Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin

Robert RAMAGE,* Jeremy GREEN, Thomas W. MUIR, Oluyinka M. OGUNJOBI, Steven LOVE and Kevin SHAW

The Edinburgh Centre for Molecular Recognition, Department of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, Scotland, U.K.

The small protein ubiquitin (76 amino acids) has been synthesized under optimized conditions by Merrifield solid-phase methodology using the *N*^α-Fmoc protecting group. The crude polypeptide mixture was purified to homogeneity by gel filtration, dialysis and a combination of cation- and anion-exchange chromatography to yield ubiquitin. Amino acid analysis, enzymic

digestion and sequencing by automated Edman degradation were used to authenticate the primary structure. Isoelectric focusing and m.s. were used to demonstrate that the final product was greater than 98 % pure with a final yield of 93 mg (4.3 %) from a single synthesis on a 0.25 nmol scale.

INTRODUCTION

The tremendous impact on biology resulting from developments in gene cloning and expression has allowed entry into a myriad of scientific problems, both academic and industrial. Despite the power of such methods, chemical synthesis can also make an important contribution to the understanding of protein folding and recognition, and to the consequences of such recognition. Incorporation of unnatural or specifically isotopically labelled α -amino acids into a protein sequence through chemical synthesis should give results complementary to those obtained by genetic manipulation. At present, many gene-based studies lead to small quantities of protein, although this can be enhanced by over-expression to give proteins containing α -amino acids for which there is tRNA. The recent use [1,2] of redundant stop codons in RNA for loading of unnatural α -amino acids offers exciting opportunities but, at present, it has some limitations in the scale of protein produced.

Thus the chemical synthesis of small proteins provides an important approach to the elucidation of aspects of their tertiary structure and function, especially with the capability of introducing unnatural and isotopically labelled α -amino acids into the primary sequence. It was decided to test the methodology initially by synthesizing a small well-characterized protein of real biological significance for which X-ray crystal data are available; ubiquitin fits these criteria admirably. We plan to build upon experiences gained in this programme to define strategies for the chemical synthesis of larger proteins (about 20 kDa) which would necessarily incorporate an element of convergence.

Ubiquitin is a highly conserved globular protein (76 amino acids, 8565 Da) (Figure 1) found in all eukaryotes [3] both as the free protein and covalently attached to many other proteins. Many diverse functions have been discovered for it since its discovery in 1975, and these all involve it being attached to other often larger proteins [4]. Probably the best understood of these is its involvement in protein degradation [5], in which ubiquitin becomes covalently linked to other proteins via a multienzyme conjugation system, usually resulting in the target protein being tagged for proteolytic digestion [6,7]. It is also likely that the

ubiquitin-degradation pathway is involved in the homeostatic response that cells deploy during times of stress, such as excessive heat [8].

Histones H2A and H2B, which are nuclear proteins involved in chromatin structure, can be both mono- and poly-ubiquitinated [9,10], and this reversible process may play a vital role in the cell cycle [4,11]. Ubiquitin is known to be a component of two membrane-bound receptors, namely lymphocyte homing receptor [12] and platelet-derived growth factor [13]. Exactly why these receptors are ubiquitinated is not yet clear, but an involvement in signal transduction has been proposed. Ubiquitin has been implicated in Alzheimer's disease, as the major constituent of the diagnostic neurofibrillary tangles is ubiquitinated [14]. More recent interest has surrounded novel C-terminal extension proteins. The identification of the genes that encode ubiquitin fused to either a 52-amino acid [15] or an 80-amino acid [16] extension protein was quickly followed by the discovery that the extension segments are located within ribosomes [17,18]. After translation, ubiquitin is quickly clipped off [19], suggesting that it may play a role in translational regulation. *Escherichia coli* has been shown to be capable of expressing an unnatural protein, comprising ubiquitin at its N-terminus and poly(ADP-ribose) at its C-terminus, in ten times greater quantity than it expresses the protein itself [20]. This again suggests that ubiquitin has a role in translational regulation.

We chose to synthesize ubiquitin chemically on a useful scale (100–200 mg) for a number of reasons. Firstly, ubiquitin, as a relatively small protein, is suitable for checking the refinements required of our synthetic work. Secondly, the tertiary structure of ubiquitin is well understood in both the crystal [21] and solution [22,23] phases, thus making it easier to authenticate the conformation of the synthetic material with that of the native protein. Thirdly, ubiquitin is a good model for the study of protein structure and folding because its tertiary structure contains all the secondary structural units commonly found within globular proteins whilst itself being small enough to facilitate a comprehensive study of folding intermediates. The ultimate objectives were to synthesize analogues of ubiquitin by the insertion of unnatural amino acids and specifically iso-

Abbreviations used: Boc, *t*-butoxycarbonyl; DCM, dichloromethane; DIC, di-isopropylcarbodi-imide; DMF, dimethylformamide; DTT, dithiothreitol; Fmoc, fluorenylmethoxycarbonyl; HOBt, hydroxybenzotriazole; IEF, isoelectric focusing; Mbh, 4,4-dimethoxybenzhydryl; OBU^t, *t*-butoxy; Pmc, 2,2,5,7,8-pentamethylchroman-6-sulphonyl; *t*-Bu, *t*-butyl; TFA, trifluoroacetic acid; Trt, trityl.

* To whom correspondence should be addressed.

