Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3582–3585, June 1985
Biochemisty

Three-dimensional structure of ubiquitin at 2.8 Å resolution

(crystal structure/protein conformation)

SENADHI VIJAY-KUMAR*†, CHARLES E. BUGG*†‡, KEITH D. WILKINSON§, AND WILLIAM J. COOK*†¶‖

Departments of ¶Pathology and ‡Biochemistry, *Comprehensive Cancer Center, and †Institute of Dental Research, University of Alabama at Birmingham, University Station, Birmingham, AL 35294; and §Department of Biochemistry, School of Medicine, Emory University, Atlanta, GA 30322

Communicated by David R. Davies, February 1, 1985

ABSTRACT **The three-dimensional structure of ubiquitin has been determined at 2.8 Å resolution. X-ray diffraction data for the native protein and derivatives were collected with an automated diffractometer. Phases were obtained by use of a single isomorphous mercuric acetate derivative. The molecule contains a pronounced hydrophobic core. Prominent secondary structural features include three and one-half turns of α-helix, a mixed β-sheet that contains four strands, and seven reverse turns. The histidine, tyrosine, and two phenylalanine residues are located on the surface of the molecule.**

Ubiquitin is a small protein believed to be universally present in living cells (1). It consists of a single 8565-Da polypeptide chain of 76 amino acids. Ubiquitin has been isolated and sequenced from a variety of sources. Insect (2), trout (3), bovine (4), and human (5) ubiquitins have identical primary structures through arginine-74. In the original sequencing studies of bovine and human ubiquitin, arginine-74 was the COOH-terminal residue. However, subsequent studies indicated that ubiquitin has 76 amino acids and terminates with Gly-Gly (6, 7). It now appears that the 76-amino acid polypeptide is the physiologically active form (7, 8). Ubiquitin that lacks the COOH-terminal Gly-Gly is most likely an *in vitro* proteolytic artifact.

Ubiquitin was initially thought to be a hormone that induced lymphocyte differentiation and activated adenylate cyclase (1, 9, 10). However, these activities were not confirmed by later workers (11, 12). More recently, interest has focused on ubiquitin–protein conjugates, in which ubiquitin is covalently linked via its carboxyl terminus to ε-amino groups of lysine residues in a wide variety of intracellular proteins. The first such conjugate described was the chromosomal protein A24 (currently designated uH2A), which consists of ubiquitin conjugated to histone H2A via an isopeptide bond involving the COOH-terminal glycine of ubiquitin and the ε-NH_2 group of lysine-119 of histone H2A (8, 13). It was subsequently shown that ubiquitin is conjugated *in vivo* to all subtypes of histone H2A and also to H2B, although at much lower levels (14, 15). The function of this conjugate is not known, but it may be involved in the transcription of active genes (16).

Other studies have shown that conjugation of ubiquitin with proteins is required for intracellular ATP-dependent protein degradation (17, 18). The proposed mechanism of ubiquitin action in this system is through isopeptide bond formation to substrates for proteolysis; one possible role of ubiquitin is to serve as a signal for attack by proteinases specific for ubiquitin–protein conjugates. Ubiquitin-mediated, nonlysosomal proteolysis has been observed in a variety of mammalian cell types (18). Most recently, ubiquitin–protein conjugation has been studied in the mouse cell line ts85, in which the conjugation is temperature sensitive (19, 20). It has been shown that this effect is due to the specific thermolability of the ts85 ubiquitin-activating enzyme. By using this cell line as a probe, it was demonstrated that the majority of short-lived proteins in this cell undergo degradation through a ubiquitin-dependent pathway. Other workers have shown that the degradation of abnormal proteins is also carried out by the ubiquitin pathway (21, 22). Thus, ubiquitin is a small protein with a large number and diversity of protein–protein interactions.

Apart from its physiological roles, ubiquitin is of interest because of its stability and unparalleled sequence conservation (2–5). It is noteworthy that the molecule is extremely resistant to tryptic digestion despite the presence of seven lysine residues and four arginine residues (4). It is also quite stable over a wide range of pH and temperature values. Studies with nuclear magnetic resonance (NMR) demonstrated no apparent denaturation over a temperature range of 23°C–80°C and a pH range of 1.18–8.48 (23). Estimations of its secondary structure have generally emphasized a highly globular compact conformation with a low percentage of α-helix and β-sheet (24, 25). In this paper, we describe the structure of ubiquitin at 2.8 Å (Fig. 1).

