Synthetic, structural and biological studies of the ubiquitin system: chemically synthesized and native ubiquitin fold into identical three-dimensional structures

Dimitriy ALEXEEV,* Stella M. BURY,* Mary A. TURNER,* Oluyinka M. OGUNJOBI,† Thomas W. MUIR,† Robert RAMAGE† and Lindsay SAWYER*‡

The Edinburgh Centre for Molecular Recognition *Department of Biochemistry, University of Edinburgh, Hugh Robson Building, George Square, Edinburgh EH8 9XD, Scotland, U.K., The Edinburgh Centre for Molecular Recognition †Department of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, Scotland, U.K.

The solid-phase chemical synthesis of ubiquitin produced a molecule with physicochemical properties similar to those of the natural protein. We have grown crystals of this synthetic ubiquitin and performed an X-ray analysis at 1.8 Å resolution in order to compare the synthetic protein with the known natural structure. The crystals were isomorphous with those of the natural protein, the *R*-factor between them being 7.1%. Difference Fourier analysis shows that the synthetic and natural structures are indistinguishable. The co-ordinates of the natural ubiquitin (1UBQ) were used as the starting point for restrained least-squares refinement (TNT program) against the synthetic X-ray data. The refinement converged to $R = 16.5\%$ and the resulting model did not change when refined against natural ubiquitin X-ray data ($R = 18.7\%$). From both the refinement and featureless difference Fourier synthesis, we conclude that the synthetic and natural protein structures are identical. A short discussion about the uses of proteins with 'non-standard' amino acid residues is included.

INTRODUCTION

Protein biosynthesis requires a sophisticated mechanism to ensure that the amino acid sequence of the final product is correct and consequently properly folded and active. This process may or may not require other proteins such as disulphide isomerase [1] or a chaperonin molecule [2], let alone any of the post-translational modifications that are necessary for full functionality. If an artificial protein or protein fragment is prepared chemically or, indeed, by recombinant methods, it cannot be predicted whether the product will be structurally and functionally identical with the original protein or fragment thereof.

Ubiquitin consists of 76 amino acid residues forming a single polypeptide chain. It is probably the most highly conserved protein known, as the primary structure is almost identical over a wide range of evolutionarily distant species [3]. The spatial structure (Figure 1) represents a compact hydrophobic core combined with a full network of hydrogen bonds created from five strands of β-sheet and a single four-turn helix [4]. The protein is rather insensitive to proteolytic digestion and is stable over a wide range of pH and temperature. Several functions have been ascribed to the protein, but the most established idea is that ubiquitin serves as a flag for ATP-dependent degradation of proteins in the cell [6,7]. This protein digestion requires that ubiquitin be linked to the target protein and it is this covalent conjugate that is recognized and subsequently digested. Both natural and synthetic ubiquitin have been compared with respect to conjugation to lysozyme and were found to be identical (R. J. Mayer, personal communication). It is one of the ultimate targets of this programme to establish, by means of tertiary-structure determination, the nature of the recognition and conjugation processes.

Solid-phase chemical synthesis of ubiquitin has been carried out, proceeding from the C-terminus using Fmoc methodology, and the integrity of the primary sequence has been shown to be accurate by application of a wide selection of chemical, physicochemical and enzymic methods. The physicochemical and biological properties of synthetic ubiquitin have been shown to be identical with those of the natural protein [8]. The next phase of the investigation required establishment of the tertiary structure of the synthetic protein in solution and the solid state. Thus we have grown crystals of this synthetic ubiquitin and performed an X-ray analysis at 1.8 Å resolution in order to compare the structure of the synthetic protein with that established for the native structure [4].

MATERIALS AND METHODS

Synthesis and purification

The chemical synthesis and purification of ubiquitin is described in detail in the preceding paper [9]. It was essential to prove that the protein isolated was ubiquitin, and therefore an extensive study was carried out on the homogeneity of the product followed by establishment of the primary sequence using electrospray m.s., Edman degradation, total acid and enzymic hydrolyses and selective enzymic fragmentation. The next phase in the programme required that a comparison be made of the solid-state (crystal) structure of samples of natural and synthetic protein.

Crystallization and data collection

Purified synthetic ubiquitin was used to grow crystals by repeated seeding, as by this technique only can crystals be reproducibly obtained [4]. Indeed, no crystals have ever been obtained in our

Abbreviations used: PEG, poly(ethylene glycol); r.m.s., root mean square.