1	Met	Gln	Ile	Phe	Val	Lys	Thr	Leu	Thr	Gly	Lys	Thr	Ile	Thr	Leu	15
	Glu	Val	Glu	Pro	Ser	Asp	Thr	Ile	Glu	Asn	Val	Lys	Ala	Lys	Ile	30
	Gln	Asp	Lys	Glu	Gly	Ile	Pro	Pro	Asp	Gln	Gln	Arg	Leu	Ile	Phe	45
	Ala	Gly	Lys	Gln	Leu	Glu	Asp	Gly	Arg	Thr	Leu	Ser	Asp	Tyr	Asn	60
	Ile	Gln	Lys	Glu	Ser	Thr	Leu	His	Leu	Val	Leu	Arg	Leu	Arg	Gly	75
	Gly															

Figure 1 Amino acid sequence of bovine ubiquitin

topically labelled amino acids which would allow investigation of the folding mechanisms of the protein and thus produce quantitative information about the energy changes taking place in such a process.

EXPERIMENTAL

General

N^α-Fmoc-amino acids and *p*-benzyloxyphenyl alcohol resin were purchased from Novabiochem (U.K.) Ltd, except for Fmoc-Arg-(Pmc) which was purchased from Raylo Chemicals. The following side-chain-protecting groups were used: Asp (OBu^t); Asn (Mbh); Gln (Mbh); His (Trt); Lys (Boc); Ser (Bu^t); Thr (Bu^t); and Tyr (Bu^t). Peptide-synthesis-grade dimethylformamide (DMF), piperidine and dioxan were purchased from Rathburn Chemicals Ltd. Di-isopropylcarbodi-imide (DIC) was purchased from Aldrich. 1-Hydroxybenzotriazole (HOBt) was purchased from Fluka and recrystallized from ethanol, dried over MgSO₄ and stored in a desiccator until required. Acetic anhydride and pyridine were distilled and stored over 4A molecular sieves. Trifluoroacetic acid (TFA) was purchased from Applied Biosystems. Scavengers used were of the best quality available, thioanisole (99%), anisole (99%), ethylmethylsulphide (99%) and urea (99%) were all purchased from Aldrich. 2β-Mercaptoethanol (99%), ammonium acetate and NH₄HCO₃ were purchased from Biochemika Fluka. Dialysis membranes were from Spectrum (Pierce and Warriener). Tos-Phe-CH₂Cl ('TPCK')-treated trypsin type XIII, pepsin, microsomal leucine aminopeptidase and prolidase were all purchased from Sigma. Amino acid analysis was carried out on an LKB 4151 Alpha Plus amino acid analyser equipped with an LKB 2220 Recording Integrator. U.v. spectra were recorded on a Varian Cary 210 spectrophotometer. M.s. was carried out using the electrospray method on a VG-Fisons BIO-Q instrument courtesy of VG-Biotech.

Chromatography

Gel chromatography and ion-exchange chromatography were carried out using a Microperpex 2132 peristaltic pump and two LKB Uvicord 2138s spectrophotometers at 226 and 277 nm. The fraction collector used was an LKB Redirac 2112. H.p.l.c. was performed using an Applied Biosystems 151A h.p.l.c. system. For analytical h.p.l.c. and preparative chromatography, Brownlee C8 columns of 100 mm × 4.6 mm internal diameter and 100 mm × 10 mm internal diameter respectively were used.

Fmoc-Gly-resin

Fmoc-Gly-OH (0.94 g, 3.16 mmol) was suspended in dry dichloromethane (DCM) (50 ml), and thionyl chloride (2.20 ml,

3.16 mmol) was added. The mixture was refluxed under nitrogen at 50 °C for 2 h when the solvent and the excess thionyl chloride was removed under vacuum. The resulting solid was then redissolved in DCM and concentrated *in vacuo*. This was repeated to give a pale yellow solid: IR (DCM) COCl 1800 cm⁻¹ Fmoc C = O 1710–1720 cm⁻¹. *p*-Alkoxybenzyl alcohol polystyrene resin (Wang resin) (2 g, 1.58 mmol substitution) was preswollen in dry DCM (25 ml) and pyridine (3 ml) for 30 min. Fmoc-Gly-Cl (from above) was dissolved in DCM (10 ml), added to the resin mixture and then sonicated for 1 h. The resin was filtered, washed with DCM (3 × 50 ml) and diethyl ether (3 × 50 ml) and dried in a desiccator overnight. The loading efficiency was determined by sonicating a sample of the resin (3–4 mg) for 10–20 min in the presence of a 20% (v/v) piperidine/DMF solution (10 ml). From the resulting u.v. absorbance of this solution at 302 nm, the coupling efficiency was calculated to be 41%, which is equivalent to 0.29 mmol of glycine/g of resin.

A sample of the Fmoc-Gly-resin was treated with *p*-bromobenzoyl chloride in order to corroborate the substitution value obtained by the u.v. method using bromine analysis. Fmoc-Gly-resin (200 mg) was suspended in anhydrous DCM (10 ml) at 0 °C with gentle stirring. Pyridine (0.1 ml) was added followed by *p*-bromobenzoyl chloride (0.120 g), and the reaction mixture was left for 20 min. The resin sample was filtered, washed thoroughly with DCM followed by several washes with diethyl ether. The sample was dried overnight at 45 °C. Bromine analysis was carried out on three samples treated under identical conditions, these being *p*-benzyloxybenzyl alcohol resin, Fmoc-Gly-resin and *p*-bromobenzoylated sample in duplicate. The results of the bromine analysis agreed with the level of loading of Fmoc-Gly obtained by the u.v. method above.

Capping of Fmoc-Gly-resin

Fmoc-Gly-resin (1.60 g) was suspended in DCM (30 ml) and cooled to 0 °C. Pyridine (0.5 ml) and benzoyl chloride (0.5 ml) were added and the mixture was stirred for 5 h. The resin was filtered, washed with DCM (3 × 50 ml) and diethyl ether (3 × 50 ml) and dried in a desiccator overnight.

