when a partially exposed site in the α-helix (Lys33) was studied, it was found that both the helical propensity and the side chain hydrophobicity correlated with changes in stability.[30] Thus, the effect on stability of a residue in the α-helix is context dependent, and cannot be predicted from helical propensities alone.

Ubiquitin's intrinsic stability makes it an ideal system for stability studies enabling a comprehensive programme of protein engineering. To date, more than 70 mutants have been made and characterised which have a range of stabilities with $\Delta\Delta G_{D-N}$s (the difference in the free energy of unfolding between wild type and mutant) ranging from −1.7 kcal mol^{-1} (more stable than wild-type) to + 4.4 kcal mol^{-1} (less stable than wild-type).

Partially structured states

Several different approaches have been used to identify and characterise partially structured states of ubiquitin. Such states provide important information on the intrinsic stabilities of substructures

 This journal is © The Royal Society of Chemistry 2006

within the protein as well as local unfolding events, both of which help to define the energy landscape for the folding/unfolding of this protein.

(i) Peptide fragment studies. Some of the earliest studies on the folding of ubiquitin focused on peptide fragments corresponding to different regions of the full-length protein. Evidence for intrinsic structure in the absence of the rest of the polypeptide chain was sought in order to identify potential nucleation sites for folding. Two peptides corresponding to residues 1–21 and 1–35 were studied by NMR and far-UV CD in water and at low pH in methanol.[31] Whilst in water there was little evidence of any structure, at low pH in methanol both peptides had considerable structure which was consistent with the formation of the β-hairpin.[31] More recent investigations have established that a peptide corresponding to residues 1–17 had some native-like structure in solution and a mutant thereof, (Thr→Asp9), was more stable and highly structured, possibly due to the introduction of a favourable interaction with Lys11.[32,33] In contrast, a fragment corresponding to the C-terminus of ubiquitin (residues 36–76) showed no evidence for structure in solution.[34] As has been shown for other proteins, the N- and C-terminal fragments can be combined and cooperatively assemble into a native-like structure.[34]

Some quite striking results were obtained when a peptide fragment encompassing residues 1 to 51 of ubiquitin was produced in order to test whether this portion had a propensity for independent self-assembly.[35] Surprisingly, the construct formed a folded symmetrical dimer that was stabilised by 0.8 M sodium sulfate. The solution structure was determined by NMR spectroscopy and each subunit found to consist of an N-terminal β-hairpin followed by an α-helix and a final β-strand, with orientations similar to intact full-length ubiquitin (Fig. 3). The dimer is formed by the third β-strand of one subunit interleaving between the hairpin and the third strand of the other to give a six-stranded β-sheet, with the two α-helices sitting on top (Fig. 3). The helix–helix and strand portions of the dimer interface also mimic related features in the structure of wild-type ubiquitin.

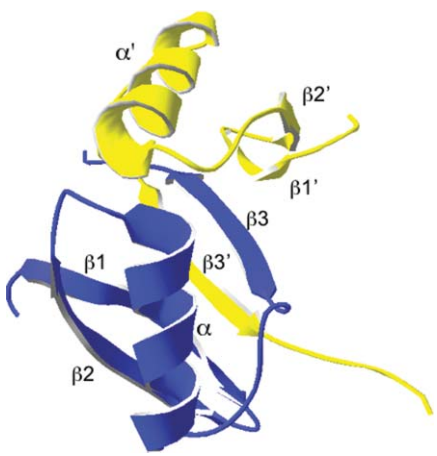

Fig. 3 Structure of the dimer formed by the 1–51 peptide.[35]

(ii) Studies in cosolvents. Organic cosolvents such as alcohols (methanol, trifluoroethanol, hexaisopropanol *etc.*) have long been known to affect the stability and structure of proteins.[36] At low concentrations, these cosolvents commonly increase the stability of the native state, however, at higher concentrations they are known to partially or fully unfold a protein. Their ability to stabilise otherwise unstable partially structured states of small proteins has resulted in their extensive use in characterising partially structured states which may be similar to folding intermediates. The groups of Evans and Williams were the first to use this approach on ubiquitin forming a stable partially structured state at low pH in a 60% : 40% (v/v) mix of methanol and water. This state (which was called the A state) was sufficiently soluble and stable to be characterised using NMR.[37] Slowly exchanging amides were assigned and vicinal coupling constants and NOESY data were used to show that the first two strands and part of the third strand remained structured. The hydrophobic core face of the loose β-sheet formed was partially covered by a weakly structured α-helix, which was considerably more flexible than in the native state.[37] Subsequently, thermal denaturation experiments monitored by NMR and DSC were performed on the A state to determine its heat capacity.[38] Surprisingly, the heat capacity change between the A state and the denatured state was close to zero indicating that the solvent ordering component of the hydrophobic effect was not an obligatory factor in determining the stability or structure of the A state.[38] These results were confirmed by a later study which proposed that van der Waals' interactions dominate under these conditions.[39]

(iii) Pressure and temperature. In addition to the fragment and cosolvent studies described above, pressure and temperature have both been used to induce partial unfolding of ubiquitin, the partially structured state formed being characterised by NMR. At low temperature (0 °C) and high pressure (30–3700 bar) a series of high-energy intermediates have been observed.[40] The major intermediate state detected has undergone a local unfolding event in the region of residues 33–42 and between residues 70–76.[40] In a follow-up study, the same group have used NOE data and torsional angle constraints to create average coordinates for the structure of ubiquitin at various pressures.[41] The results revealed that the α-helix can swing in and out by more than 3 Å with a simultaneous reorientation of the C-terminal region of the protein. Spin relaxation analysis established that these conformational fluctuations are occurring on a 10 μs timescale.[41]

