Synthetic, structural and biological studies of the ubiquitin system: synthesis and crystal structure of an analogue containing unnatural amino acids

Steven G. LOVE*, Tom W. MUIR*, Robert RAMAGE*‖, Kevin T. SHAW*, Dmitriy ALEXEEV†, Lindsay SAWYER†, Sharon M. KELLY‡, Nicholas C. PRICE‡, Jane E. ARNOLD§, Maureen P. MEE§ and R. John MAYER§

*Department of Chemistry, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, Scotland, U.K., †Structural Biochemistry Group, University of Edinburgh, Swann Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, Scotland, U.K., ‡Department of Biological and Molecular Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, U.K., and §Department of Biochemistry, University of Nottingham Medical School, Queens Medical Centre, Nottingham NG7 2UH, U.K.

Ubiquitin is a 76-amino acid protein involved in the targeting for destruction of proteins in the cell. The protein can readily be synthesized chemically affording an extra dimension to studies of protein stability. Ubiquitin with various modifications to the hydrophobic core has been synthesized. In particular, two core amino acids have been replaced by aminobutyric acid (Val-26) and norvaline (for Ile-30) and the product crystallized. The refined crystal structure shows an overall contraction of the molecule and the side chain of Nva-30 rotates relative to Ile-30. However, the side chain rotation is not sufficient to compensate for the effect of the loss of the methyl group and hence a small cavity is introduced into the structure, which decreases the stability of the protein. The biological behaviour of the modified protein is unaltered. The observed changes in stability are of the magnitude expected for the removal of methyl groups from the hydrophobic core of a protein. Interestingly, the effect appears to be independent of the position of the removed methyl group. The intact structure, but not its stability, is important for recognition by the biological conjugating system.

INTRODUCTION

Until recently the effect of amino acid substitutions on protein folding has been limited to replacement by another one of the twenty DNA-specified amino acids. Chemical synthesis of small proteins can utilize a wide variety of novel α-amino acids and hence allow sophisticated probing of the protein structure, rather than merely removing undesired side-chain functional groups.

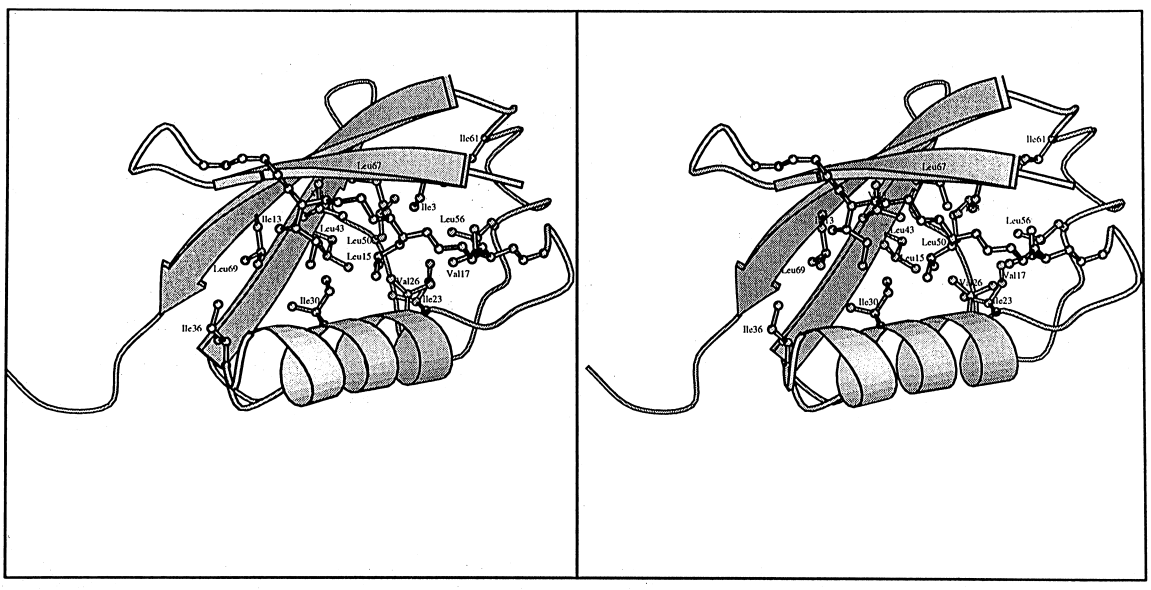


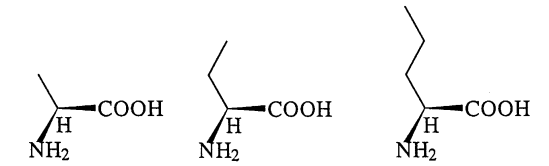
Figure 1 A MOLSCRIPT [15] stereodrawing of ubiquitin showing the hydrophobic residues that form the core

Abbreviations used: Abu, aminobutyric acid; Ub(des-Core), ubiquitin(Val5Gly, Val26Ala, Ile30Ala, Leu43Gly, Leu50Gly, Leu56Gly, Leu67Gly); DMF, dimethylformamide; Fmoc, fluorenylmethoxycarbonyl; Nva, norvaline; TFA, trifluoroacetic acid; Tbfmoc, tetrabenzofluorenylmethoxycarbonyl; TMSbr, trimethylsilyl bromide; GuHCl, guanidine hydrochloride; r.m.s., root mean square.

‖ To whom correspondence should be addressed.

Not only has recent synthetic work led to the total synthesis of D-HIV protease [1] and also of larger proteins using the methods of chemical ligation [2,3] but the syntheses have also facilitated studies with model peptide systems.

It has been shown that N^{α}-aminobutyric acid (Abu) behaves in model peptides in a similar manner to alanine (see the structures below), i.e. it stabilizes the α-helix to a greater extent than, Leu, Ile, Val, Met and Phe [4–10], and, indeed, unusual α-amino acids do appear in naturally occurring peptides such as cyclosporin. However, what is as yet unknown about amino acids such as Abu and norvaline (Nva) is how they affect the side-chain packing and hence the stability of a protein core, given the increased degrees of freedom and flexibility that these linear side-chain amino acids have in comparison with Val and Ile respectively. The effect of linear side chains on the process of protein folding is also unknown.