METHODS

Ubiquitin from human erythrocytes was isolated and purified as described (7). The crystallization technique was the same as described by Cook *et al.* (26); all crystals were obtained by seeding at pH 5.6. The crystals, which grow as large rectangular prisms, belong to orthorhombic space group $P2_12_12_1$ with $a = 50.73$ Å, $b = 42.75$ Å, and $c = 28.78$ Å; there is one molecule of ubiquitin per asymmetric unit. A small amount of ubiquitin was treated with trypsin to obtain a 74-amino acid polypeptide that lacked the glycylglycine COOH terminus. Crystals of this were initially obtained by seeding with native ubiquitin crystals. These crystals were then used as seeds to obtain large single crystals of the 74-amino acid polypeptide (ubiquitin-74).

The single heavy-atom derivative used in the solution of the structure was prepared by soaking crystals for 6 days at 4°C in a 32% (wt/vol) solution of PEG-4000 that contained 0.05 M cacodylate (pH 7.0) and 1 mM mercuric acetate. Four other derivatives [K_2PtCl_4, $K_2Pd(NO_2)_4$, KI_3, and $Hg(C_2H_3O_2)_2$ soaked at pH 6.5] were also prepared, but two of these (Pt and Pd) were isomorphous only at low resolution, and the other two (Hg and I) were single-site derivatives with low occupancies; therefore, none of these other derivatives was used in the isomorphous replacement phasing.

Data for native and derivative crystals were collected with a Picker FACS-1 diffractometer at room temperature using an ω step-scan procedure and nickel-filtered CuKα radiation. The reflections were divided into 2θ shells containing 100 reflections each and were collected from high to low resolution. Friedel mates were generated during data collection and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‖To whom reprint requests should be addressed.

```
                                5                                  10
Met - Gln - Ile - Phe - Val - Lys - Thr - Leu - Thr - Gly -
                               15                                  20
Lys - Thr - Ile - Thr - Leu - Glu - Val - Glu - Pro - Ser -
                               25                                  30
Asp - Thr - Ile - Glu - Asn - Val - Lys - Ala - Lys - Ile -
                               35                                  40
Gln - Asp - Lys - Glu - Gly - Ile - Pro - Pro - Asp - Gln -
                               45                                  50
Gln - Arg - Leu - Ile - Phe - Ala - Gly - Lys - Gln - Leu -
                               55                                  60
Glu - Asp - Gly - Arg - Thr - Leu - Ser - Asp - Tyr - Asn -
                               65                                  70
Ile - Gln - Lys - Glu - Ser - Thr - Leu - His - Leu - Val -
                               75
Leu - Arg - Leu - Arg - Gly - Gly
```

FIG. 1. Amino acid sequence of ubiquitin.

were collected at negative 2θ values in groups of 10 reflections, alternating with their unique counterparts. The ω scan width was typically 0.8°–1.2° and was selected so that, for most reflections, half of the scan width contained the peak and the other half corresponded to background. A typical scan step size was 0.02°–0.03°.

The crystals were always mounted with the a crystallographic axis parallel to the capillary axis. Absorption corrections were applied according to the method of North *et al.* (27). To monitor and correct for decomposition effect, three to six standard reflections were measured periodically. During the collection of data to 2.8 Å resolution, the intensities of the standard reflections showed a <5% decrease.

A unique octant of data, plus the complete set of Friedel mates, was collected for the native protein, ubiquitin-74, and the Hg derivative. The integrated intensities were obtained in a two-step process during which several of the shells were processed together. Peak positions were determined by the method of Tickle (28). In the first pass through the data, peak widths were obtained for the stronger reflections by using the methods of Lehman and Larsen (29). These peak widths were then expressed as a function of the diffractometer angles ϕ and χ. In the second pass through the data, all reflections were assigned peak widths that were calculated from the ϕ, χ function. The steps outside the peaks were used to determine the background level for each individual reflection.

The overall quality of the data sets was evaluated by comparing the intensities of Friedel mates. We computed the reliability index

$$R_{\text{sym}} = \frac{\Sigma|I(h) - I(\bar{h})|}{\Sigma(I(h) + I(\bar{h}))},$$

in which the sums included all Friedel mates for the native data and ubiquitin-74 and those Friedel mates from the centric zones for the Hg derivative. The resultant R_{sym} values were 0.018 for the native data, 0.014 for ubiquitin-74, and 0.031 for the Hg data. The data for ubiquitin-74 and for the derivatives were scaled to the native data by means of relative Wilson plots (30), followed by equating the average F^2 values for the individual shells of data. The mean fractional isomorphous differences for the Hg derivative and for ubiquitin-74 [$(\Sigma|\,|F_{PX}| - |F_P|\,|)/\Sigma|F_P|$, where F_{PX} is the structure factor for the derivative or for ubiquitin-74 and F_P is the native structure factor] are 0.274 and 0.097, respectively.