Cold denaturation is rarely used to study unfolding transitions as, for the majority of proteins, denaturation usually occurs below the freezing temperature of the aqueous solution. Recently, Wand and coworkers have developed a novel method which overcomes this difficulty, and have studied the cold denaturation of ubiquitin in micelles at −35 °C.[42] Monitoring the unfolding transition by NMR, they found that the unfolding is relatively non-cooperative under these conditions and an ensemble of partially structured states was observed. Within this ensemble, the N-terminal region of ubiquitin along with the α-helix, in addition to the 3_{10}-helix, remained native-like and folded.[42] In contrast, a study on the pressure-assisted cold denaturation of ubiquitin at 225 MPa and −16 °C, showed H/D exchange kinetics consistent with a random coil.[43]

NMR: H/D exchange and hydrogen bonding

In addition to its use in determining the structure of the native state of ubiquitin, NMR methods have been used in conjunction with H/D exchange methods to provide information on the folding

pathway of this protein. The first studies used pulsed-quenched flow H/D exchange to investigate potential intermediates on the folding pathway.[44] These experiments showed that ubiquitin folds rapidly and cooperatively with the majority of amide protons being protected against labeling in a phase on the 10 ms timescale, with little evidence for protection of amides in any faster phase.[44] A later study used a variation on this method to detect and characterise early intermediates.[45] In this case, no significantly protected structure was detected suggesting that secondary structural elements may only be populated marginally ahead of, or at the same time as, the major cooperative folding event.[45] This is consistent with the current models of ubiquitin folding which show it to be two-state under these conditions;[46,47] the protection observed in the original studies may have resulted from aggregation known to occur during folding under these conditions (Figs 4a and b).[47]

Recent H/D exchange studies on ubiquitin[48] have established the experimental conditions required for EX1 and EX2 regimes of exchange. (The mechanism of exchange is classified as EX1 when the intrinsic rate of exchange is sufficiently fast that the observed rate of exchange is determined solely by the opening or unfolding rate of the protein; an EX2 mechanism applies when the intrinsic rate of exchange is slower than the closing (folding) rate. In this case, the observed rate constant depends upon the equilibrium constant between the open and closed forms and the intrinsic rate constant for exchange.) Exchange occurs by the EX1 mechanism at high pHs where the intrinsic rates of exchange of amide protons are rapid. Under these conditions, the exchange rates can be used to calculate unfolding rate constants and, combined with EX2 measurements at low pH, used to calculate folding rates.[48] The rate constants obtained by this method are in excellent agreement with those measured using other techniques (see kinetics section) and with saturation transfer experiments.[48]

NMR techniques have also been used to examine the hydrogen bonding groups in ubiquitin and, in particular, their response to temperature.[49] A residue-specific analysis has established that not all hydrogen bonds are affected to the same degree by temperature and thermal expansion. The N-terminus of β-strand 5 was found to be the least thermally stable, as expected as this region sticks out from the main body of the protein. In contrast, the end of β1–β2 is stable even at elevated temperatures. Hydrogen bonds in the α-helix are also particularly strong.