Alanine (Ala) Aminobutyric acid (Abu) Norvaline (Nva)

Bovine ubiquitin provides an ideal system for investigating such amino acid substitutions in a protein. The overall tertiary structure is well understood in solution [11,12] and also in the crystalline state [13,14]. A major contributor to the structure of ubiquitin (Figure 1) is the α-helix which is packed on to a sheet of four β-strands. This α-helix–β-sheet packing forms a compact hydrophobic core which involves three residues from the α-helix (Ile-23, Val-26 and Ile-30) and eleven of the hydrophobic residues from the β-sheet (Ile-3, Val-5, Ile-13, Leu-15, Val-17, Ile-36, Leu-43, Leu-50, Leu-56, Ile-61 and Leu-67). These residues which make up the hydrophobic core of the protein are shown in the stereo diagram Figure 1. The importance of the helix–sheet interaction to the stability of the core is emphasized by the ubiquitin analogue, Ub(des-Core), in which the hydrophobic core amino acids were replaced by glycine or alanine and the resulting product was devoid of tertiary structure [16]. This highlights the crucial role played by the hydrophobic interactions in the development of the tertiary structure of ubiquitin.

Stop-flow NMR studies in solution have been carried out in order to define the events involved in ubiquitin folding [17]. From these studies it is found that the folding process is extremely fast, with formation of the secondary-structural elements and completion of the overall tertiary structure being highly co-operative. It appears that association of the α-helix and the β-sheet occurs first and probably is required for the stabilization of the α-helix.

We have shown that we can synthesize ubiquitin chemically [18] and that it is identical in all respects with bovine ubiquitin, as judged by X-ray crystallography [14]. There are several reports of unnatural amino acids being incorporated into a natural sequence using recombinant methods, but the expression of the resulting proteins has not been sufficient for detailed structural investigation [19]. However, synthetic proteins with unnatural amino acids incorporated have been crystallized, and the first reported structure was that of HIV protease, where the cysteine residues were replaced by Abu to overcome handling problems [20]. Now, with the successful and efficient synthesis of ubiquitin, we have an ideal method for investigating the effects of the incorporation of unnatural amino acids into the protein so that we can modify the core progressively and by the smallest increments (e.g. single $-CH_3/-CH_2-$, but also $-CH_2F$, $-CHF_2$, $-CF_3$). With the helix-stabilizing effects of linear-chain amino acids having already been demonstrated in model peptide helices [10], we have replaced Val-26 by Abu (Val26Abu) and Ile-30 by norvaline (Ile30Nva) in the natural sequence, at the same time demonstrating the application of new methodology designed to effect the speedy purification of the synthetic protein.

Thus in this paper we illustrate the potential of using non-coded amino acids to perturb protein structure in a manner that is not possible with conventional amino acid substitution by recombinant techniques. We analyse the effects of unnatural amino acid substitutions on the structure, stability and activity of ubiquitin. Such perturbations add a further dimension to studies of protein folding and may provide information as to why the particular set of 20 amino acids coded by DNA has arisen.

MATERIALS AND METHODS

Protein synthesis

The protein was synthesized on a polystyrene-based Wang resin as described previously [18]. The dried peptide resin was Fmoc-deprotected for 15 min with 20% piperidine in dimethylformamide (DMF). The resin was washed with DMF (3×30 ml) then trichloromethane (3×30 ml). To the washed dried peptide resin was added tetrabenzofluorenylmethoxycarbonyl chloride (Tbfmoc-Cl) and di-isopropylethanolamine (six fold equivalents of both compared with the peptide N-terminus) in trichloromethane and the mixture sonicated for 3 h at 25 °C, protected from the light. After filtering and washing the resin, the protein resin was checked for the loading of the Tbfmoc group by deprotecting a small sample in 20% piperidine in dioxan and monitoring the resulting deprotection solution at 364 nm.

Cleavage of the peptide from the resin was carried out using trimethylsilyl bromide (TMSBr) (1.35 ml), thioanisole (1.2 ml), *m*-cresol (0.2 ml), ethanedithiol (0.6 ml) and trifluoroacetic acid (TFA) (7.48 ml). Before the addition of the TMSBr, the TFA solution and the scavengers were cooled to 0 °C. The reaction mixture was left at 0 °C under N_2, in the dark for 4.5 h. The crude Tbfmoc protein was isolated after precipitation with ether and gel filtration (Sephadex G-50; 1000 mm × 30 mm) using aq. 30% acetic acid/2% 2-mercaptoethanol, and then dissolved in 10% acetonitrile in aq. 0.1% TFA containing 2% 2-mercaptoethanol and stored in the dark until required. The crude Tbfmoc protein was purified by preparative HPLC on a C_8 Aquapore column (250 mm × 2.6 mm), monitoring the A_{364}, and running a gradient of 0.1% TFA in 10% water/90% acetonitrile to 90% water/10% acetonitrile. Deprotection (20% piperidine in DMF) gave the analogue of ubiquitin, [Abu^{26},Nva^{30}]Ub, which was further purified [21] to afford the synthetic protein (33.5 mg). The mass spectrum exhibited MH^+8538 (expected MH^+8537).

Tryptic digestion of [Abu^{26},Nva^{30}]Ub was carried out as previously described [21]. Peptide fragments were isolated by HPLC using a C_{18} Vydac column (250 mm × 4.6 mm) with monitoring at 205 nm. Solvents used were 0.1% TFA/0.1% triethanolamine in water and 0.1% TFA/0.1% triethanolamine in 40% acetonitrile/water. The isolated peptide fragments were identified by amino acid analysis, which showed them to be consistent with the required sequence. The protein [Abu^{26},Nva^{30}]Ub was sequenced for the first 30 residues to provide unequivocal amino acid assignment for the 30 N-terminal amino acids including the two unnatural α-amino acids.