‡ To whom correspondence should be addressed.

The atomic co-ordinates for the synthetic ubiquitin structure have been deposited with the Protein Data Bank at Brookhaven with the accession number 2UBQ.

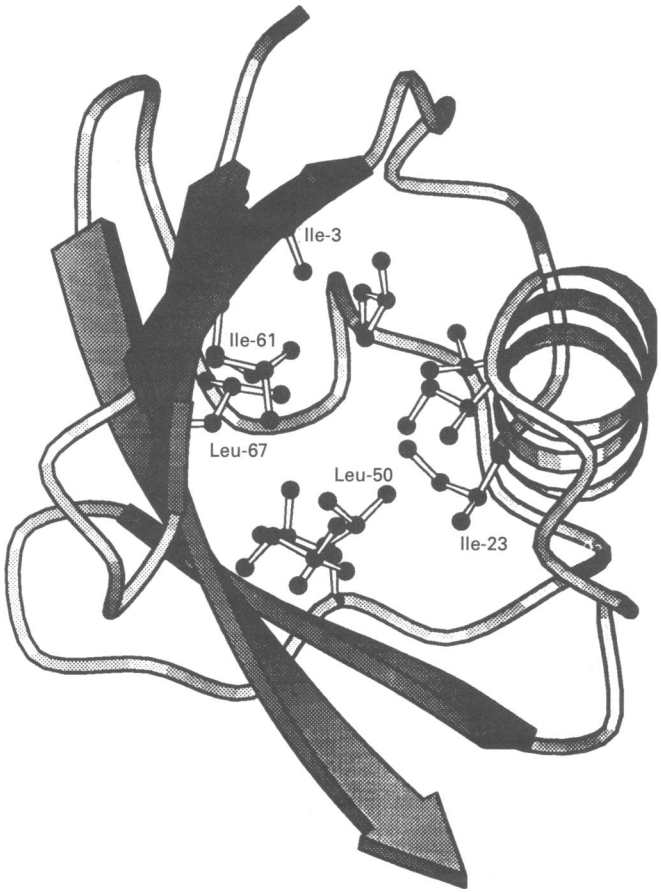

Figure 1 A cartoon of the main-chain fold of ubiquitin, drawn by MOLSCRIPT [5]

The β-strands are shown as arrows and the α-helix as a coil. The side chains of the hydrophobic core between the helix and sheet referred to in the text are also shown.

laboratory without seeding [10]. Several large seed crystals were kindly supplied by Dr. W. J. Cook (University of Alabama Medical School). One such seed crystal was carefully added to a single 10 μl drop containing 5 μl of concentrated synthetic ubiquitin solution (20 mg/ml) and 5 μl from the well solution, containing 30% (w/v) poly(ethylene glycol) (PEG) 4000 in 50 mM cacodylate/HCl (pH 5.6). Small single crystals formed independently of the seed crystal in the original drop and were transferred to fresh drops of the same composition for further growth. These crystals were then removed from the second drop and put into fresh ones containing PEG 4000 at a concentration of 25% (w/v). This PEG concentration was gradually increased to 38% in order to stabilize the crystals.

X-ray precession photographs confirmed that the synthetic ubiquitin crystallized in the same space group, $P2_12_12_1$, as natural protein and with identical parameters $a = 50.83$, $b = 42.74$, $c = 28.98$ Å. A data set to a maximum resolution of 1.6 Å was collected from one crystal using a Siemens–Nicholet–Xentronics area detector on a Rigaku (RU200) rotating anode generator operating at 40 kV and 100 mA and fitted with a graphite monochromator to select Cu-Kα radiation. There were 14404 observations representing 6322 unique reflections to 1.6 Å resolution. The data were reduced using the XENGEN v.1.3 program to give an unweighted absolute value for R_{merge} on intensity of 5.3%. Because of the poor quality of the weak reflections at high resolution, the data were restricted to 1.8 Å resolution, thus providing 5058 out of a possible 6050 reflections, which represents 85% of the expected diffraction pattern. After a $1\sigma(I)$ cut-off was applied, 4349 reflections in the range 6.0–1.8 Å were left for model refinement. In comparison, the final refinement of the original model of natural ubiquitin [4] was based on 5554 reflections in the same range with a similar $\sigma(I)$ cut-off, indicating that the weak reflections in that study were measured more accurately.