Check of Fmoc-Gly-resin for Gly-Gly

A sample of the resin was treated with 20% piperidine in DMF (30 ml) for 20 min to remove the Fmoc group. Glycine was removed from the resin by treatment with aq. 95% TFA for 1 h. After removal of the TFA *in vacuo*, the residue was triturated with ether to give a white powder which was isolated by filtration and dissolved in citric acid buffer, pH 2.2, for amino acid analysis. There was no indication of any glycylglycine: retention time for glycine was 6.19 min and for glycylglycine 39.2 min.

Synthesis of ubiquitin

Synthesis of ubiquitin was carried out on a 0.25 mmol scale on an Applied Biosystems 430A peptide synthesizer equipped with an Applied Biosystems 757 absorbance detector linked to a Hewlett-Packard HP3396A integrator for on-line monitoring of the deprotection solution containing the piperidine-Fmoc adduct. All amino acids were triple-coupled as the symmetrical anhydride (2 mmol of amino acid) followed by two HOBt active-ester (1 mmol of amino acid) couplings using DIC in dioxan as the activating agent. Glycine was single-coupled using 2 mmol of symmetrical anhydride, and histidine was triple-coupled using the HOBt active-ester method. Each coupling step lasted 30 min with a capping step using acetic anhydride in DMF for 6 min on

the completion of the coupling step. This was followed by the removal of the Fmoc group with successive 5, 3, 3 and 1 min treatments with 20 % piperidine in DMF, with extensive washing with DMF between each deprotection. The synthesis was monitored using real-time assessment of the deprotection solution containing the chromophoric Fmoc-piperidine adduct at 302 nm in a continuous flow mode from the synthesizer to the u.v. monitor. This provided a continuous record of every deprotection step throughout the synthesis. The final amino acid was left with the Fmoc protecting group attached so that u.v. monitoring could be carried out manually as described above for the loading process.

Cleavage of the peptide from the resin

A sample of the peptidyl-resin (about 1.33 g) was swollen in DMF, then treated with 20 % piperidine in DMF and sonicated for 15 min. After being washed with DMF (3 × 50 ml) and DCM (3 × 50 ml), the resin was dried in a desiccator overnight. A mixture of anisole, thioanisole and ethylmethylsulphide (1 ml each) was added to the resin in a round-bottomed flask and stirred under nitrogen for 30 min before the addition of water (1 ml) and TFA (16 ml). After stirring for 3.5 h under nitrogen, the TFA was removed *in vacuo* and triturated with diethyl ether containing 2 % 2-mercaptoethanol (50 ml) to give a white solid which was isolated by filtration. The white solid was immediately dissolved in a solution containing 8 M urea, 50 mM ammonium acetate and 10 mM dithiothreitol (DTT) at pH 4.5 (15 ml) and left stirring overnight.

Purification of ubiquitin

Gel filtration was carried out using a high-resolution G-50 Sephadex (fine) column (800 mm × 30 mm) with a solution of 8 M urea, 50 mM ammonium acetate and 10 mM DTT at pH 4.5 as the eluent. The solution containing the crude peptide was loaded on the column, and the eluent monitored at 226 and 277 nm. Three major peaks were isolated, their fractions combined and sequentially dialysed (2000 Da Spectropor) in nitrogen-saturated solutions containing 6 M urea, 50 mM ammonium acetate and 10 mM DTT at pH 4.5, 4 M urea, 50 mM ammonium acetate and 10 mM DTT at pH 4.5 and 2 M urea, 50 mM ammonium acetate and 10 mM DTT at pH 4.5 for 24 h each step. As a final step the peptide was dialysed against 50 mM ammonium acetate at pH 4.5 for 4 × 24 h. The dialysed materials were then lyophilized to concentrate the solutions to approximately 30–50 ml. These fractions were assayed by isoelectric focusing (IEF) on phast gel pI range 3–9, analytical reversed-phase h.p.l.c. and amino acid analysis which indicated that the first fraction contained the largest amount of ubiquitin.

Ion-exchange chromatography

Cation-exchange chromatography was carried out using a CM-Sepharose (CL6B beads; Pharmacia) column (320 mm × 18 mm). Fraction 1 from the gel-filtration column was loaded on the column which was equilibrated in 50 mM ammonium acetate at pH 4.5. A pH gradient from 50 mM ammonium acetate at pH 4.5 to 50 mM ammonium acetate at pH 5.5 was run overnight before the column was run isocratically at 50 mM ammonium acetate, pH 5.5, for 4–5 h where the first protein fractions began to be eluted from the column. A salt gradient from 50 mM to 0.3 M ammonium acetate at pH 5.5 was then applied to elute the remaining protein material from the column. This gave five

distinct fractions which were isolated and repeatedly lyophilized to remove the excess salt. Amino acid analysis, analytical h.p.l.c. and IEF showed that the third fraction contained the required material.

Anion-exchange chromatography was carried out on the third fraction from the CM-Sepharose column using a DEAE-Sepharose column (320 mm × 18 mm), carefully equilibrated with 50 mM NH_4HCO_3 at pH 9.3. The protein was loaded in the NH_4HCO_3 buffer (15 ml) and a salt gradient of 50 mM to 0.3 M NH_4HCO_3 at pH 9.3 was run producing six peaks with a combined product weight of 41.8 mg, of which one fraction weighing 17.8 mg was judged to be identical with authentic bovine ubiquitin by analytical h.p.l.c., amino acid analysis and IEF.

Further purification of the remaining peptidyl-resin gave an overall yield of 93 mg (4.3 %) based on the expected yield of product from the resin.