Other modified ubiquitins, [Ala^{26},Ala^{30}]Ub and Ub(des-Core),

Table 1 Refinement statistics for the X-ray structure

$R_{merge} = \Sigma_h|I_h - \langle I_h\rangle|/\Sigma_h|I_h|$; $R_{scale} = 2\Sigma_h|I_h' - I_h|/\Sigma_h|I_h + I_h'|$

	Mean	r.m.s. deviation	Maximum	Target
Bond length (Å)	0.001	0.017	0.059	0.02
Bond angle (°)	0.29	1.53	6.93	3.00
Torsion angle (°)	5.00	23.26	46.44	15.00
Chiral volume (Å^3)	0.012	0.015	0.030	0.01
Planar groups (Å)	0.022	0.026	0.064	0.02
Bad contacts (Å)	0.086	0.106	0.258	0.10

were synthesized, purified and characterized in a similar manner [16,22].

Iodination procedure

Ubiquitin and its analogues were radiolabelled by the chloramine T method [23] using 37 MBq Na^{125}I/mg of protein.

Conjugation studies

Rabbit reticulocyte DEAE-cellulose ubiquitin-depleted fraction (FII) was prepared by the method of Hershko et al. [24]. Samples (30 μg) were incubated with ^{125}I-ubiquitin (10^6 c.p.m.) in 12.5 mM Tris/HCl buffer, pH 7.5, containing 3 mM $MgCl_2$, 6 mM ATP, 3 mM dithiothreitol, for 2 h at 37 °C in a final volume of 40 μl. Control samples were incubated with EDTA (9 mM) in the absence of ATP and $MgCl_2$. Lysozyme (50 μg) was added to incubation mixtures where indicated.

Reactions were stopped by the addition of 2.5% (v/v) SDS and 1.25% (v/v) 2-mercaptoethanol and boiled for 2 min. Samples were subjected to PAGE (10% acrylamide) in the presence of 0.1% SDS at 20 mA by the method of Laemmli [25]. Gels were dried and the ubiquitin conjugates visualized by autoradiography.

Crystallography

The purified synthetic analogue [Abu^{26},Nva^{30}]Ub was used to grow crystals by repeated seeding, since by this technique crystals can be reproducibly obtained [14]. A seed crystal of native ubiquitin was carefully added to a single 10 ml drop containing 5 ml of mutant ubiquitin solution (20 mg/ml) and 5 ml from the well solution [30% (w/v) poly(ethylene glycol) 4000 in 50 mM cacodylate/HCl, pH 5.6]. Small single crystals, independent of the seed crystal in the original drop, were transferred into fresh drops of the same composition for further growth. The poly(ethylene glycol) concentration was gradually increased to 35% in order to stabilize the crystals.

A data set to a maximum resolution of 1.7 Å (0.17 nm) was collected from one crystal using a Siemens–Nicholet–Xentronix area detector on a Rigaku rotating anode generator RU200 operating at 40 kV, 100 mA and fitted with a graphite monochromator to select Cu-Kα radiation. The space group $P2_12_12_1$ was unaltered from the native but there was a small change in the unit cell parameters: native, $a = 50.83$, $b = 42.74$, $c = 28.98$ Å; mutant, $a = 50.48$, $b = 42.95$, $c = 28.82$ Å. There were 14404 observations representing 6355 unique reflections with $I > \sigma(I)$ (85% complete) to 1.7 Å after reduction using the Xengen v.1.3 program [26] to give an unweighted absolute value for the R_{merge} (see Table 1) on intensity of 5.3%. A total of 5028 reflections in the range between 7.0 and 1.9 Å (96% complete) was used for model refinement. Comparison of the synthetic mutant and native ubiquitin structure factors gave an $R_{scale} = 5.5\%$ [27]. The Fourier synthesis with coefficients ($|F_{mut}| - |F_{nat}|$) and phases derived from native ubiquitin model [14], when inspected using the program O [28] running on an Silicon Graphics Indy 2 molecular graphics workstation, showed one positive and three negative peaks, the maximum being around 5 times the root mean square (r.m.s.) deviation of the map. The molecular model of native

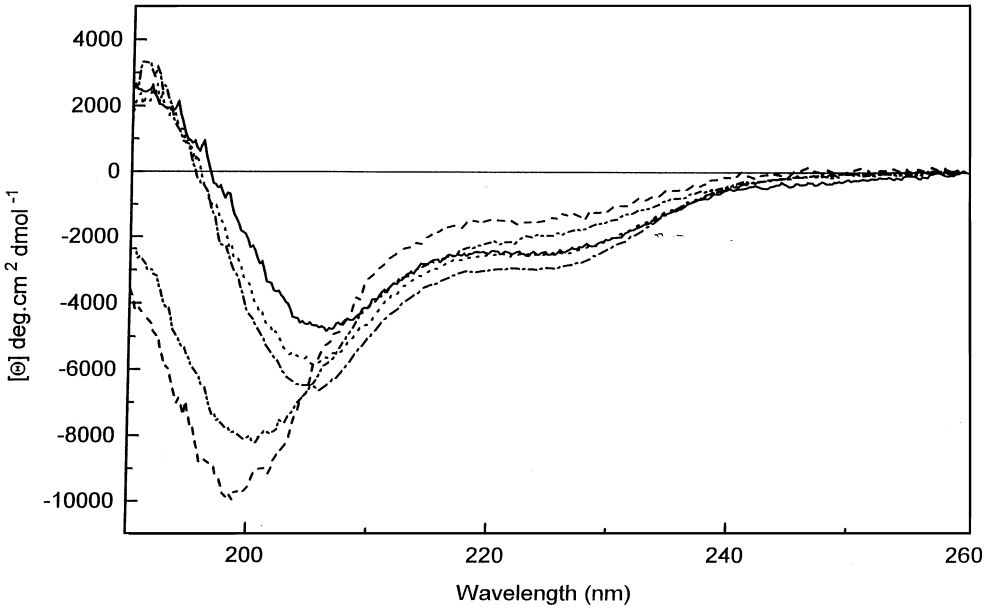


Figure 2 CD spectra for native and various mutant forms of ubiquitin

——, Synthetic native; – – –, Ub(des-Core); · · · ·, [Abu^{26},Nva^{30}]Ub; –·–·–·, [Ala^{26},Ala^{30}]Ub; –·–·, [Ala^{26}]Ub.