In order to minimize the possibility that significant differences in structure were the result of different ambient conditions between Edinburgh and Alabama, we collected data from a crystal of natural ubiquitin to a similar resolution using a Siemens–Stoe AED-2 four-circle diffractometer and graphite-monochromated Cu-Kα radiation from a sealed beam tube operating at 50 kV and 35 mA. Some 4295 reflections with $I > 2.\sigma(I)$ were collected as background–peak–background by ω-scan, with an R_{merge} of 4.2% for the data in the same resolution range as above. This is data set Nat1. In a separate experiment, data were also collected on a natural ubiquitin crystal on the area detector under the same conditions as described above. There were 17462 reflections collected which produced 6510 unique reflections to 1.8 Å resolution with an $R_{merge} = 4.7\%$. However, the outer shell between 1.9 and 1.8 Å was only sparsely populated, and the data set was restricted to 1.9 Å resolution, comprising 4449 reflections with $I > 2\sigma(I)$, which represents 84.2% of the possible reflections. This is data set Nat2.

Data set comparison

Comparison of the synthetic and natural ubiquitin structure factors with those calculated from the published model for natural ubiquitin and 58 associated water molecules (1UBQ in the Brookhaven Databank [11,12]) revealed that data collected for the synthetic protein (Syn) appeared to conform to the model (Mod) somewhat better than the data collected from our native ubiquitin crystals: $R_{Syn-Mod} = 18.3\%$, $R_{Nat1-Mod} = 19.8\%$ and $R_{Nat2-Mod} = 20.5\%$. After model temperature factor refinement, the R-factor fell to 17.1% for the synthetic data set, which is close to the final R-factor published for the original model of 17.6% [4].

To establish whether or not synthetic ubiquitin was identical with the native protein, we compared experimental data sets directly rather than each of them to the model structure factors. Although $R_{syn-Nat1}$ was found to be 10.4%, a value somewhat higher than expected, $R_{Nat2-Nat1}$ was found to be 11.1%. These values, however, reflect the comparison of a data set collected serially (Nat1) with those collected on an area detector. The identity of the synthetic and natural data sets is shown by $R_{Syn-Nat2}$, which was 7.1%, a value typical of data sets collected on two crystals from the same tube.

To estimate the extent of any difference in structure, we have compared $(2|F_{Syn}|-|F_{Nat1}|)$ and $(|F_{Syn}|-|F_{Nat1}|)$ Fourier syntheses, using phases derived from the published natural ubiquitin model, before refinement with the data collected in our study. The maximum peak in the difference Fourier map was found to be five times lower than root mean square (r.m.s.) deviation of the $(2|F_{Syn}|-|F_{Nat1}|)$ synthesis showing that the synthetic and natural protein structures are indistinguishable at this resolution.

Model refinement

To provide a final check that the structures were identical, we used the data set for the synthetic protein, collected as described above and the molecular model of native ubiquitin taken from

Table 1 Mean and r.m.s. deviations of the refined model parameters from the standard values used in he refinement

Max. is the maximum observed deviation from ideal. The σ value is that used as the target restraint for each class.

	Mean	R.m.s. deviation	Max.	σ
Bond length (Å)	0.009	0.011	0.040	0.020
Bond angles (°)	0.052	0.990	5.531	3.000
Torsion angles (°)	5.664	24.23	51.77	15.00
Chiral volume (Å^3)	0.012	0.014	0.027	0.020
Planar groups (Å)	0.013	0.016	0.040	0.020
Bad contacts (Å)	0.358	0.627	0.215	0.100

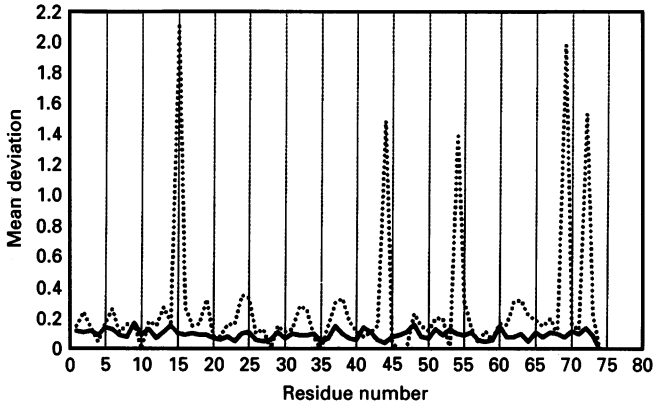

Figure 2 Deviations of the residue-averaged main-chain (solid line) and side-chain (dashed line) atoms between the structures of natural ubiquitin determined previously [4] and the synthetic protein used in this work