Total enzymic digestion

Protein (1 mg) was dissolved in performic acid (100 μl) at 0 °C for 1 h after which the solution was lyophilized before being redissolved in 8 M urea/5 % formic acid (100 μl) and left for 2 h. This solution was dialysed against 5 % formic acid for 24 h before being lyophilized. The solid was taken up in 5 % formic acid (100 μl) and digested with pepsin (1 %, w/v; 25 μl) for 24 h after which a fresh sample of pepsin was added and the digestion allowed to proceed for a further 24 h. The solution was lyophilized, dissolved in 0.2 M NH_4HCO_3 buffer, pH 7.75, and microsomal leucine aminopeptidase (2 units/ml, 15 μl) and prolidase (200 units/ml, 28.1 μl) were added with MnCl_2 (50 mM) and left to digest for 48 h. After repeated lyophilization, the resulting residue was dissolved in citric acid buffer, pH 2.2, for amino acid analysis.

IEF and Western-blot analysis

IEF was carried out using an LKB flat-bed chamber 2117 Multipor fitted with a Pharmacia power pack ECPS 3000/150. Broad-range pH 3.5–9.5 PAG plate gels (Pharmacia) were run at 28 mA, 1300 V, at 5 W for 930 Vh. Samples were dissolved in 50 mM Tris/HCl, pH 8.0, to give 1 mg/ml concentrations of which 20 μl was loaded on the gel. After completion of the run, the gel was either stained with Coomassie Blue or prepared for electroblotting by being placed in a transfer buffer consisting of 250 mM Tris/HCl, pH 8.3, 1.5 M glycine and 20 % (v/v) methanol. After removal of the plastic backing from the gel, the protein was electroeluted on to nitrocellulose for 2 h at 0.6–1.2 mA. The nitrocellulose was treated with 10 mM Tris/HCl, pH 8.0, containing 400 mM NaCl (TBS) and 20 % (w/v) dried milk overnight. After being washed with TBS (3 × 5 min), the nitrocellulose was placed in a TBS and 5 % milk solution with rabbit polyclonal antibodies (5 μl) and shaken overnight. The nitrocellulose was then shaken overnight in a TBS solution with anti-rabbit IgG horseradish peroxidase (10 μl) before being washed (3 × 5 min) in TBS. Antibody binding to nitrocellulose was visualized by staining with a mixture of dianisidine (0.5 ml, 5 mg/ml), imidazole (1 mg, 0.1 M, pH 7.4), 30 % H_2O_2 (0.1 ml) and water (8.4 ml).

Sequencing

The purified product was sequenced by automated Edman degradation from residues 1–64 inclusively. No deletion peptides were detected. Sequencing was carried out on a 477A Applied

Biosystems protein sequencer by the Wellmet Protein Characterisation Unit at the University of Edinburgh.

Tryptic mapping

The ubiquitin samples (0.25 mg each) were submitted to oxidation with freshly prepared performic acid [100 vol. of H_2O_2 /formic acid (1:19, v/v), room temperature, 2 h] [24]. Oxidation was permitted to take place at 0 °C for 1 h. The oxidant was evacuated under reduced pressure overnight. The samples were reconstituted with 0.2 M ammonium acetate, pH 8.5 (100 μl), followed by addition of diphenylcarbamoyl chloride-treated trypsin (1 mg/ml; 15 μl). The stoppered tubes were incubated at 37 °C for 4 h, followed by freezing for 20 min. These samples were evacuated under reduced pressure overnight before reconstitution with 0.1 M NH_4OH in preparation for spotting on Whatman 3 MM paper for electrophoresis.

Analytical high-voltage paper electrophoresis was carried out at pH 6.5 at a potential gradient of 52 V/cm for 45 min in a toluene-cooled tank. When the positions of the bands had been located by the use of ninhydrin solution [0.1 % ninhydrin/acetone plus collidine/acetic acid (1:2, v/v)], preparative experiments were conducted to elute excised peptides for further analysis under the conditions used for analytical runs. In the preparative experiments, employing 3 mg of protein each, the positions of the peptides were located with guide strips which included 10 mm at each end of the sample zone and a narrow strip at the centre of the paper. These were cut out, sprayed with ninhydrin solution and heated at 105 °C for about 5 min to develop the stain. The electrophoretograms were reconstructed by attaching the stained strips to the unstained portion. The regions of the paper containing peptide bands were outlined with a rule (about 1 ml) and pencil, labelled and cut out. The peptides were eluted from the paper with 0.1 M NH_4OH using an elution apparatus. The poorly resolved bands near the origin (T7 and T8) were resubmitted to electrophoresis at pH 2.1 (acetic acid/formic acid) (70 V/cm, Whatman 3 MM) and pH 3.5 (pyridinium acetate) respectively, followed by elution of the resolved bands. The samples were processed by total hydrolysis in 6 M HCl in durham tubes (8 mm $\times$ 35 mm) and resubmitted to high-voltage paper electrophoresis at pH 2.1 (70 V/cm Whatman 1 MM). The amino acid composition was determined by comparison with amino acid standards. It was found, during the course of this work, that the LKB α analyser could detect the weak solutions eluted from these bands, thereby adding a further dimension, i.e. the quantitative analysis of these peptides.

N-Terminus determination by dansylation

The dry peptides (at useful concentrations) were placed in labelled durham tubes (6 mm $\times$ 30 mm), and 0.1 M NaHCO_3 (10 μl) was added to each sample. This was dried in a desiccator for 10 min. Water (10 μl) was added followed by dansyl chloride (10 μl ; 2.5 mg/ml acetone), and samples were covered with nescofilm to prevent evaporation of the acetone and incubated for 1 h at 37 °C. The samples were dried in a desiccator for 30 min, 6 M HCl (50 μl) was added to each sample and hydrolysis was carried out overnight (107 °C). The samples were again dried in a desiccator and 50 % pyridine (10 μl) was added. The samples were spotted on paper and electrophoresis was carried out at pH 4.4 (pyridinium acetate) for 2 h (75 V/cm) on a flat plate. The N-termini were identified by comparison with dansylated amino acid standards.