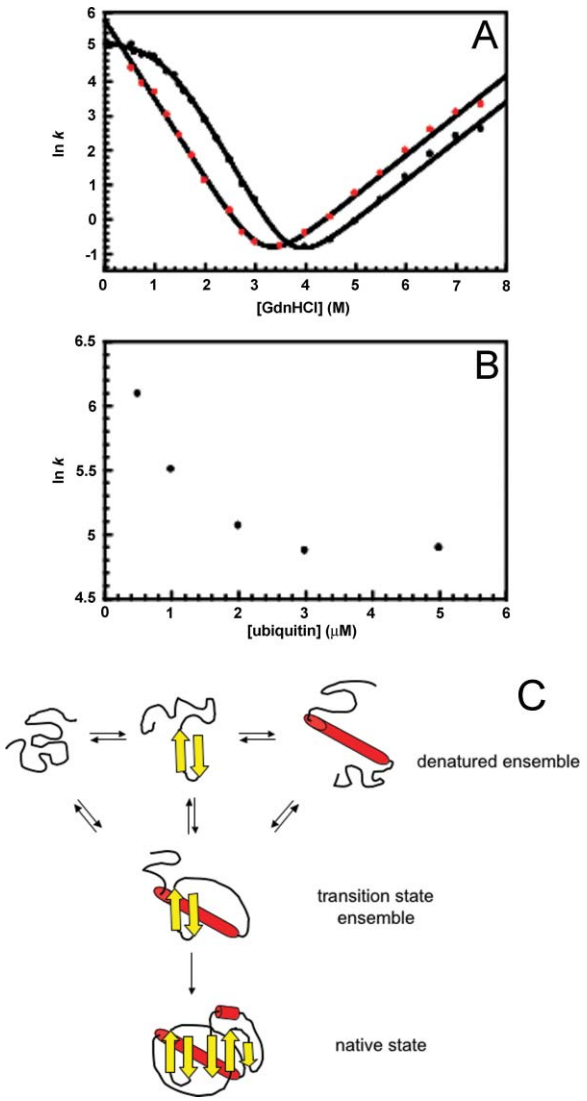

Fig. 4 Unfolding and refolding kinetics of ubiquitin at pH 5.0. **A**. Tagged F45W (red open circles) and a non-tagged F45W mutant (black closed circles). The latter is identical to the protein used in other studies.[22,46] The V-shaped chevron plot is indicative of the simple two-state folding of a protein[85] whilst rollover at low denaturant concentrations can be indicative of the population of an intermediate state[86] or of transient aggregation during folding.[87] **B**. Protein concentration dependence of the refolding rate constant for the non-tagged F45W mutant at 0.5 M GdmCl at 25 °C, clearly showing that transient aggregation is occurring. Data taken from Went *et al.*, (2004).[47] **C**. Putative scheme for the folding pathway of ubiquitin based on protein engineering and other results.[24] It is likely that there is some residual structure in the denatured state of ubiquitin in the region of the first β-hairpin and the α-helix. Such structure will be transient, flickering in-and-out. By the rate-limiting transition state this structure has consolidated—the helix is almost fully formed and the β-hairpin is partially structured. In this state, the secondary structure is stabilised by interactions between the β-hairpin and the helix. The C-terminal region remains relatively unstructured until after the TS barrier has been crossed. It is formed rapidly in a downhill process post-TS.

Folding kinetics

(i) Two- or three-state? In recent years, the kinetic model of folding for ubiquitin has become somewhat controversial. The original experiments of Roder and coworkers on a fluorescently engineered variant of ubiquitin (F45W) reported that the protein folded with three-state kinetics populating an intermediate state during folding at 25 °C (but interestingly not at 8 °C).[50] A follow-up study by the same group using protein engineering techniques to probe the folding pathway showed that mutants of residues in the hydrophobic core which destabilised the native state of the protein also destabilised the intermediate state and two-state kinetics were observed.[22] This was consistent with other studies of that time, and many subsequent studies, which have shown that it is possible to switch from three- to two-state kinetics by destabilising the native state of a protein either by mutation or by changing an experimental condition such as pH and temperature. However, a later investigation by the Sosnick group combining different experimental approaches questioned the three-state nature of the folding of ubiquitin.[46] Recent work from my own group has established the cause of the apparent

 This journal is © The Royal Society of Chemistry 2006

discrepancy. Ubiquitin has a strong tendency to aggregate during refolding at low concentrations of chemical denaturant and this gives rise to what appears to be an intermediate state but which is artefactual (Fig. 4a,b).[47]

Rapid-mixing experiments have also been used to measure the solvent isotope effects on unfolding and refolding rates.[51] In this case, stability and rate constants in H_2O and D_2O were measured and compared. The refolding rate constants remain unchanged between solvents, and there was only a small effect observed on the unfolding rate constant. These results have been interpreted in terms of hydrogen bond formation in the rate-limiting transition state ensemble (TSE)—it being concluded that hydrogen bonds are largely absent in the TSE. However, it was noted that, in this case, the approach may be at its limit due to the small changes in stability of the system between solvents.

Recently, the folding of human ubiquitin was compared with that for the highly homologous yeast protein and the structurally related Raf-RBD which has no sequence similarity.[52] In all cases, the kinetics were two-state and the folding rates were all within an order of magnitude. This, in conjunction with the fact that the folding/unfolding energy barriers displayed a similar temperature dependence, and sensitivity to a stabilising salt and to mutation, led to the conclusion that the folding pathways were the same. In contrast, recent work from my own group, suggests that, whereas the folding pathways of proteins with the same β-grasp fold as ubiquitin which are thought to be related through a distant common ancestor are similar, unrelated proteins with the same fold (such as Raf-BD or protein L/G) do not necessarily share the same folding mechanism.[53]

(ii) Intermediate states. The folding of ubiquitin is two-state under most conditions, however, an intermediate can be stabilised and become populated during folding using a number of methods, for example, by the use of a stabilising salt such as sodium sulfate.[47] Some very interesting results from the Searle group have also shown that it is possible to populate an intermediate state by overstabilising an element of secondary structure (see the later section on protein engineering). Kinetic measurements at low temperatures in high viscosity solutions have also been made.[54] In this case, there is strong evidence for an intermediate state not observed at higher temperatures. Under these conditions, folding rates are sufficiently slow that burst phases can be measured accurately.

In recent years, technical advances in other biophysical techniques such as mass spectrometry (MS) and infra-red (IR) have meant that ubiquitin folding kinetics can be monitored with other probes. For example, the two-state model for the unfolding and refolding of ubiquitin has recently been tested by using H/D exchange techniques in conjunction with mass spectrometry.[55] Refolding from an acid–methanol induced denatured state is monitored both by the extent of H/D exchange of the backbone amide groups after different refolding times, and by the charge-state distribution (CSD) which is characteristic of denatured or native protein. When taken individually the results from both probes are consistent with a two-state model, however, a careful comparison revealed an additional species with a CSD characteristic of the native protein but with non-native H/D exchange behaviour, suggesting a transient intermediate might be present.

(iii) Non-exponential kinetics. In addition to the stopped-flow and continuous-flow experiments described above, temperature-jump methods have also been employed to probe ubiquitin's folding kinetics.[56] Temperature-jump was used to rapidly initiate the folding of a destabilised mutant of ubiquitin from a cold-denatured state. Remarkably, highly non-exponential kinetics were observed under some conditions. This was attributed to a complex energy landscape in which the folding of a fraction of the molecules was downhill, there being no substantial energy barrier—these molecules residing on the native side of the rate-limiting transition state barrier after the temperature jump.[56]

Unfolding kinetics of ubiquitin measured with non-linear IR spectroscopy including 2D IR spectroscopy and dispersed vibrational echo (DVE) spectroscopy, have revealed a series of conformational changes on the nanosecond to millisecond timescale during thermal unfolding, some of which also demonstrate non-exponential kinetics.[57] For example, from 100 ns to 0.5 ms, the results are consistent with partial unfolding of the β-sheet and non-exponential kinetics are observed. By modeling the amide I vibrations of ubiquitin, it is proposed that this results from the unfolding of the less stable strands 3–5, before the unfolding of the N-terminal β-hairpin which forms part of the hydrophobic core. This downhill folding is followed by a slower exponential phase corresponding to barrier-crossing kinetics.