the Brookhaven Data Bank as the starting point for restrained refinement. Careful inspection of the $(2|F_{Syn}|-|F_{Nat1}|)$ density map revealed that the conformation of some of the side chains and the positions of some of the water molecules could be improved. The model was therefore refined using Release 4 of the TNT restrained least-squares program [13] against the synthetic X-ray data set. A set of 4349 reflections was used for TNT refinement, and the set of 5031 reflections including poorly measured weak reflections was used for electron-density calculation in the resolution range 6.0–1.8 Å. After every round of refinement (ten cycles), the $(2|F_o|-|F_c|)$ and difference Fourier maps were inspected and adjusted using program FRODO [14] running on an ESV 10/20 molecular graphics workstation. Water molecules were added during this process. Because the last residues at the C-terminus are very flexible and therefore poorly defined, both in our and the original electron-density maps, throughout the refinement their positions and occupancies were fixed as in the original model. The positions of the two C-terminal residues previously determined were based on the difference map between the whole ubiquitin molecule (76 residues) and that with the last two residues removed.

Each round of refinement consisted of five cycles of positional refinement for the non-hydrogen atoms in the protein and water molecules, followed by five subsequent cycles of temperature factor and occupancy refinement. All positive peaks in the $(2|F_o|-|F_c|)$ and $(|F_o|-|F_c|)$ Fourier maps higher than the $1.\sigma$ level that were not already covering model atoms were examined. Water molecules were added to the model only if they occupied stereochemically reasonable positions. After six rounds of refinement, the model had converged to $R = 16.5\%$ and remained close to the starting structure with the r.m.s. deviations from the starting model being 0.110 and 0.535 Å respectively for the main- and side-chain atoms. Table 1 summarizes the results of the refinement.

To validate the proposed minor modifications to the published co-ordinates, the model produced as described above was refined against the native data set, Nat2. Three cycles of refinement in TNT were performed to obtain an accurate scale factor at which point $R = 18.5\%$, and the r.m.s. deviations between the model for synthetic ubiquitin and that obtained by refinement of this model against the natural data set were 0.076 and 0.091 Å for the main- and side-chain atoms respectively.

RESULTS

The data analysis, Fourier syntheses and refinement studies all indicate that the structures of the synthetic and natural ubiquitin molecules are identical. For the synthetic structure, refined against the data collected in this study of the synthetic protein, the mean deviation from the ideal bond lengths is less than 0.005 Å, r.m.s. = 0.01 Å, the maximum deviation being 0.04 Å or 2.0σ. Bond-angle deviations are 0.06°, r.m.s. = 1.03° and maximum deviation is 5.8°, that is 1.9σ. There are no intramolecular contacts shorter than 2.5 Å. There is one short intermolecular contact between the C-terminal carboxy group which arises because the terminal conformation was fixed as in the natural structure refinement.

The hydrogen-bond network and dihedral angles of the main chain are very similar to those of the starting model. Figure 2 shows the mean displacement of the main-chain and side-chain atoms of the refined model relative to the starting model, i.e. the natural structure. Only a few side chains moved or were repositioned significantly during refinement. Large movements shown in Figure 2 with mean side-chain displacements of more than 1 Å correspond to the rotation of the branched residues (Leu-15, Ile-44, Arg-54, Leu-69, Arg-72). Figure 3 depicts some typical electron-density maps from which it can be seen that such a side-chain rotation affects the electron density only slightly. Figure 4 shows that the temperature factors for these same side chains are high. The location of a water molecule was deemed to be genuine if its occupancy remained above 0.3, its temperature factor remained below 35 Å^2 and it made at least one sensible hydrogen bond. By this procedure, 82 water molecules were located in the course of the refinement compared with the 58 molecules of the starting model.

DISCUSSION

From our analysis of the natural and synthetic crystal structures, using the data sets collected in this study, we conclude that the two structures are indistinguishable. Examination of the difference Fourier syntheses of the X-ray data collected for the synthetic and natural ubiquitin establishes that the structures are identical. This conclusion is supported not only by the isomorphous nature of the crystals which allows 'epitaxial' growth but also by the result of careful analyses involving c.d., n.m.r., m.s. and fluorescence studies [8]. The differences that do exist concern minor adjustments of the published model for a few side-chain conformations and for the protein hydration shell. Our refinements using both synthetic and native data sets collected under the same conditions (Syn and Nat2) show that the changes that we observe are consistent in both data sets.