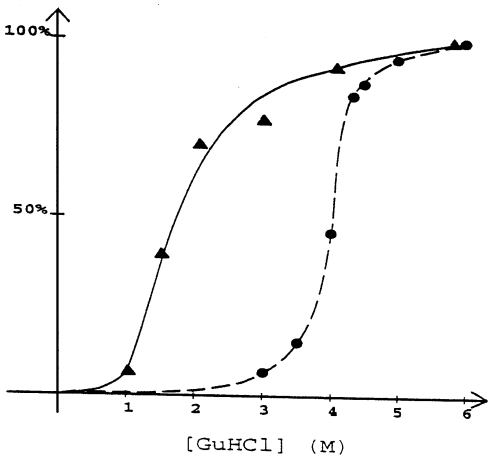


Figure 3 Percentage variation of the ellipticity at 222 nm relative to that in 6 M GuHCl, as a function of the GuHCl concentration at 25 °C and pH 5.0

●, Native protein; ▲, [Abu^{26},Nva^{30}]Ub. The [GuHCl] for the mutant protein is 1.9 M at the midpoint of the transition and that for the native protein is 4.1 M.

ubiquitin was used as a starting model with the substitution of Abu for Val-26 and Nva for Ile-30. This model was refined using release 4 of the TNT restrained least-squares procedure [29]. Because the last three residues at the C-terminus are flexible and therefore poorly defined in both our [14] and the original [13] electron-density maps, throughout the refinement their positions and occupancies were fixed as in the original model. Table 1 summarizes these results of the refinement. The final *R*-factor was 18.5%. XPLOR [30] was used to energy-minimize the final models after TNT refinement. The same starting parameters, including water molecules, were used in each case.

Circular dichroism

Measurements were carried out using a JASCO J-600 spectropolarimeter on synthetic ubiquitin and its analogues which afforded the curves shown in Figure 2. Application of the CONTIN algorithm of Provencher and Glöckner [31] gave the factions of secondary structure. The CD of samples of native and Abu/Nva-mutated protein in 50 mM acetate buffer, pH 5.0, were monitored at 208 nm as a function of the temperature between ambient and 85 °C. Instrumental limitations prevented the melting of native ubiquitin being observed. The CD at 222 nm was also monitored at 25 °C as a function of guanidine hydrochloride (GuHCl) concentration in 25 mM acetate buffer, pH 5.0. The ellipticity at 250 nm was used as a baseline for these measurements. The results of these titrations are shown in Figure 3.

RESULTS AND DISCUSSION

The novel purification protocol for synthetic ubiquitin which had been modified at the N-terminus by the Tbfmoc group making it ideal for separation by hydrophobic interaction chromatography [21], produced a protein that was incorrectly folded, as judged by NMR spectroscopy, in comparison with previous purifications of ubiquitin unmodified at the N-terminus [18]. Thus we believe that the α-helix and the first strands of the β-sheet are mutually stabilizing and that productive folding requires the normal N-terminus [22]. The correct refolding of the protein produced by the Tbfmoc purification was effected in a separate step after removal of the protecting group.

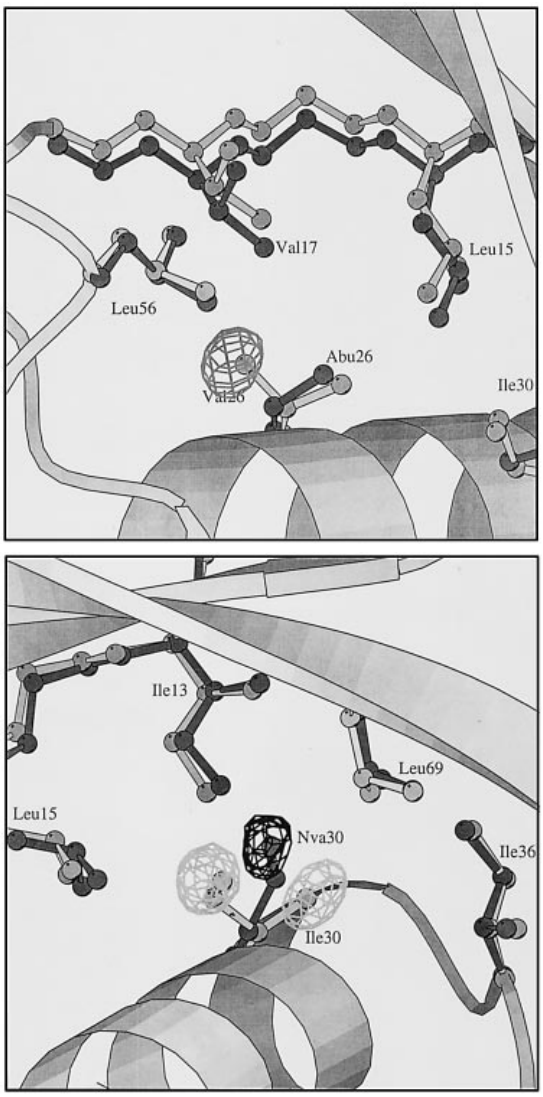


Figure 4 Difference electron-density between the mutant structure and the native ubiquitin in a map calculated with coefficients ($|F_{mut}| - |F_{nat}|$) and phases derived from the ubiquitin model

Both positive and negative contours drawn at the 3.0 σ level are shown. The positive contour is bold (round Nva-30) and the negative ones are pale. Top, the region around Abu^{26} with the negative peak corresponding to the loss of one Val methyl group, CG2. Bottom, the region at Nva-30 showing the negative contours for the loss of methyl CG2 and the shift of the ethyl group from its native position, CD1, to the positive contour between the negative peaks.