In addition to the studies above, protein engineering techniques have been used in conjunction with rapid-reaction kinetics such as stopped-flow and continuous-flow to study the folding pathway of ubiquitin. Results from these studies are discussed in the next section.

Protein engineering

As with just about any other study on protein structure and function, protein engineering has played an important role in our understanding of the structure, stability, folding and dynamics of ubiquitin. The first protein engineering on ubiquitin was work done by the Roder group who engineered a tryptophan at position 45 in the protein in order to be able to use fluorescence to probe the state of the protein in folding studies.[50] Over the past ten years, this F45W variant has been extensively used in folding studies by many other groups. The Roder group were also the first to make mutants of the hydrophobic core residues in order to investigate the role of the core in folding and stability.[22] Other groups have also engineered cavity creating mutations into the core of ubiquitin, in this case they were introduced at various distances from the tryptophan in order to probe the effect on fluorescence.[58] These studies showed that loosening of the structure near the tryptophan resulted in hyperfluorescent species. Thus indicating that hyperfluoresence observed in burst phases may not be due to the formation of compact folding intermediates but due to a pre-transitional conformationally loosened state.

Other protein engineering studies have focused on the N-terminal β-hairpin and, in particular, on the β-turn connecting strands 1 and 2.[59] The sequence NPDG was introduced into a peptide corresponding to the β-hairpin of ubiquitin and resulted in a misalignment of the two strands and non-native interactions between side chains.[59] In contrast, when the sequence was engineered into the full-length protein, a native-like β-hairpin was formed, however the protein was significantly destabilised and

slower to fold.[60] In this case, cosolvents were observed to stabilise not only the non-native conformer but also native-like structure in the denatured ensemble and to accelerate folding.[60] In another study, deletion of Gly10, which forms a β-bulge in the native-like structure, is found to greatly destabilise the native structure and the β-hairpin.[61]

In recent work from the Searle group, an autonomously folding 14-residue β-hairpin unit was engineered into the first β-hairpin of ubiquitin.[62] This stabilised the protein and a ϕ-value analysis (see below) of the β-hairpin established that it was largely formed in the transition state ensemble (TSE). At low concentrations of denaturant, non-linearity was observed in the kinetic chevron plot indicative of a compact collapsed intermediate state which appears to arise from overstabilisation of local interactions. Mutations within the extended β-hairpin (but not elsewhere in the structure) were found to restore two-state behaviour. In an extension to this work, the Searle group then engineered a similar stable mini-hairpin sequence into a turn region in the C-terminal region of the protein. A single Leu→Phe mutation within this sequence resulted in premature collapse of the denatured ensemble and formation of a similar compact intermediate state.[63]

Protein engineering has long been used to study folding pathways and to characterise transition state ensembles.[64,65] Perhaps the best known form of this is ϕ-value analysis developed by Fersht and coworkers.[66] Here, a non-disruptive mutation is made and its effect on the energy levels of the native and transition states is measured and compared. The Sosnick group recently developed an alternative approach where they have extensively engineered metal-binding sites into ubiquitin by the introduction of pairs of histidine residues on the surface.[67] Metal binding is then used to perturb the energy levels in a manner similar to mutation for ϕ-value analysis, and the resulting Ψ-value analysis is used to characterise the structure of the TSE for folding. Their results indicated that ubiquitin folds through a single folding nucleus, the TSE having a common core that contains heterogeneous features on its periphery. The obligate common core has part of the α-helix docked against four out of five correctly aligned β-strands. These results were then compared with the ϕ-values obtained for a limited number of hydrophobic core mutants.[23]

A comprehensive ϕ-value analysis by my own group was published last year.[24] Twenty seven non-disruptive mutations were made throughout the protein structure and a range of ϕ-values from zero to one was observed. Medium and high values were found only in the N-terminal region of the protein, whilst the C-terminal region had consistently low values. In the TSE, the main α-helix appears to be fully formed, and the helix is stabilised by packing against the first β-hairpin which is partially structured. In striking comparison, the C-terminal half of the protein is largely unstructured in the TSE. Ubiquitin, thus, has a relatively polarized folding nucleus (Fig. 4c).

The two different approaches, ϕ- and Ψ-value analysis, generate similar but different results. Both studies indicate that the N-terminal region of ubiquitin, including the first β-hairpin and the α-helix, are structured in the TSE, but whereas Ψ-value analysis suggests that regions of the C-terminus are also present,[67] ϕ-value analysis shows little evidence for this.[24] The reason for this discrepancy is at present unclear—it has been suggested that mutations may change the flux through the TSE,[23] however, the Ψ-value method has recently been re-evaluated and may need further testing.[68,69]

Ubiquitin has been used as a model system to study the effect of loop insertions on the structure and stability of proteins. The Robertson group has engineered loop sequences corresponding to loop regions in structural homologues of ubiquitin into two sites in the protein—the 9–10 loop and the 35–36 loop (Fig. 1).[70] They observed that the effect of the loop is largely dependent upon the position of the insert and not on the sequence or the length of the insert. For example, inserts into the 35–36 loop result in greater structural perturbation than inserts into the 9–10 loop, possibly due to the intrinsic flexibility/stability of the two loops in the wild-type structure.