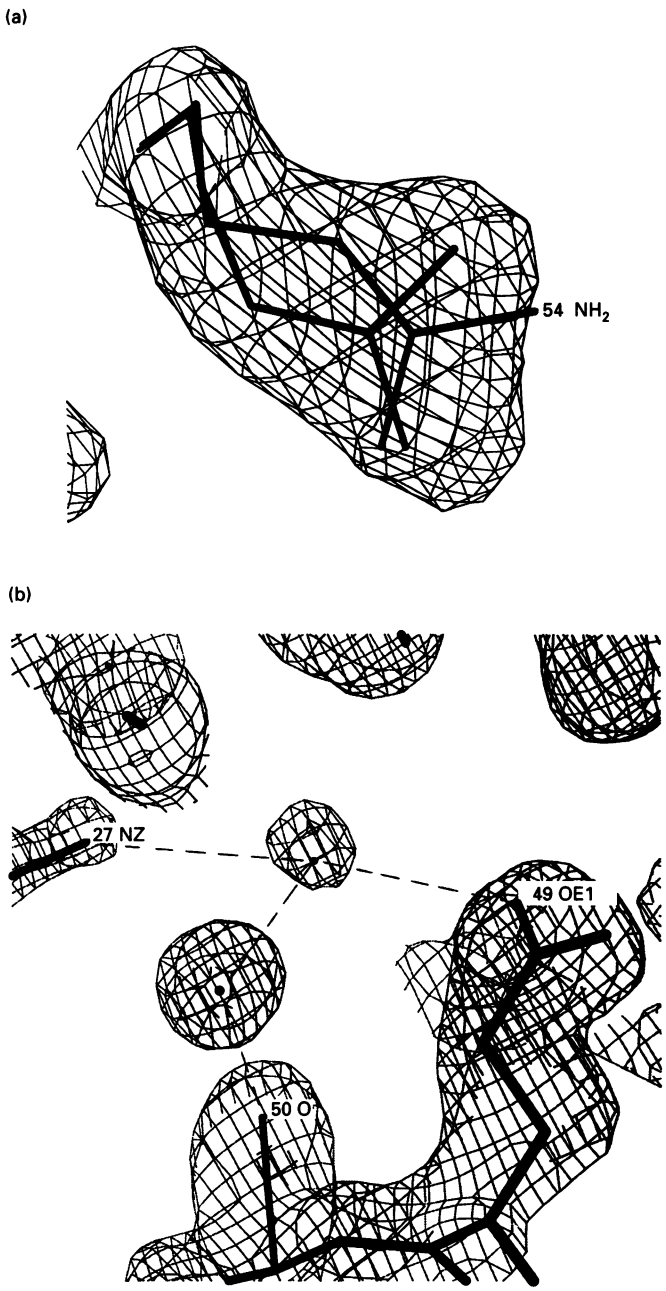

Figure 3 (a) Diagram showing two alternative interpretations of the side chain of Arg-54 and its electron density calculated from the data for the synthetic protein and (b) two well-defined water molecules hydrogen-bonded to main- and side-chain atoms in the crystal

(**a**) The original interpretation [4] is the one with the terminal nitrogen atom (NH_2) projecting out of the electron density. It is highly likely that both conformations exist *in vivo*. (**b**) NZ, OE1 and O refer to N_ζ, Oϵ1 and O atoms of residues 27, 49 and 50 respectively.