Tryptic digestion was carried out by the method of Cox et al. [24]. Protein (5 mg) was dissolved in Tris buffer, pH 8.2 at 37 °C,

containing 6.5 M urea (390 mg/ml) and 10 mM CaCl_2 . Trypsin (5 % w/w with the protein) was added and the digestion was allowed to proceed at 37 °C for 4 h before a further portion of trypsin (5 %, w/w) was added. After 8 h the protein solution was removed from the incubator and the pH adjusted to 2 with 6 M HCl before the protein solution was stored at -20 °C. Approx. 50–100 μg quantities of the solution were injected on the h.p.l.c. column and eluted using the following solutions: A, 0.1 % TFA plus 0.1 % triethanolamine in water; B, 0.1 % TFA plus 0.1 % triethanolamine in aq. 40 % acetonitrile. The gradient used was 10 % B for 2 min, then a gradient of 10–70 % B for 50 min, 70–100 % B for 2.5 min and 100 % B for 5 min before re-equilibration at 10 % B ready for the next injection. U.v. monitoring was carried out at 205 nm at an absorbance setting of 1.5. The peptide fragments were isolated in 1 ml fractions, lyophilized and subjected to amino acid analysis for identification.

RESULTS AND DISCUSSION

During the preliminary synthetic studies, there emerged inadequacies in the protocols described for the purification of ubiquitin. As our aim was to produce a crystalline product, we sought to develop a more effective purification procedure using commercially available ubiquitin of about 80 % purity, bearing in mind that the impurities therein would be different from those present in the crude synthetic material. Two features of the original mode of purification were retained, i.e. gel filtration and dialysis, but modifications had to be incorporated to render the synthetic material soluble, in contrast with the natural material. The 8 M urea used for this purpose is known to denature ubiquitin and therefore, advantageously, any misfolded synthetic ubiquitin would be deconvoluted. DTT was also included to prevent undesired aerial oxidation of the methionine residue which would prevent proper folding of ubiquitin [25]. Dialysis treatment of the desired fraction from gel filtration was primarily used as a means of refolding the ubiquitin by controlled removal of the denaturant. The most selective purification was achieved by cation-exchange chromatography on the acidic side of the isoelectric point of ubiquitin (pI 6.7), followed by anion-exchange chromatography. The dialysed ubiquitin was solubilized at pH 4.5 in 50 mM ammonium acetate and loaded on to the cation-exchange column. Optimum purification and recovery were achieved by a combination of pH gradient (4.5–5.5) and salt gradient (50 mM to 0.3 M ammonium acetate 5.5). We found that anion-exchange chromatography had to be performed at a pH higher than 9.0 in order to obtain any retention, and hence chromatographic discrimination, of ubiquitin. It was found that pH 9.3 was optimal but that great care had to be taken with column equilibration to obtain reproducible results. All of these purification stages were monitored by h.p.l.c. and IEF.

The solid-phase synthesis of ubiquitin was conducted using the N^{α} -Fmoc-protecting-group strategy designed by Carpino and Han [26] and developed extensively by Atherton et al. [27]. The synthesis of ubiquitin has been refined over three syntheses [28,29].

Loading of Fmoc-Gly-OH on the *p*-alkoxybenzyl alcohol resin (Wang resin) [30] was effected using the acid chloride method [31,32] in DCM using ultrasonication to achieve mixing. No glycylglycine could be detected by amino acid analysis and the degree of loading of Fmoc-Gly of 0.29 mmol/g of resin was established by u.v. analysis of a loaded resin sample, deprotected by 20 % piperidine in DMF.

The guanidino group of arginine was protected by the Pmc group which was designed [33] with ubiquitin synthesis in mind

because of the reported tendency of the terminal Gly-Gly residues to be cleaved under mild conditions. The imidazole of histidine was protected by the trityl group, and the amide functions of Asn and Gln were protected by the 4,4-dimethoxybenzhydryl (Mbh) group. In general, the amino acids were coupled once as the symmetrical anhydride using DIC in dioxan and twice as the hydroxybenzotriazole (DIC/HOBt) active ester. The exceptions to this were Fmoc-His (Trt), coupled three times as the HOBt active ester to avoid the racemization that occurred in early synthetic studies. Fmoc-Asn (Mbh) and Fmoc-Gln (Mbh) were also coupled three times as the HOBt active ester. Fmoc-Gly was coupled once in a single symmetrical anhydride step to avoid incorporation of glycylglycine which can form as a result of a repeated symmetrical anhydride coupling.

After each coupling, acetic anhydride/pyridine in DMF was added and the capping reaction was allowed to proceed for 6 min. Each N^α -Fmoc deprotection stage involved successive treatments with 20% piperidine in DMF for periods of 5, 3, 3 and 1 min. In order to enable real-time assessment of the synthesis to be achieved, a diluted sample of the deprotection solution containing the chromophoric Fmoc-piperidine adduct was monitored at 302 nm in a continuous flow mode from the synthesizer to a u.v. detector. In this way, the efficiency of each coupling and deprotection cycle could be monitored and an estimate of the overall yield of the protein at the completion of the synthesis could be gauged. This also maximizes the yield of resin-bound protein at the completion of the synthesis, as there is no resin loss because of the need for stepwise ninhydrin analysis. On completion of the assembly of the protected ubiquitin, the final weight of peptidyl resin was 3.01 g compared with the initial polystyrene resin of 1.15 g. As no low coupling steps were detected, on the basis of the u.v. measurements, the overall yield of the peptide assembly on the resin was expected to be approximately 45%.

The Fmoc group was left attached to the final amino acid, methionine, thus a quantitative assessment could be made of the amount of the final amino acid coupled to the peptidyl-resin. Duplicate u.v. analysis of the Fmoc-piperidine adduct suggested that Fmoc-Met coupled to the peptidyl-resin with an efficiency of $53 \pm 5\%$.