Atomic force microscopy

In addition to the myriad of techniques discussed above, ubiquitin has recently been the subject of a number of mechanical unfolding studies using atomic force microscopy (AFM) techniques. AFM has been used to test the mechanical strength of polyubiquitin molecules with different linkages.[71] The mechanical strength of N–C linked polyubiquitin was found to differ from a Lys48–C linked polyubiquitin chain. This has established that the mechanical strength of proteins can be modified by the linkage of the domain and may have important biological consequences for ubiquitin which is found in nature in different complexes linked through Lys63, Lys48, Lys29 or Lys11.

Force-clamp AFM has also been used to investigate the mechanical strength, unfolding and refolding properties of a chain of multiple ubiquitin molecules.[72] High forces are used to sequentially unfold domains of polyubiquitin, quenching to lower forces then allows refolding to occur. Various phases are observed which are attributed to elastic recoil and folding processes. It should be noted that folding events are monitored indirectly by unfolding the polyubiquitin chain again after different refolding times.[73]

Using the force-clamp technique, the kinetics of the unfolding of ubiquitin have also been measured at a constant force (in contrast to other AFM experiments which typically measure extension/unfolding at a constant velocity). In this case, N–C linked ubiquitin was used, and the force-driven unfolding was shown to be a Markovian process that depends exponentially on stretching force.[74] Although an ensemble average of the single molecule experiments was well described by a two-state model, unfolding events were observed on a single molecule level which deviate and indicate variant unfolding pathways. These events have a low frequency and are, therefore, not represented in the ensemble average. Recent Monte Carlo simulations of the mechanical unfolding were in agreement with these studies showing that unfolding could occur as a single-step process or through intermediate steps,[75] and that there was a distinct order to the unfolding events.

Dynamics

The dynamics of both the side chains and the backbone of ubiquitin have been studied using a variety of NMR techniques. In a recent paper from the Palmer group, NMR spin relaxation experiments were used to characterise the backbone dynamics on the microsecond timescale and chemical exchange processes were

 This journal is © The Royal Society of Chemistry 2006

identified which affect residues Ile23, Asn25, Thr55 and Val70.[76] The exchange processes affecting residues 23, 25 and 55 appear to result from disruption of the N-cap of the α-helix and possibly from repacking of the side chain of Ile23.

NMR techniques in conjunction with molecular dynamic (MD) methods have also been used to simultaneously determine both the structure and the dynamics of ubiquitin.[77] This novel method, which is called dynamic ensemble refinement (DER), uses experimentally determined structure order parameters (s^2) for the native state of ubiquitin, in addition to distance information from NOESY experiments as restraints in MD simulations. The order parameters contain information on molecular motion on the picosecond to nanosecond timescale. The method requires that a set of ubiquitin conformations is simultaneously consistent with both NOE data and order parameters in order to generate an ensemble of structures. It is found that there is considerable conformational heterogeneity throughout the protein structure. Each structure in the ensemble has a tightly packed core, however, even side chains in the core are found to populate multiple rotameric states and can, therefore, be considered to have liquid-like characteristics (Fig. 5).

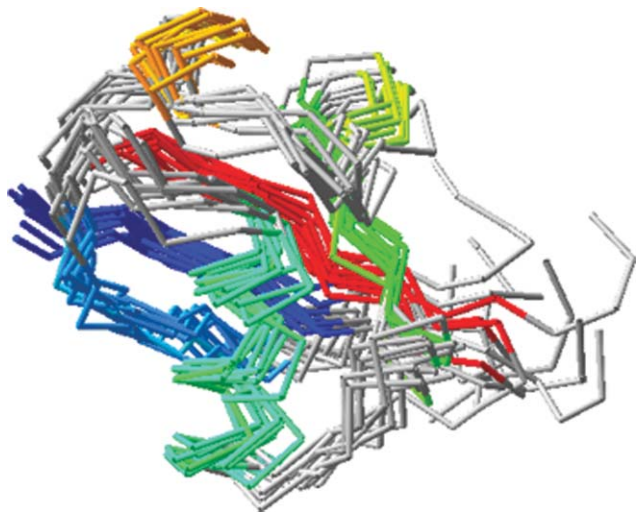

Fig. 5 C_α trace of ten representative conformations of ubiquitin obtained from the dynamic ensemble refinement method.[77] β-Strands 1, 2, 3, 4 and 5 are shown in dark blue, light blue, bright green, yellow and red, respectively. The major α-helix is in greenish-blue, and the 3_{10}-helix in orange.

Redesign of the hydrophobic core

The structural and stability work that had been carried out on ubiquitin by the mid 1990s meant that it was by then an excellent system with which to test *de novo* design strategies that were being developed by computational chemists and biochemists. The Handel group was one of the first to use ubiquitin to test their design algorithms. The work focused on packing arrangements in the hydrophobic cores of proteins, and the program RoC (Repacking of Cores) was used to design nine variants containing between three to eight mutations of hydrophobic core residues.[78] These variants were made and their structure and stability characterised.[78,79] All the designed variants were found to be more stable than controls in which hydrophobic core residues were chosen randomly, however, none was as stable as the wild-type protein. The stabilities of the mutants measured were used to evaluate and improve the core packing algorithm. The solution structure of one of the variants with seven mutations in the hydrophobic core revealed that although the backbone conformation was very similar to the wild-type, the side chain conformations, in general, were in statistically less favourable conformations, thus explaining the lower stability.[51] Further NMR characterisation of this mutant showed that overall the level of dynamics is similar to that of the wild-type, however, the mutations caused a redistribution in the positions of core residues that are dynamic,[80] correlating with the tendency of these residues to populate unfavourable rotameric states. Thus, it seems that strain from poor side chain conformations promotes increased flexibility as a mechanism to relieve unfavourable steric interactions. A subsequent high resolution structural analysis revealed that the variant is in slow exchange between two conformations—the dynamic response and the lower stability are coupled to greater strain and mobility in the core.[81]