The Hg sites were located by use of an isomorphous difference Patterson map, and these sites were refined by least-squares analysis using data from the centric zones. The reliability index based on 344 centric reflections between 10 and 3 Å $[\Sigma|F_H(\text{obs}) - |F_H(\text{calc})|\,|/\Sigma F_H(\text{obs})$, in which $F_H(\text{obs}) = |F_{PH} - F_P|]$ is 0.45. The final model for the Hg derivative includes two sites that are ≈6 Å apart. The major site forms close contacts with two symmetry-related molecules of ubiquitin. This site is ≈3 Å from the carboxyl group of aspartate-32 and ≈3 Å from one of the two nitrogen atoms in the imidazole ring of histidine-68. The minor site is ≈3 Å from the other nitrogen atom in the imidazole ring of histidine-68. Various statistics describing the single isomorphous-replacement phasing are summarized in Fig. 2. The figure of merit is 0.76 for the complete data set.

Phase angles for the native data were calculated from the single isomorphous derivative including anomalous dispersion effects. The enantiomer was assigned by using cross-difference Fourier maps, computed with phase sets obtained from both possible enantiomeric arrangements of Hg atoms, to examine the heavy atom sites in the K_2PtCl_4, $K_2Pd(NO_2)_4$, KI_3 and $Hg(C_2H_3O_2)_2$ derivatives. An electron density map for the protein was calculated including the 1734 native structure factors with $d > 2.8$ Å and using centroid phase angles with figure-of-merit weights. The map was contoured on acetate sheets at a scale of 0.474 cm/Å, with a spacing between sheets along c corresponding to 1.16 Å.

RESULTS

The electron density map clearly revealed the molecular boundary, and essentially all of the polypeptide chain could be readily traced. Most of the individual amino acids were identifiable, particularly the proline residues and those with aromatic side chains. By use of the primary structure, we were able to follow the entire sequence by inspection of the acetate sheets. Only the last two residues at the COOH terminus did not lie in continuous density, which suggests that this part of the molecule may have large thermal motion or may be disordered. To fit this portion of the molecule, we calculated a difference Fourier map using the single isomorphous-replacement phases and amplitudes that were the difference between structure factors for ubiquitin-74 and those for native ubiquitin. The only significant electron density in this difference map was in the region of the COOH terminus, and we were able to obtain a reasonable fit of the Gly-Gly segment to this density.

Guide coordinates for Cα, Cβ, and the end atom for each of the larger side chains were taken from the single isomorphous-replacement map for the first 74 residues. Guide coordinates for the terminal glycylglycine dipeptide were taken from the difference Fourier map. These guide coordinates

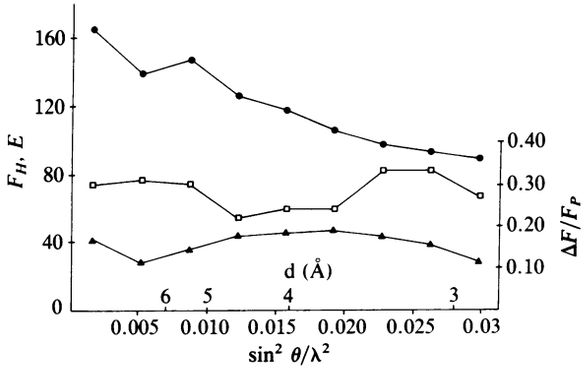

FIG. 2. Statistics for the Hg derivative as a function of resolution. ●, F_H is the calculated heavy-atom contribution to the structure factors; ▲, E is lack-of-closure error; □, $\Delta F/F_P$ is equal to $|F_{PH} - F_P|/F_P$ where F_{PH} and F_P are the structure factors for the heavy-atom derivative and native protein, respectively.