The N^α -Fmoc protecting group of the peptidyl resin was removed as required with 20% piperidine/DMF in a sonicator over 10 min. In order to remove the side-chain-protecting groups and cleave the peptide from the resin, a sample of the peptidyl-resin (approx. 1 g) was treated with a mixture of thioanisole (5%, 1 ml), ethylmethylsulphide (5%, 1 ml) and anisole (5%, 1 ml) for 20 min under nitrogen before the addition of water (5%, 1 ml) and TFA (16 ml) at room temperature for 3.5 h. The optimal cleavage time (3.5 h) was established by conducting a small (50 mg)-scale experiment using the same concentrations of cleavage reagents on a reduced scale. Samples of the TFA mixture were removed from these trial cleavages and triturated with ether (containing 2%, 2-mercaptoethanol) to remove the scavengers and precipitate the peptide. The peptide was isolated by filtration and dissolved in aqueous acetic acid for analysis by h.p.l.c.

After the deprotection stage, the TFA was removed *in vacuo* and the residue was stirred with diethyl ether (50 ml) containing 2% 2-mercaptoethanol and then filtered to isolate the insoluble peptide. The resulting white solid was dissolved in a nitrogen-saturated solution containing 8 M urea, 10 mM DTT and 50 mM ammonium acetate at pH 4.5 for 16 h. The protein solution was subjected to gel filtration on a Sephadex G-50 (fine) column (800 mm $\times$ 30 mm) previously equilibrated in the above nitrogen-saturated solution (Figure 2). The protein fractions eluted from

Figure 2 G-50 Sephadex gel-filtration profile

The shaded area is the protein fraction isolated for subsequent purification.

the column were isolated and dialysed slowly against decreasing concentrations of urea and DTT until the final dialysis step against 50 mM ammonium acetate at pH 4.5. On completion of the dialysis, the solution was concentrated *in vacuo* to 30–50 ml. The fractions were analysed by amino acid analysis (Table 1), IEF and analytical h.p.l.c.

The desired fraction was further purified by cation-exchange chromatography on a CM-Sepharose column with a pH gradient from 50 mM ammonium acetate, pH 4.5, to 50 mM ammonium acetate, pH 5.5, then isocratically at pH 5.5 before a linear gradient from 50 mM to 0.3 M ammonium acetate, pH 5.5, was applied to elute the protein (Figure 3). The fractions were isolated and combined on the basis of the analytical h.p.l.c. profiles and elution times. These combined fractions were lyophilized repeatedly to remove the ammonium acetate, and the resulting lyophilized protein was analysed by amino acid analysis (Table 1), IEF and analytical h.p.l.c.

The requisite fraction from the CM-Sepharose column was dissolved in 50 mM NH_4HCO_3 at pH 9.3 and loaded on a pre-equilibrated DEAE-Sepharose column. A linear gradient of 50 mM to 0.3 M NH_4HCO_3 at pH 9.3 was applied and the fractions corresponding to the two peaks A and B were isolated (Figure 4) and lyophilized repeatedly to provide two samples of purified ubiquitin, which gave h.p.l.c., IEF and amino acid analyses (Table 1) identical with authentic bovine ubiquitin (Sigma).

The analysis also showed that the two distinct protein fractions isolated from the DEAE-Sepharose column were of identical chemical composition. We designated them type A, which was found to be identical with natural bovine ubiquitin, and type B (Figure 4). Natural ubiquitin at pH 9.3 was also found to equilibrate to A and B on the DEAE-Sepharose column. This interesting finding is discussed in the following paper [34].

Further characterization of the primary structure of the protein was carried out by automated sequencing Edman degradation which was able to identify the sequence to residue 64 without ambiguity. This technique alone, starting as it does at the portion of the molecule synthesized last, almost suffices to prove that the primary structure was free of deletions. Further characterization was effected by total enzymic digestion and subsequent amino acid analysis (Table 1).

Bovine and synthetic ubiquitin samples were submitted to performic acid oxidation [24] before separate enzymic digestion with trypsin, chymotrypsin, thermolysin and *Staphylococcus*

Table 1 Amino acid composition of the best fractions of synthetic ubiquitin at various stages of the purification process

Acid hydrolyses were carried out at 100 °C in 6 M HCl with 5 mg/ml Na₂SO₃ for 40 h in a sealed tube. Enzymic hydrolyses of the DEAE-Sepharose-purified ubiquitin were carried out as described in the Experimental section. Values in parentheses denote the theoretical values for Asp and Glu respectively. The enzymic digest of bovine ubiquitin is shown for comparison with the synthetic ubiquitin. ND, Not determined.

Amino acid	Expected values	Gel filtration	CM-Sepharose	DEAE-Sepharose	Enzymic digest	
					Bovine ubiquitin	Synthetic ubiquitin
Asx	7	7.44	7.40	7.54	4.21 (4)	4.90 (4)
Thr	7	5.12	4.12	5.88	9.30	8.85
Ser	3	2.64	0.78	2.29	4.67	4.63
Glx	12	10.82	13.62	14.00	7.72 (6)	6.92 (6)
Pro	3	ND	3.30	3.54	2.32	2.40
Gly	6	5.66	5.83	6.29	6.24	7.05
Ala	2	1.98	1.79	2.05	1.81	2.11
Val	4	4.64	4.00	4.02	4.27	4.17
Met	1	0.39	0.95	0.95	Met O	Met O
Ile	7	5.83	6.53	6.78	6.87	6.65
Leu	9	8.11	9.29	9.22	10.35	10.66
Tyr	1	1.45	1.26	0.98	0.98	1.46
Phe	2	1.65	2.11	2.06	2.70	2.50
His	1	0.99	1.02	0.94	1.05	0.92
Lys	7	5.28	7.02	7.04	6.36	6.30
Arg	4	3.74	4.06	4.05	3.93	3.56