In a completely different approach, the Woolfson group have made multiple hydrophobic core variants of ubiquitin using a library of core mutants and an efficient and effective selection procedure.[82] Stable variants from a library of ubiquitin hydrophobic core mutants were synthesized as hexahistidine-tagged fusions and were displayed on the surface of phages. These protein–phages were immobilized onto Ni-coated surfaces and the bound fusion-phages were treated with protease to remove unstable or poorly folded proteins. Stable phage fusions were eluted and infected into *E. coli*, which allowed amplification for further selection, sequencing, or protein expression. Many stable variants were selected using this screen, some of which had up to seven mutations in the hydrophobic core, and it was found that ubiquitin was surprisingly tolerant to substitution of core residues.[83] This is somewhat surprising given the sequence conservation of ubiquitin and the ubiquitin structural superfamily. Interestingly, and in agreement with the Handel group studies, no variant was found which was more stable than wild-type. The structure, folding and dynamics of two of the variants from the screen—one overpacked with seven substitutions (U4), the other underpacked with just two mutations (U7)—was undertaken.[84] Both proteins were well folded and exhibited backbone dynamics similar to the wild-type. A crystal structure of U4 revealed that there were almost no changes in the position of the backbone and that the mutations had been accommodated by small movements in the side chains of both mutated and non-mutated residues, Fig. 6. Despite it being less stable than wild type, U4 was found to fold with similar rates.

Summary

With only 76 residues, ubiquitin is one of the smallest proteins which is not a domain of a larger protein, is not always present as part of a large oligomeric complex and which is functional as a monomer. Its small size, stability and availability have made it the subject of an extensive number of biophysical studies aimed at understanding the fundamental principles governing protein structure, stability, folding and dynamics. There is no doubt that ubiquitin has proved itself as a model system time and time again. Despite its small size and relatively simple architecture,

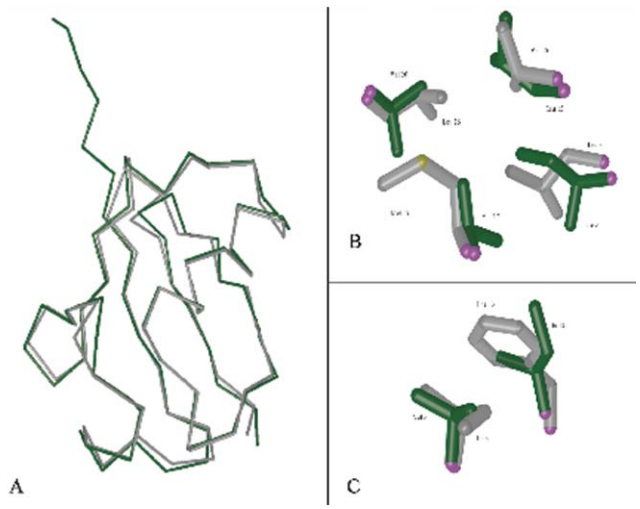

Fig. 6 Diagrams of the crystallographic structures of wild-type (green) and U4 (grey) ubiquitin. **A**. Comparison of the C_α backbones and (**B**. and **C**) side chain orientations of the mutated residues in the mutant U4. **B**. Residues 3, 15, 17, and 26. **C**. Residues 5 and 13. Data taken from Benitez-Cardoza *et al.*, (2003).[84]

ubiquitin shows many of the complexities of larger proteins, including response to mutation, dynamic behaviour, complex folding kinetics (exponential and non-exponential behaviour) *etc.* As such, it is not only a model for how small proteins may fold, but captures many of the features of how larger proteins may fold as well (such as the kinetic traps induced by overstabilisation of secondary structure). I am confident that ubiquitin has not yet revealed all of its many secrets, and that, with advances in experimental techniques and methodologies, we will learn more about the complex energy landscape it can explore. Some important fundamentals question still remain, particularly with regards to linking what we know about the protein's physical properties to its function *in vivo.* Perhaps, we will even find the reason why this robust little protein is so highly conserved.

Acknowledgements

I would like to thank the members of my own research group who have worked on ubiquitin or ubiquitin-like proteins for their interesting and stimulating discussions: Heather Went, Claudia Benitez-Cardoza, Neelan Marianayagam, Andy Brown, Andy Roberts and Anna Mallam. Dr Katherine Stott and Dr Miri Hirshberg provided invaluable expertise in NMR and X-ray crystallography, for the work on the multiple core mutants. I also want to thank Prof. Dek Woolfson for the initial collaboration that brought us into the ubiquitin folding field. I am also grateful to everyone in the ubiquitin folding field for numerous interesting discussions over the last five years and to Bill Broadhurst for providing many of the references for the early NMR work.