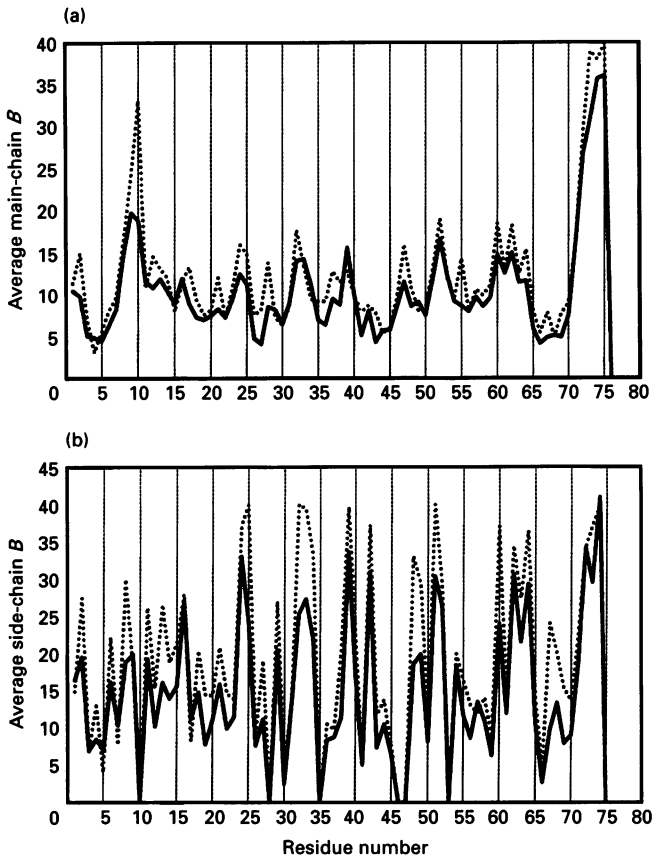

Figure 4 Residue-averaged temperature factors, *B* (Å^2), for main-chain (a) and side-chain (b) atoms for the refined model of synthetic ubiquitin (dashed line) and the natural protein (solid line).

We found that five amino acids had side chains with significantly changed conformation during the refinement, and are therefore the major contributors to the r.m.s. deviation obtained (0.535 Å). These are Leu-15, Ile-44, Arg-54, Leu-69 and Arg-72. Each of these side chains is flexible, and it is entirely reasonable that under the conditions of our experiment they can adopt essentially isoenergetic conformations. Indeed, as can be seen in Figure 3(a), the alternative conformation makes an acceptable fit to the final $2|F_o|-|F_c|$ Fourier map, suggesting that the surface side chains occupy both possible conformations *in vivo*. The conformation of Arg-54 in our refined model seems preferable because a short intermolecular contact involving this residue is removed. We stress, however, that our interpretation of these flexible side-chain conformations is certainly subjective and thus may or may not represent a real, albeit unimportant, difference between the structures determined in Edinburgh and Alabama. In effect, the work described in this study amounts to a re-refinement of the native structure.

The process of allocating solvent molecules is, to some extent, a matter of deciding when to stop. We have found 50 of the 58 original water molecules of the starting model in our electron-density maps and were able to locate 34 additional ones. All of these additional water molecules found in the synthetic-structure refinement remained with occupancies above 0.6 when the model was refined into the data set, Nat2. Most water molecules were added to the bulk solvent between the neighbouring protein molecules in the unit cell. As such, it is conceivable that several of them with low occupancies and high temperature factors could be random errors in the map. However, the peaks all returned, if omitted, in the same positions in difference Fourier maps, they do not disappear when refined against the natural data (Nat2) and they all satisfy the criteria described in the Materials and methods section. Water molecules form hydrogen-bonded chains which are linked to the protein with a few hydrogen bonds. In contrast with the first hydration shell, where

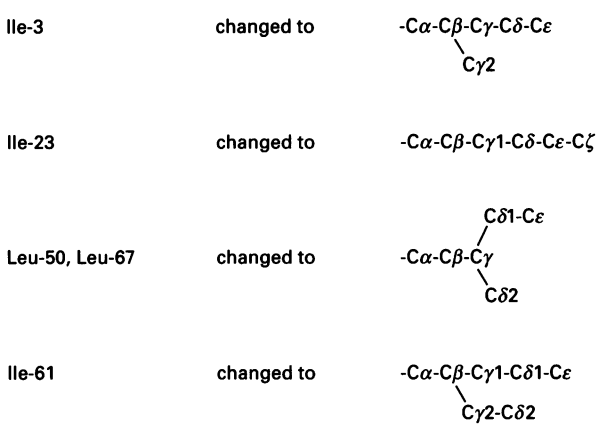

Figure 5 Short series of synthetic α-amino acids with side chains that can occupy the same internal space as those of residues 3, 23, 50, 61 and 67

the electron density is compact and spherical (Figure 3b), the temperature factors of these outer water molecules are around 40 Å^2, and electron-density peaks are anisotropic suggesting that the water molecules are quite mobile.