Ubiquitin with two unnatural amino acids in the core was synthesized and purified quickly and efficiently using the affinity purification method where the final N-terminal coupling includes Tbfmoc rather than Fmoc, which provides an efficient means of separating the complete protein from any deletion and capped peptides.

The X-ray structure of the protein analogue has a mean deviation from ideal bond lengths of less than 0.001 Å, r.m.s. = 0.016 Å with the maximum deviation being 0.059 Å or 2.9 σ. Mean bond angle deviations are 0.29°, r.m.s. = 1.53° and maxi-

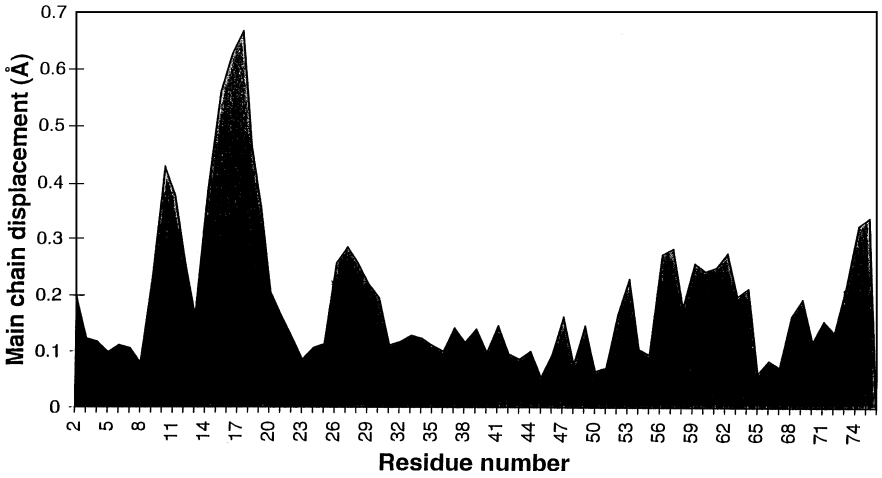


Figure 5 R.m.s. deviation of the main chain atoms of the mutant ubiquitin from their positions in the native protein

mum deviation is 6.9°, that is 2.3 σ. The hydrogen-bond network and dihedral angles of the main chain are very similar to those of the starting model. Figure 4 depicts the principal negative and positive electron-density features in the map, centred around residue 26 (top) and 30 (bottom), from which it can be seen that there is a rotation about χ_1 (of $\sim$ 30°) which repositions the side chain of Nva-30 into the positive density. The other negative peaks correspond to the loss of the two-CH_3 groups on conversions Val26Abu and Ile30Nva. The conformation of the Nva side chain indicates that it is eclipsed, and so presumably the other minor readjustments such as the slight rotation of Ile-13 produce more favourable hydrophobic packing interactions. Figure 5 shows the mean displacement of the main-chain and side-chain atoms of the refined model relative to the native structure. The principal movements of the main chain involve the region between 14 and 20, and the whole structure has contracted slightly to compensate for the loss of the two methyl groups (see Figure 4, top). It is interesting to note that the decrease in the unit cell volume that we observe corresponds to twice the volume of the missing methyl groups, which is what might be expected in a crystal that contains typically 50 % solvent. There is always the possibility that small changes in cell dimensions are the result of an inaccurate crystal to detector distance used in the initial data processing. In this case, using the native cell parameters leads to a final *R*-factor that is about 0.5 % higher than when the correct parameters are used. Thus in this case we believe that the cell parameter changes are real.

The substitutions form cavities

Figure 6 shows the differences in the internal cavities caused by the loss of the two $-CH_3$ groups. The atoms are assumed to be spheres of standard van der Waals' radius. The volumes in the native protein core occupied by the methyl groups that are absent from the mutant protein are shown in the top panels of Figure 6 with Val-26 on the left and Ile-30 on the right. The bottom diagram shows the real cavities created by the mutations: the rotation of the Nva-30 side chain (right panel) splits the cavity created by the mutation Ile30Nva into two parts. The amino acids of the β-strand opposite the helix (Val-17, Leu-15, Ile-13) move into the cavities as partial compensation. However, the volumes that are vacated by the methyl groups of Ile-30 and Val-26 are only partially filled by the movements of the neighbouring amino acids. The volumes filled by these movements are shown in the middle panels of Figure 6. Thus the cavities formed are only partially filled, which leads to a net destabilization of the structure. The other compensating movements of the remainder of the structure are smaller and are not shown.

The melting curves measured for native and mutant ubiquitin by CD at 208 nm (results not shown) indicated that the melting temperature T_m for the mutant that occurred around 75 °C is significantly below that of the native where no change was observed below 85 °C. This wavelength is appropriate for monitoring the breakdown of α-helix. However, convincing evidence for the destabilization of the core by the loss of the two methyl groups came from GuHCl denaturation, the results of which are shown in Figure 3. The wavelength chosen for this experiment, 222 nm, is characteristic of both α-helix and β-sheet and would therefore report on the overall loss of secondary structure of the protein. It also avoids the artifacts introduced by the high absorbance of GuHCl at lower wavelengths. The CD experiment is therefore monitoring the same process as that in the melting experiment described above. The mid-point of the sigmoidal denaturation curve for the native protein is about 4 M GuHCl, and our curve is superimposable on that of Roder and coworkers [32]. The shape of the curve for the double-mutant form is similar to that of the native but with the mid-point shifted to 1.9 M GuHCl, indicating a significant destabilization. The destabilization ($\Delta\Delta G$) is of the order of 14 kJ/mol for the loss of the two-CH_3 groups and is close to the value in Table 2 derived from the fluorescence change observed when the Phe45Trp/Val26Ala double mutant of ubiquitin is titrated with GuHCl [33].