References

1 M. Hochstrasser, *Cell*, 2006, **124**, 27–34.
2 R. E. Lenkinski, D. M. Chen, J. D. Glickson and G. Goldstein, *Biochim. Biophys. Acta*, 1977, **494**, 126–130.
3 K. D. Wilkinson and T. K. Audhya, *J. Biol. Chem.*, 1981, **256**, 9235–9241.
4 M. Oliveberg and E. I. Shakhnovich, *Curr. Opin. Struct. Biol.*, 2006, **16**, 68–70.
5 D. Baker and W. A. Eaton, *Curr. Opin. Struct. Biol.*, 2004, **14**, 67–69.
6 *Acc. Chem. Res.*, 2002, **35**, 325–489, special issue on molecular dynamics simulations of biomolecules.
7 S. Vijay-Kumar, C. E. Bugg, K. D. Wilkinson and W. J. Cook, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 3582–3585.
8 S. Vijay-Kumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531–544.
9 G. Cornilescu, J. L. Marquardt, M. Ottiger and A. Bax, *J. Am. Chem. Soc.*, 1998, **120**, 6836–6837.
10 P. D. Cary, D. S. King, C. Crane-Robinson, E. M. Bradbury, A. Rabbani, G. H. Goodwin and E. W. Johns, *Eur. J. Biochem.*, 1980, **112**, 577–588.
11 D. L. Stefano and A. J. Wand, *Biochemistry*, 1987, **26**, 7272–7281.
12 P. L. Weber, S. C. Brown and L. Mueller, *Biochemistry*, 1987, **26**, 7282–7290.
13 D. J. Ecker, T. R. Butt, J. Marsh, E. J. Sternberg, N. Margolis, B. P. Monia, S. Jonnalagadda, M. I. Khan, P. L. Weber and L. Mueller, *J. Biol. Chem.*, 1987, **262**, 14213–14221.
14 P. L. Wintrode, G. I. Makhatadze and P. L. Privalov, *Proteins: Struct., Funct., Genet.*, 1994, **18**, 246–253.
15 G. I. Makhatadze, M. M. Lopez, J. M. Richardson and S. T. Thomas, *Protein Sci.*, 1998, **7**, 689–697.
16 B. Ibarra-Molero, V. V. Loladze, G. I. Makhatadze and J. M. Sanchez-Ruiz, *Biochemistry*, 1999, **38**, 8138–8149.
17 V. V. Loladze, B. Ibarra-Molero, J. M. Sanchez-Ruiz and G. I. Makhatadze, *Biochemistry*, 1999, **38**, 16419–16423.
18 V. V. Loladze and G. I. Makhatadze, *Protein Sci.*, 2002, **11**, 174–177.
19 G. I. Makhatadze, V. V. Loladze, D. N. Ermolenko, X. F. Chen and S. T. Thomas, *J. Mol. Biol.*, 2003, **327**, 1135–1148.
20 M. Sundd and A. D. Robertson, *J. Mol. Biol.*, 2003, **332**, 927–936.
21 S. T. Thomas and G. I. Makhatadze, *Biochemistry*, 2000, **39**, 10275–10283.
22 S. Khorasanizadeh, I. D. Peters and H. Roder, *Nat. Struct. Biol.*, 1996, **3**, 193–205.
23 T. R. Sosnick, R. S. Dothager and B. A. Krantz, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17377–17382.
24 H. M. Went and S. E. Jackson, *Protein Eng., Des. Sel.*, 2005, **18**, 229–237.
25 M. Nei, I. B. Rogozin and H. Piontkivska, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 10866–10871.
26 V. V. Loladze, D. N. Ermolenko and G. I. Makhatadze, *J. Mol. Biol.*, 2002, **320**, 343–357.
27 S. T. Thomas, V. V. Loladze and G. I. Makhatadze, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10670–10675.
28 D. N. Ermolenko, S. T. Thomas, R. Aurora, A. M. Gronenborn and G. I. Makhatadze, *J. Mol. Biol.*, 2002, **322**, 123–135.
29 D. N. Ermolenko, J. M. Richardson and G. I. Makhatadze, *Protein Sci.*, 2003, **12**, 1169–1176.
30 V. V. Loladze and G. I. Makhatadze, *Proteins: Struct., Funct., Bioinf.*, 2005, **58**, 1–6.
31 J. P. L. Cox, P. A. Evans, L. C. Packman, D. H. Williams and D. N. Woolfson, *J. Mol. Biol.*, 1993, **234**, 483–492.
32 R. Zerella, P. Y. Chen, P. A. Evans, A. Raine and D. H. Williams, *Protein Sci.*, 2000, **9**, 2142–2150.
33 R. Zerella, P. A. Evans, J. M. C. Ionides, L. C. Packman, B. W. Trotter, J. P. Mackay and D. H. Williams, *Protein Sci.*, 1999, **8**, 1320–1331.
34 M. Jourdan and M. S. Searle, *Biochemistry*, 2000, **39**, 12355–12364.
35 D. Bolton, P. A. Evans, K. Stott and R. W. Broadhurst, *J. Mol. Biol.*, 2001, **314**, 773–787.
36 R. M. Ionescu and C. R. Matthews, *Nat. Struct. Biol.*, 1999, **6**, 304–307.
37 M. M. Harding, D. H. Williams and D. N. Woolfson, *Biochemistry*, 1991, **30**, 3120–3128.
38 D. N. Woolfson, A. Cooper, M. M. Harding, D. H. Williams and P. A. Evans, *J. Mol. Biol.*, 1993, **229**, 502–511.
39 M. Jourdan and M. S. Searle, *Biochemistry*, 2001, **40**, 10317–10325.
40 R. Kitahara and K. Akasaka, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3167–3172.
41 R. Kitahara, S. Yokoyama and K. Akasaka, *J. Mol. Biol.*, 2005, **347**, 277–285.
42 C. R. Babu, V. J. Hilser and A. J. Wand, *Nat. Struct. Mol. Biol.*, 2004, **11**, 352–357.