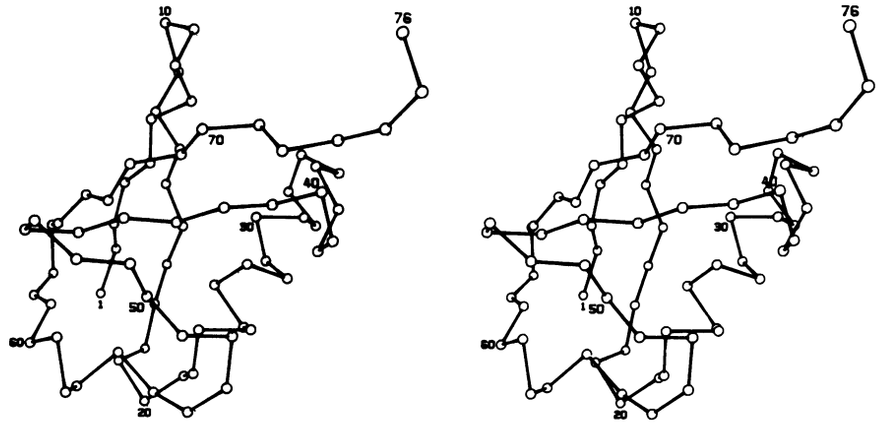

FIG. 3. Stereo drawing of the α-carbon backbone of ubiquitin. This drawing was prepared by using the computer program ORTEP (32).

were used in conjunction with the computer graphics program FRODO (31) to generate atomic coordinates for the entire molecule. The coordinates were adjusted to approximate ideal geometry by the model building program included in the FRODO package. The model was then inspected and improved by use of the FRODO interactive graphics system. This improved model was further refined by several additional cycles of regularization. The reliability index for the final graphics model based on the 2.8 Å data and using an overall isotropic temperature factor of 12 $Å^2$ is 0.46.

A tracing of the α-carbon backbone is shown in Fig. 3. The most prominent secondary structural features are 3.5 turns of α-helix involving residues 23–34 and a mixed β-sheet that contains four strands. The two inner strands are parallel and include residues 1–7 and 64–72. The two outer strands are antiparallel to the inner strands and include residues 11–17 and 40–44. The α-helix occurs in the sequence between the two outer strands of the β-sheet and packs against the sheet. These features are depicted schematically in Fig. 4. The molecule appears to contain seven reverse turns that involve residues 8–11, 18–21, 37–40, 45–48, 51–54, 57–60, and 62–65. There is an unusual bend in the molecule involving residues 36–40, which include a Pro-Pro sequence. Another important structural feature is a large loop involving residues 51–59; this loop is stabilized by a hydrogen bond between the hydroxyl oxygen atom of the tyrosine-59 side chain and the peptide N atom of glutamate-51.

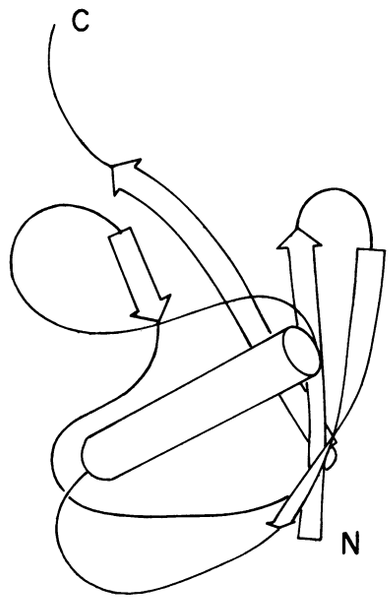

FIG. 4. Schematic diagram of ubiquitin. The cylinder and ribbons represent the α-helix and β-strands, respectively. The NH_2 and COOH termini are labeled N and C, respectively. This drawing was produced by a computer program written by Lesk and Hardman (33).

DISCUSSION

As predicted by NMR and chemical studies, ubiquitin is a compact, globular protein. The protein has a pronounced hydrophobic core; of the 21 valine, leucine, isoleucine, and methionine residues, 16 are buried within the interior of the molecule. This hydrophobic core may be a reason for the marked stability of ubiquitin. In addition to this, there is also a large amount of secondary structure. On the basis of circular dichroism results, Cary *et al.* (24) predicted 28% α-helix and 12% β-sheet, while Jenson *et al.* (25) predicted 6% α-helix and 10% β-sheet. In the crystal structure, however, the single α-helix contains 16% of the residues, the β-sheet corresponds to 37% of the chain, and the seven reverse turns represent 37% of the chain. Thus, ≈90% of the polypeptide chain is involved in hydrogen-bonded secondary structure.

NMR studies (24, 25) suggest that the tyrosine and histidine residues, as well as one of the phenylalanine residues, are in hydrophobic environments and may be buried. We find that all of these residues are located on the surface of the protein. The histidine side chain is clearly accessible to solvent, as evidenced by the finding that the mercury derivative reacts with the histidine residue. The nearest side chains around the histidine are provided by lysine-6, isoleucine-44, and threonine-66. Although the two phenylalanine residues are located on the surface of ubiquitin, one of these (phenylalanine-45) is in a shallow hydrophobic pocket composed of alanine-46, isoleucine-61, and leucine-67. The environment of the single tyrosine (residue 59) is of some interest, because iodination of the tyrosine renders ubiquitin much more susceptible to cleavage by trypsin (unpublished observations). The tyrosine residue is involved in a reverse turn that is composed of residues 57–60. In addition, we find that the tyrosine ring spans the large loop involving residues 51–59 and contributes to the stability of the loop by formation of a hydrogen bond between N of glutamic acid-51 and OH of the tyrosine side chain. Examination of the structure indicates that diiodination of the tyrosine ring would result in steric hindrance that might disrupt the stabilizing interactions that involve this residue.