Further evidence for the importance of the interactions involving both of these residues to the stabilization of the helix and the sheet in ubiquitin is provided by the observation that the Val26Ala mutation, in which there is also a loss of two $-CH_3$ groups but in this case from the same residue, is destabilized by a similar amount (Table 2 and Figure 2).

If the double mutation Val26Ala/Leu30Ala is made (a loss of five $-CH_3$ or $-CH_2$- groups) the protein aggregates and cannot be obtained in a soluble form suitable for NMR characterization in aqueous solution. However, it was possible to obtain CD and fluorescence measurements, which indicated loss of helix and significant destabilization (Figure 2 and Table 2).

Similar findings for the loss of a methyl or methylene group have been described in the case of repressor of primer protein,

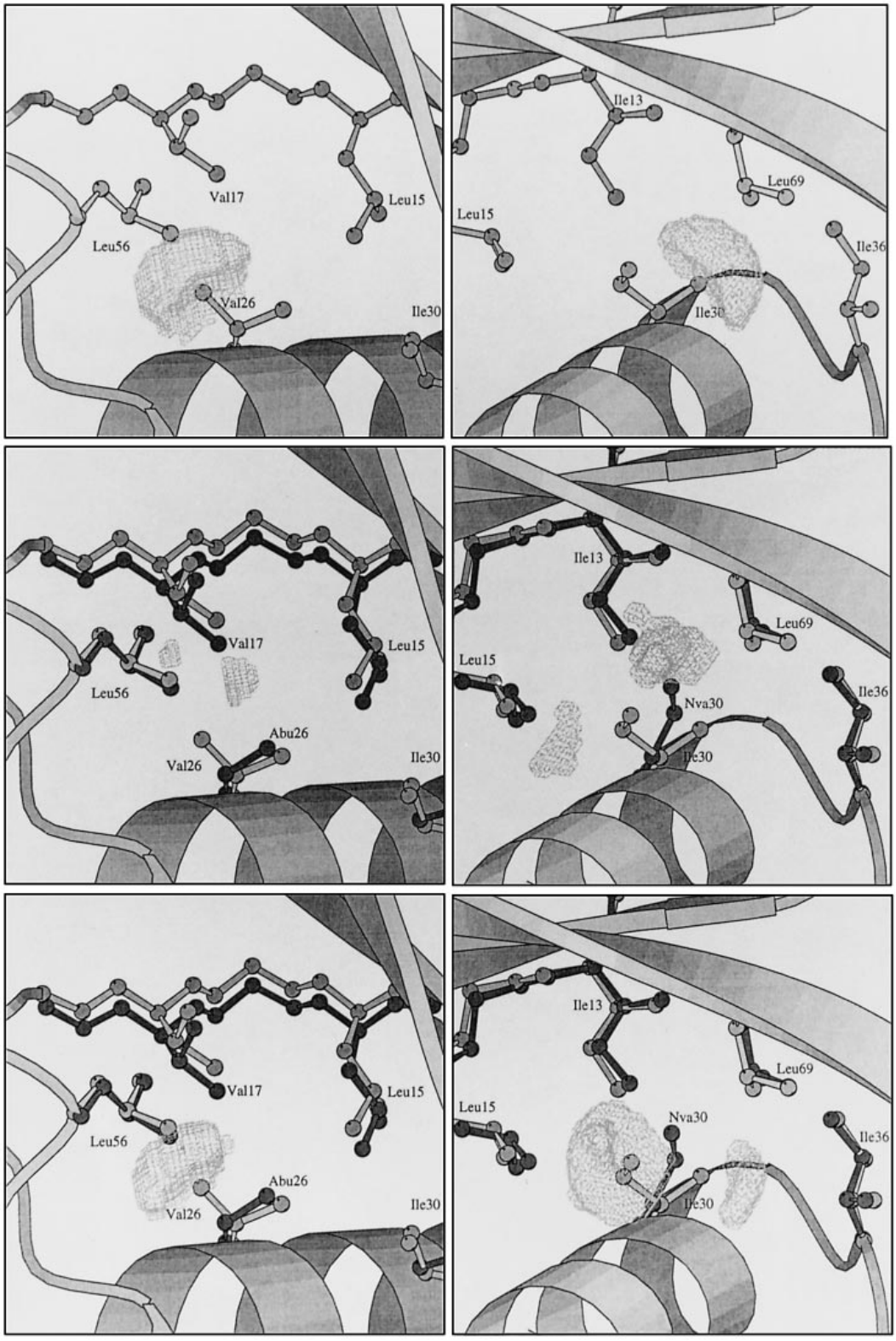


Figure 6 MOLSCRIPT stereodrawing showing cavities in ubiquitin structure created by Val26Abu and Ile30Nva mutations

Top, native ubiquitin structure showing the cavities that would be created by the removal of the methyl groups from residues Val-26 (left) and Ile-30 (right) if the remainder of the molecule did not move. Bottom, native and mutant structures of ubiquitin showing the real cavities created by the loss of the core methyl groups by mutations Val26Abu (left) and Ile30Nva (right). Middle, compensatory movements of the neighbouring side chains partially fill the cavities shown in the top panels when the two methyl groups are removed from the otherwise native structure. The views are around Val-26 (left) and Ile-30 (right).

Table 2 Data for the ubiquitin analogues studied

The difference in the number of methyl or methylene groups different from the native structure. CD spectra are shown in Figure 2. The percentages of sheet and helix are those derived by the method of Provencher and Glöckner [31]. *Data from Khorosanizadeh et al. [33]. The dashed lines in the structure indicate the methyl/methylene groups that have been removed.