 This journal is © The Royal Society of Chemistry 2006

43 J. Jonas, *Biochim. Biophys. Acta*, 2002, **1595**, 145–159.
44 M. S. Briggs and H. Roder, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 2017–2021.
45 S. T. Gladwin and P. A. Evans, *Folding Des.*, 1996, **1**, 407–417.
46 B. A. Krantz and T. R. Sosnick, *Biochemistry*, 2000, **39**, 11696–11701.
47 H. M. Went, C. B. Benitez-cardoza and S. E. Jackson, *FEBS Lett.*, 2004, **567**, 333–338.
48 T. Sivaraman, C. B. Arrington and A. D. Robertson, *Nat. Struct. Biol.*, 2001, **8**, 331–333.
49 F. Cordier and S. Grzesiek, *J. Mol. Biol.*, 2002, **317**, 739–752.
50 S. Khorasanizadeh, I. D. Peters, T. R. Butt and H. Roder, *Biochemistry*, 1993, **32**, 7054–7063.
51 B. A. Krantz, L. B. Moran, A. Kentsis and T. R. Sosnick, *Nat. Struct. Biol.*, 2000, **7**, 62–71.
52 A. Vallee-Belisle, J.-F. Turcotte and S. W. Michnick, *Biochemistry*, 2004, **43**, 8447–8458.
53 N. J. Marianayagam, A. G. Brown, H. M. Went, A. L. Mallam, C. G. Benitez-Cardoza and S. E. Jackson, submitted.
54 Z. Qin, J. Ervin, E. Larios, M. Gruebele and H. Kihara, *J. Phys. Chem. B*, 2002, **106**, 13040–13046.
55 J. Pan, D. J. Wilson and L. Konermann, *Biochemistry*, 2005, **44**, 8627–8633.
56 J. Sabelko, J. Ervin and M. Gruebele, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6031–6036.
57 H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim and A. Tokmakoff, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 612–617.
58 J. Ervin, E. Larios, S. Osvath, K. Schulten and M. Gruebele, *Biophys. J.*, 2002, **83**, 473–483.
59 M. Jourdan, S. R. Griffiths-Jones and M. S. Searle, *Eur. J. Biochem.*, 2000, **267**, 3539–3548.
60 G. W. Platt, S. A. Simpson, R. Layfield and M. S. Searle, *Biochemistry*, 2003, **42**, 13762–13771.
61 P.-Y. Chen, B. G. Gopalacushina, C.-C. Yang, S. I. Chan and P. A. Evans, *Protein Sci.*, 2001, **10**, 2063–2074.
62 R. Bofill, E. R. Simpson, G. W. Platt, M. D. Crespo and M. S. Searle, *J. Mol. Biol.*, 2005, **349**, 205–211.
63 R. Bofill and M. S. Searle, *J. Mol. Biol.*, 2005, **353**, 373–384.
64 A. Matouschek, J. T. Kellis, Jr., L. Serrano and A. R. Fersht, *Nature*, 1989, **342**, 122–126.
65 C. R. Matthews, *Methods Enzymol.*, 1987, **154**, 489–511.
66 A. R. Fersht, A. Matouschek and L. Serrano, *J. Mol. Biol.*, 1992, **224**, 771–782.
67 B. A. Krantz, R. S. Dothager and T. R. Sosnick, *J. Mol. Biol.*, 2004, **337**, 463–475.
68 A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17327–17328.
69 B. A. Krantz, R. S. Dothager and T. R. Sosnick, *J. Mol. Biol.*, 2005, **347**, 1103.
70 D. M. Ferraro, E. K. Hope and A. D. Robertson, *J. Mol. Biol.*, 2005, **352**, 575–584.
71 M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser and J. M. Fernandez, *Nat. Struct. Biol.*, 2003, **10**, 738–743.
72 J. M. Fernandez and H. Li, *Science*, 2004, **303**, 1674–1678.
73 R. B. Best and G. Hummer, *Science*, 2005, **308**, 498.
74 M. Schlierf, H. B. Li and J. M. Fernandez, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 7299–7304.
75 A. Irback, S. Mitternacht and S. Mohanty, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13427–13432.
76 F. Massi, M. J. Grey and A. G. Palmer, *Protein Sci.*, 2005, **14**, 735–742.
77 K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson and M. Vendruscolo, *Nature*, 2005, **433**, 128–132.
78 G. A. Lazar, J. R. Desjarlais and T. M. Handel, *Protein Sci.*, 1997, **6**, 1167–1178.
79 E. C. Johnson, G. A. Lazar, J. R. Desjarlais and T. M. Handel, *Structure*, 1999, **7**, 967–976.
80 E. C. Johnson and T. M. Handel, *J. Biomol. NMR*, 1999, **15**, 135–143.
81 G. A. Lazar, E. C. Johnson, J. R. Desjarlais and T. M. Handel, *Protein Sci.*, 1999, **8**, 2598–2610.
82 M. D. Finucane, M. Tuna, J. H. Lees and D. N. Woolfson, *Biochemistry*, 1999, **38**, 11604–11612.
83 M. D. Finucane and D. N. Woolfson, *Biochemistry*, 1999, **38**, 11613–11623.
84 C. Benitez-Cardoza, K. Stott, M. Hershberg, H. M. Went, D. N. Woolfson and S. E. Jackson, *Biochemistry*, 2003, **43**, 5195–5203.
85 S. E. Jackson and A. R. Fersht, *Biochemistry*, 1991, **30**, 10428–10435.
86 A. Matouschek, J. T. Kellis, Jr., L. Serrano, M. Bycroft and A. R. Fersht, *Nature*, 1990, **346**, 440–445.
87 M. Silow and M. Oliveberg, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 6084–6086.