Figure 3 Cation-exchange chromatography on a CM-Sepharose column (300 mm × 18 mm)

The sample was loaded in 50 mM ammonium acetate at pH 4.5, then a gradient of pH 4.5–5.5 was applied before an isocratic elution period with 50 mM ammonium acetate at pH 5.5. A salt gradient of 50 mM to 0.3 M ammonium acetate at pH 5.5 was then used to elute the synthetic ubiquitin as shown by the shaded area.

aureus V8 protease. This procedure was adopted in the expectation that methionine would be converted into the sulphone, thereby denaturing the structure and rendering it more open to enzymic attack. Enzymic digestion with trypsin [35] followed by isolation of the resulting peptide fragments by h.p.l.c. (Figure 5), and subsequent analysis of the peptide fragments by amino acid analysis (Table 2) showed that all the isolated fragments were in agreement with the primary sequence of the protein. The fragmentation patterns obtained from the digestion of natural and synthetic ubiquitin with chymotrypsin, thermolysin and

Figure 4 Anion-exchange chromatography on a DEAE-Sepharose column (300 mm × 18 mm)

The protein was eluted from the column using a linear gradient of 50 mM–0.3 M NH₄HCO₃ at pH 9.3 to elute the two forms of ubiquitin, A and B.

Staphylococcus V8 protease confirmed the identity of the primary structure.

Electrospray m.s. of the purified synthetic protein showed it to be of identical mass (8564.8) with that of bovine ubiquitin (8565) (Figure 6).

An antibody-binding assay was carried out to compare the immunoreactivity of bovine ubiquitin and synthetic ubiquitin. After IEF, the protein was electroblotted on to nitrocellulose. The protein was then incubated in the presence of rabbit and anti-ubiquitin IgG polyclonal antibodies. On development and staining for antibodies, it was observed that the synthetic ubiquitin exhibited antibody recognition identical with that of bovine ubiquitin. In *in vitro* assays involving conjugation to

Figure 5 Tryptic map of synthetic ubiquitin

Ubiquitin (5 mg) in buffer containing 6.5 M urea, 10 mM CaCl_2 , 50 mM Tris/HCl, pH 8.2, at 37 °C was incubated with 5% (w/w) trypsin for 4 h, before a further portion of trypsin (5% w/w) was added. On completion of the digestion (8 h), the pH was adjusted to 2 with 6 M HCl, and the digest stored at -20 °C. Tryptic mapping was performed on a C18 Vydac column (250 mm $\times$ 4.1 mm internal diameter) using an ABI 151 h.p.l.c. system. Solution A was 0.1% TFA and triethanolamine in water. Solution B was 0.1% TFA and triethanolamine in 40% acetonitrile/water. The gradient used was 10% B for 2 min, 10–70% B over 50 min, then 70–100% B over 2.5 min. Peptide fragments were monitored at 205 nm.

Table 2 Identification of the tryptic peptides from synthetic ubiquitin

The peptide sequence was determined by amino acid analysis of the isolated peptide. Residue numbers are given in parentheses.

Peptide	Peptide sequence
T0	Ala-Lys (28–29)
T1	Ile-Gln-Asp-Lys (30–33)
T2	Leu-Arg-Gly-Gly (73–76)
T3	Thr-Leu-Thr-Gly-Lys (7–11)
T4	Gln-Leu-Glu-Asp-Gly-Arg (49–54)
T5	Glu-Gly-Ile-Pro-Pro-Gln-Gln-Arg (34–42)
T6	Thr-Leu-Ser-Asp-Tyr-Asn-Ile-Gln-Lys (55–63)
T7	Leu-Ile-Phe-Ala-Gly-Lys (43–48)
T8	Met-Gln-Ile-Phe-Val-Lys (1–6)
T9	Glu-Ser-Thr-Leu-His-Leu-Val-Leu-Arg (64–72)
T10	Thr-Ile-Thr-Leu-Glu-Val-Glu-Pro-Ser-Asp-Thr-Ile-Glu-Asn-Val-Lys (12–27)

lysozyme by the E1 enzyme, native and synthetic ubiquitin exhibited identical activities (M. P. Mee, J. E. Arnold and J. Mayer, personal communication).

Conclusion

The preliminary reports on the synthesis of ubiquitin by us and subsequently by Briand et al [36] quoted yields of about 1%. Although this synthesis afforded 4.3% of pure ubiquitin in useful quantities (93 mg), we were disappointed in the apparent disproportionate losses incurred during the deprotection/purification stages compared with the estimated 45% efficiency of the assembly phase. We consider that optimization of the deprotection conditions, which are to some extent dependent on the sequence, is vital to the ultimate success of the synthesis. The

Figure 6 Electrospray m.s. plot showing the calculated masses for the synthetic protein at 8564.8 plus the sodium salt at 8586.0

Fraction B from the DEAE-Sepharose column gave an identical mass.

present range of protecting groups may have to be refined further.

The work described here represents the chemical synthesis of a small protein, ubiquitin, together with proof of its homogeneity and elucidation of the primary structure (sequence). This synthesis, and indeed the synthesis of several ubiquitin analogues, have been performed several times and important lessons have been learned that should prove valuable for the total synthesis and proper characterization of significantly larger proteins. Synthetic ubiquitin has been crystallized, and the X-ray-crystallographic data identity of synthetic with natural crystalline ubiquitin is reported in the following paper [34]. The products of chemical synthesis of proteins are not commonly isolated in crystalline form. Other notable examples of crystalline proteins synthesized by chemical methods are human immunodeficiency virus 1 protease [37] and ribonuclease A [38].