Ubiquitin has a tightly packed hydrophobic core, but careful inspection reveals that there are some empty holes inside it. Protein-engineering studies of other proteins, most notably T4 lysozyme [15], subtilisin [16] and dihydrofolate reductase [17], show that increased stability can be achieved if such gaps are filled. Further, it has been shown that it is also possible to replace a bulky hydrophobic group in the core with a smaller one and leave the whole protein intact, thereby creating a pocket which may be large enough to accommodate a solvent molecule. Such a gap could form a perfect trap for a ligand conforming to the shape of the absent hydrophobic group(s), so that, when the protein is refolded after deprotection, the ligand forms a nucleus around which the polypeptide folds, effectively binding it.

Ubiquitin with seven internal residues replaced by Gly or Ala has been synthesized and shown to be not capable of folding into its native conformation [8].

Conversely, we are investigating the possibility of increasing the stability of ubiquitin by filling the existing gaps present in the core. Synthetic proteins need not be confined to the natural amino acids and, as a consequence, it is possible to conserve the overall protein structure and introduce a small number of unusual amino acids to occupy the available extra space. Such arguments are equally applicable to modifications of an enzyme active site. The effects of these atypical amino acid substitutions need to be carefully evaluated, and, in doing so, an extra dimension is introduced to the problem of protein folding.

In the case of ubiquitin, it is possible to extend specific side chains to match some of the gaps found in the core. An examination of the core of ubiquitin (Figure 1) showed that modifications of Ile-3, Ile-23, Leu-50, Ile-61 and Leu-67 could lead to tighter packing. A model was produced and optimized using TNT with no X-ray term to relieve any close interatomic clashes, with the positions of all protein atoms fixed except those of the five amino acids cited above. TNT used in this manner amounts to an energy minimization of the groups within the hydrophobic core. The final conformation does not have interatomic contacts shorter then 3.3 Å and fills the overall core in a stereochemically reasonable manner. Thus there should be no increase in the energy of this structure and there may well be a small decrease. The modifications proposed are presented in Figure 5 and are at present being synthesized to that we can (a) confirm the hypothesis about the stability and (b) compare the folding kinetics with those of the native protein [18].

We thank Dr. W. J. Cook for kindly supplying the natural ubiquitin crystals, Professor Neil Isaacs and Dr. Andy Freer for help with data collection on the synthetic protein and Dr. Sandy Blake, Dr. Paul Taylor and Dr. Alan McAlpine for help with data collection on the natural protein. We are grateful to the Science and Engineering Research Council, Merck Sharp and Dohme and Applied Biosystems for financial support.

REFERENCES

1 Freedman, R. B., Baneid, N. J., Hawkins, H. C. and Paver, J. L. (1989) Biochem. Soc. Symp. **55**, 167–192
2 Ellis, R. J. and van der Vries, S. M. (1991) Annu. Rev. Biochem. **60**, 321–347
3 Rechsteiner, M. (1988) Ubiquitin, Plenum Publishing Corp., New York
4 Vijay-Kumar, S., Bugg, C. E. and Cook, W. J. (1987) J. Mol. Biol. **194**, 531–544
5 Kraulis, P. J. (1991) J. Appl. Crystallogr. **24**, 531–544
6 Finley, D., Bartel, B. and Varshavsky, A. (1989) Nature (London) **338**, 394–401
7 Hershko, A. and Ciechanover, A. (1992) Annu. Rev. Biochem. **61**, 761–807
8 Muir, T. W. (1992) Ph.D. Thesis, University of Edinburgh
9 Ramage, R., Green, J., Muir, T. W., Ogunjobi, O. M., Love, S. and Shaw, K. (1994) Biochem. J. **299**, 151–158
10 Jancarik, J. and Kim, S.-H. (1991) J. Appl. Crystallogr. **24**, 409–411
11 Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M (1977) J. Mol. Biol. **112**, 535–542
12 CCP4 Program Package for Protein Crystallography, SERC, Daresbury Laboratory, U.K.
13 Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987) Acta Crystallogr. **A43**, 489–501
14 Jones, T. A. (1978) J. Appl. Crystallogr. **11**, 268–272
15 Hurley, J. H., Baase, W. A. and Matthews, B. W. (1992) J. Mol. Biol. **224**, 1143–1159
16 Cunningham, B. C. and Wells, J. A. (1987) Protein Eng. **1**, 319–325
17 Thillet, J., Absil, J., Stone, S. R. and Pictet, R. (1988) J. Biol. Chem. **263**, 12500–12508
18 Briggs, M. S. and Roder, H. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 2017–2021

Received 9 July 1993/9 November 1993; accepted 15 November 1993