Structure	Protein	Difference	Stability [$\Delta\Delta G$ (kJ/mol)]	CD (%)			Activity (Conjugates)
				α-Helix	β-Sheet	Random	
	Val^{26}, Ile^{30}	0	0	16	55	29	Yes
	Ile^{26}, Ile^{30*}	+1	+4.6	–	–	–	–
	Leu^{26}, Ile^{30*}	+1	+1.3	–	–	–	–
	Abu^{26}, Nva^{30}	−2	−13.9	11	56	33	Yes
	Ala^{26}, Ile^{30*}	−2	−13.9	20	62	18	–
	Gly^{26}, Ile^{30*}	−3	−26.5	–	–	–	–
	Ala^{26}, Ala^{30}	−5	−29.4	6	54	39	Yes
	des-Core	−24	−120	0	52	48	No

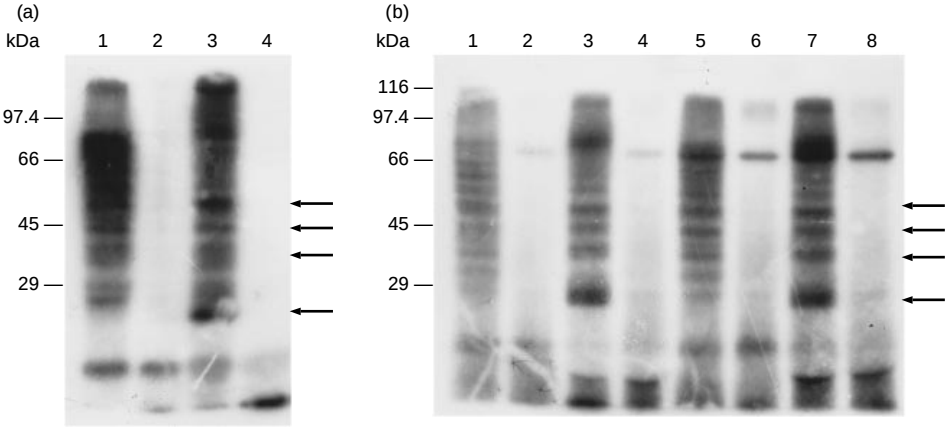


Figure 7 Conjugation of ubiquitin and ubiquitin analogues to endogenous and exogenous proteins

The conjugations were carried out as described in the Materials and methods section with all ubiquitin samples present at 10^6 c.p.m. The arrows indicate the addition of multiple ubiquitin molecules to endogenous protein or to lysozyme, where present. (**a**) Lane 1, ^{125}I-labelled bovine ubiquitin + ATP; lane 2, ^{125}I-labelled bovine ubiquitin − ATP; lane 3, ^{125}I-labelled bovine ubiquitin + ATP + lysozyme (50 μg); lane 4, ^{125}I-labelled bovine ubiquitin − ATP + lysozyme (50 μg). (**b**) Lane 1, ^{125}I-[Ala^{26},Ala^{30}]Ub + ATP; lane 2, ^{125}I-[Ala^{26},Ala^{30}]Ub − ATP; lane 3, ^{125}I-[Ala^{26},Ala^{30}]Ub + ATP + lysozyme (50 μg); lane 4, ^{125}I-[Ala^{26},Ala^{30}]Ub − ATP + lysozyme (50 μg); lane 5, ^{125}I-[Abu^{26},Nva^{30}]Ub + ATP; lane 6, ^{125}I-[Abu^{26},Nva^{30}]Ub − ATP; lane 7, ^{125}I-[Abu^{26},Nva^{30}]Ub + ATP + lysozyme (50 μg); lane 8, ^{125}I-[Abu^{26},Nva^{30}]Ub − ATP + lysozyme (50 μg).

which is a four-helical bundle [34]. Further, a recent analysis of mutants Leu → Val and Leu → Ala concludes that the loss of a single methylene group destabilizes the protein by around 5 kJ/mol, in keeping with similar observations made for monomeric protein [35], where the values for Leu → Ala and Val → Ala are given as 13.0 and 9.6 ± 6.3 kJ/mol. Thus our values, obtained for mutants containing unnatural amino acids, are consistent with those of other workers.

Effects of modification on biological conjugation

We have examined the effects of subtle amino acid modification on the protein structure and have also tested the effects of these changes on the biological activity of the conjugating system in rabbit reticulocytes. Ubiquitin conjugates are readily formed in an ATP-dependent manner with both endogenous reticulocyte proteins (Figure 7a, lane 1; Figure 7b, lanes 1 and 5) and the well-characterized exogenous protein, lysozyme (Figure 7a, lane 3; Figure 7b, lanes 3 and 7). Ladders of ubiquitin–lysozyme conjugates are seen between molecular masses 16.5 and 66 kDa (Figure 7a, lane 3), which may correspond to mono-, di-, tri- and tetra-ubiquitin–lysozyme species. This is more clearly seen in Figure 7(b). The ubiquitin analogues Val26Abu/Ile30Nva and, more surprisingly, Val26Ala/Ile30Ala, also gave ubiquitin–lysozyme ladders (Figure 7b, lanes 3 and 7). This indicates that both analogues are recognized similarly to native ubiquitin by

reticulocyte activating and conjugating enzymes, even though in the *bis*-Ala mutant, the structure suffers considerably more disruption.