We thank the University of Edinburgh/SERC for support of the Edinburgh Centre for Molecular Recognition, and Merck Sharp & Dohme, Applied Biosystems and ICI for general support. It is a pleasure to acknowledge the technical support given by Mr. Brian Whigham. In addition, we acknowledge the assistance given by Professor Richard Ambler and Margaret Daniel of the Department of Molecular Biology. We acknowledge the sequencing service given by Dr. Linda Gilmore and WELMET.

REFERENCES

- Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. and Schultz, P. G. (1989) *Science* **244**, 182–188
- Holler, E. (1978) *Angew. Chem. Int. Ed. Engl.* **17**, 648–656
- Hersko, A. and Ciechanover, A. (1982) *Annu. Rev. Biochem.* **251**, 335–364
- Goldstein, G., Scheid, M., Hammerling, U., Boyse, E. A., Schlesinger, D. H. and Niall, H. D. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 11–15
- Hough, R. F., Pratt, G. W. and Rechsteiner, M. (1988) in *Ubiquitin* (Rechsteiner, M., ed.), pp. 101–134, Plenum Publishing Corp, New York

- 6 Hough, R., Pratt, G. and Rechsteiner, M. (1986) *J. Biol. Chem.* **261**, 2400–2480
- 7 Waxman, L., Fagan, J. M. and Goldberg, A. L. (1987) *J. Biol. Chem.* **262**, 2451–1457
- 8 Findlay, D., Ciechanover, A. and Varshausky, A. (1984) *Cell* **37**, 43–55
- 9 Hunt, H. T. and Dayhoff, M. O. (1977) *Biochem. Biophys. Res. Commun.* **74**, 650–655
- 10 Nickel, B. E. and Davie, J. R. (1987) *Biochemistry* **28**, 964–968
- 11 Glotzer, M., Murray, A. W. and Kirschner, M. W. (1991) *Nature (London)* **349**, 132–138
- 12 Siegelmen, M., Bond, M. W., Gallatin, T., St John, T., Smith, H. T., Fried, V. A. and Weissman, I. L. (1986) *Science* **231**, 823–829
- 13 Yarden, Y., Escobedo, J. A., Kuang, W. J., Yang-Feng, T. L., Daniel, T. O., Tremble, P. M., Chen, E. Y., Ando, M. E., Harkins, R. N., Francke, U., Fried, V. A., Ulrich, A. and Williams, L. T. (1986) *Nature (London)* **232**, 226–232
- 14 Mori, H., Kondo, J. and Ihara, Y. (1987) *Science* **235**, 1641–1644
- 15 Salvesen, G., Lloyd, C. and Farley, D. (1987) *Nucleic Acids Res.* **15**, 5485
- 16 Lund, P. K., Moats-Staats, B. M., Simmons, J. G., Hoyt, E., D'Ercole, A. J., Montin, F. and Van Wyk, J. J. (1985) *J. Biol. Chem.* **260**, 394–401
- 17 Finlay, D., Barte, B. and Varshavsky, A. (1989) *Nature (London)* **338**, 394–401
- 18 Redman, K. L. and Rechsteiner, M. (1989) *Nature (London)* **338**, 438–440
- 19 Monia, B. P., Ecker, D. J., Jonnalagadda, S., March, J., Gotlib, L., Butt, T. R. and Crooke, S. T. (1989) *J. Biol. Chem.* **264**, 4093–4103
- 20 Cherney, B. W., Chaudhry, B., Bhatia, K., Butt, T. R. and Smulson, M. (1991) *Biochemistry* **30**, 10420–10427
- 21 Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D. and Cook, W. J. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3582–3585
- 22 Weber, P. L., Brown, S. C. and Mueller, L. (1987) *Biochemistry* **26**, 7282–7290
- 23 Distefano, E. L. and Wand, A. J. (1987) *Biochemistry* **26**, 7272–7281
- 24 Bamezai, W., Banez, M. A. T. and Breslow, E. (1990) *Biochemistry* **29**, 5389–5396
- 25 Cox, M. J., Shapira, R. and Wilkinson, K. D. (1986) *Anal. Biochem.* **154**, 345–352
- 26 Carpino, L. A. and Han, G. Y. (1970) *J. Am. Chem. Soc.* **92**, 5748–5749
- 27 Atherton, E., Gait, M. J., Sheppard, R. C. and Williams, B. J. (1979) *Bioorganic Chem.* **8**, 351–370
- 28 Ramage, R., Ogunjobi, O. M. and Green, J. (1989) *Tetrahedron Lett.* **30**, 2149–2152
- 29 Ogunjobi, O. Y. and Ramage, R. (1990) *Biochem. Soc. Trans.* **18**, 1322–1323
- 30 Wang, W. W. (1973) *J. Am. Chem. Soc.* **95**, 1328–1333
- 31 Fisher, E. (1903) *Ber. Deutsch. Chem. Ges.* **36**, 2094–2106
- 32 Carpino, L. A., Cohen, B. J., Stephens, K. E., Sadat-Aalace, S. Y., Tien, J.-H. and Langridge, D. C. (1986) *J. Org. Chem.* **51**, 3732–3734
- 33 Ramage, R. and Green, J. (1987) *Tetrahedron Lett.* **28**, 2287–2290
- 34 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
- 35 Benoiton, N. L. and Chen, F. M. F. (1986) *Peptides*, 127–130
- 36 Briand, J.-P., Van Drosselar, A., Raboy, B. and Muller, S. (1989) *Peptide Res.* **2**, 381–388
- 37 Miller, M., Schneider, J., Sathymaryara, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Self, L., Kent, S. B. H. and Wlodawer, A. (1989) *Science* **246**, 1149–1152
- 38 Yajima, H. and Fujii, N. (1989) *Biopolymers* **20**, 1859–1867