Conclusions

We have analysed the effects of progressively shortening the hydrophobic side chains from the core of a small protein. To set an upper limit to the destabilization, the first experiment was the draconian removal of all (24 methylene groups) of a section of the core which completely destroyed both structure and biological activity [16]. The estimated destabilization is $24 \times 5 = 120$ kJ/mol, which is several times larger than the stability of native ubiquitin (about 40 kJ/mol), so that it is not surprising that little structure and biological activity are detectable, notwithstanding the presence of all of the hydrogen-bonding capacity of the native ubiquitin. This suggests that folding is initiated by hydrophobic interactions, which produce a molten globule to which subsequent hydrogen-bonding adds further stability. Second, the removal of five methylene groups in [Ala^{26},Ala^{30}]Ub significantly reduced the solubility such that an NMR analysis of the solution structure was impossible [22]. This structure has a destabilization of around 30 kJ/mol, which is close to that of native ubiquitin. The protein is still biologically active, however, and we conclude that it can form a near-native structure. The least disruptive changes (two $-CH_3$ groups; 14 kJ/mol) involved the atypical amino acids Nva and Abu and gave a soluble product that was less stable than native ubiquitin but retained full biological activity. Finally, the destabilization resulting from the loss of two $-CH_3$ groups appears to be independent of their exact position within the core [33], i.e. the length and volume of the side chain must maintain side-chain interactions between the α-helix and β-sheet structures. In this case, the greater flexibility of the linear side chains of Abu and Nva does not lead to any greater destabilization than that for the natural amino acids.

We have shown using chemical synthesis that it is possible to construct a crystalline native-like protein containing amino acid residues of a type not specified by the genetic code and that the structure and biological activity of this particular protein analogue is little altered from that of the native protein.

We thank Professor Neil Isaacs and Dr. Andy Freer for help with data collection and Dr. Linda Gilmore and WELMET with protein sequencing. We are grateful to the EPSRC and BBSRC, Merck Sharp and Dohme and Applied Biosystems Inc. for financial support. R.J.M. and R.R. thank the MRC for support of some of this study. We are also pleased to acknowledge the help of J. Wilken and B. Whigham.

Received 2 August 1996/12 December 1996; accepted 17 December 1996

REFERENCES

1 Milton, R. C., Milton, S. C. F. and Kent, S. B. H. (1992) Science **256**, 1445–1448
2 Dawson, P. E., Muir, T. W., Clark-Lewis, I. and Kent, S. B. H. (1994) Science **266**, 776–779
3 Liu, C. F. and Tam, J. P. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 6584–6588
4 Muir, T. W., Williams, M. J., Ginsberg, M. H. and Kent, S. B. H. (1994) Biochemistry **33**, 7701–7708
5 Padmanabhan, S., Marqusee, S., Ridgeway, T. and Laue, T. M. (1990) Nature (London) **344**, 268–270
6 Chou, P. Y. and Fasman, G. D. (1978) Annu. Rev. Biochem. **47**, 251–276
7 Thornton, J. M., Jones, D. T., MacArthur, M. W., Orengo, C. M. and Swindells, M. B. (1995) Philos. Trans. R. Soc. London Ser. B **348**, 71–79
8 O'Neil, K. T. and DeGrado, W. F. (1990) Science **250**, 646–651
9 Hurley, J. H., Baase, W. A. and Matthews, B. W. (1992) J. Mol. Biol. **224**, 1143–1159
10 Padmanabhan, S. and Baldwin, R. L. (1991) J. Mol. Biol. **219**, 135–137
11 Weber, P. L., Brown, S. C. and Mueller, L. (1987) Biochemistry **26**, 7282–7290
12 Di Stefano, D. L. and Wand, A. J. (1987) Biochemistry **26**, 7272–7281
13 Vijay-Kumar, S., Bugg, C. E. and Cook, W. J. (1987) J. Mol. Biol. **194**, 531–544
14 Alexeev, D., Bury, S. M., Turner, M. A., Ogunjobi, O. M., Muir, T. W., Ramage, R. and Sawyer, L. (1994) Biochem J. **299**, 159–163
15 Kraulis, P. (1991) J. Appl. Crystallogr. **24**, 946–950
16 Muir, T. W. (1992) Ph.D. Thesis, University of Edinburgh
17 Briggs, M. S. and Roder, H. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 2017–2021
18 Ramage, R., Green, J., Muir, T. W., Ogunjobi, O. M., Love, S. and Shaw, K. (1994) Biochem J. **299**, 151–158
19 Cornish, V. W., Mendel, D. and Schultz, P. G. (1995) Angew. Chem. Int. Ed. Engl. **34**, 621–633
20 Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. and Kent, S. B. H. (1989) Science **245**, 616–621
21 Brown, A. R., Irving, S. L. and Ramage, R. (1993) Tetrahedron Lett. **34**, 7129–7132
22 Wilken, J. (1995) Ph.D. Thesis, University of Edinburgh
23 Moore, A. T., Williams, K. E. and Lloyd, J. B. (1977) Biochem. J. **164**, 607–616
24 Hershko, A., Heller, H., Elias, S. and Ciehanover, A. (1983) J. Biol. Chem. **258**, 8206–8214
25 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
26 Howard, A. J., Gilliland, G. L., Finzel, B. F., Poulos, T. L., Ohlendorf, D. D. and Salemme, F. R. (1987) J. Appl. Crystallogr. **20**, 383–387
27 Collaborative Computing Project No. 4 (1994) Acta Crystallogr. **D50**, 760–763
28 Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Acta Crystallogr. **A47**, 110–119
29 Tronrud, D. E., Ten Eyck, L. F. and Mathews, B. W. (1987) Acta Crystallogr. **A43**, 489–501
30 Brunger, A. T., Krukowski, A. and Erickson, J. W. (1990) Acta Crystallogr. **A46**, 585–593
31 Provencher, S. W. and Glöckner, J. (1981) Biochemistry **20**, 33–37
32 Khorasanizadeh, S., Peter, I. D., Butt, T. R. and Roder, H. (1993) Biochemistry **32**, 7054–7063
33 Khorasanizadeh, S., Peters, I. D. and Roder, H. (1996) Nature Struct. Biol. **3**, 193–205
34 Steif, C., Hinz, H.-J. and Cesarini, G. (1995) Protein Struct. Funct. Genet. **23**, 83–96
35 Shortle, D., Stites, W. E. and Meeker, A. K. (1990) Biochemistry **29**, 8033–8041