It has been found that the single amino-terminal methionine of ubiquitin is buried and cannot be readily alkylated (unpublished observations). This finding is consistent with our current model, in which the sulfur of methionine-1 is hydro-

gen bonded to the backbone N of lysine-63 and is somewhat shielded from solvent.

Ubiquitin forms conjugates with other proteins via isopeptide bonds between the COOH-terminal glycine residue of ubiquitin and ε-amino groups of lysine residues on the other proteins. In the only characterized ubiquitin–protein complex, ubiquitin–histone H2A, it has been reported that there are probably no allosteric, noncovalent interactions between ubiquitin and histone H2A (25). This is easy to reconcile with the crystal structure, because the COOH terminus is protruding from the structure and is not interacting with the rest of the molecule by any hydrogen bonding or hydrophobic interactions. Thus, this portion of the molecule has considerable freedom of motion and should be accessible by enzymes involved in formation or cleavage of the isopeptide bond.

This research was supported by National Institutes of Health Research Grants GM-27144, CA-13148, and DE-02670. W.J.C. is the recipient of National Institutes of Health Research Career Development Award CA-00696.

1. Goldstein, G., Scheid, M., Hammerling, U., Boyse, E. A., Schlesinger, D. H. & Niall, H. D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 11–15.
2. Gavilanes, J. G., de Buitrago, G. G., Perez-Castells, R. & Rodriguez, R. (1982) *J. Biol. Chem.* **257,** 10267–10270.
3. Watson, D. C., Levy, W. B. & Dixon, G. H. (1978) *Nature (London)* **276,** 196–198.
4. Schlesinger, D. H., Goldstein, G. & Niall, H. D. (1975) *Biochemistry* **14,** 2214–2218.
5. Schlesinger, D. H. & Goldstein, G. (1975) *Nature (London)* **255,** 423–424.
6. Goldknopf, I. L. & Busch, H. (1980) *Biochem. Biophys. Res. Commun.* **96,** 1724–1731.
7. Wilkinson, K. D. & Audhya, T. K. (1981) *J. Biol. Chem.* **256,** 9235–9241.
8. Busch, H. & Goldknopf, I. L. (1981) *Mol. Cell. Biochem.* **40,** 173–187.
9. Kagan, W. A., O'Neill, G. J., Incefy, G. S., Goldstein, G. & Good, R. A. (1977) *Blood* **50,** 275–288.
10. Kagan, W. A., Siegal, F. P., Gupta, S., Goldstein, G. & Good, R. A. (1979) *J. Immunol.* **122,** 686–691.
11. Low, T. L. K., Thurman, G. B., McAdoo, M., McClure, J., Rossio, J. L., Naylor, P. H. & Goldstein, A. L. (1979) *J. Biol. Chem.* **254,** 981–986.
12. Low, T. L. K. & Goldstein, A. L. (1979) *J. Biol. Chem.* **254,** 987–995.
13. Hunt, L. T. & Dayhoff, M. O. (1977) *Biochem. Biophys. Res. Commun.* **74,** 650–655.
14. West, M. H. P. & Bonner, W. M. (1980) *Biochemistry* **19,** 3238–3245.
15. West, M. H. P. & Bonner, W. M. (1980) *Nucleic Acids Res.* **8,** 4671–4680.
16. Levinger, L. & Varshavsky, A. (1982) *Cell* **28,** 375–385.
17. Hershko, A. & Ciechanover, A. (1982) *Annu. Rev. Biochem.* **51,** 335–364.
18. Ciechanover, A., Finley, D. & Varshavsky, A. (1984) *J. Cell. Biochem.* **24,** 27–53.
19. Finley, D., Ciechanover, A. & Varshavsky, A. (1984) *Cell* **37,** 43–55.
20. Ciechanover, A., Finley, D. & Varshavsky, A. (1984) *Cell* **37,** 57–66.
21. Hershko, A., Eytan, E., Ciechanover, A. & Haas, A. L. (1982) *J. Biol. Chem.* **257,** 13964–13970.
22. Chin, D. T., Kuehl, L. & Rechsteiner, M. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5857–5861.
23. Lenkinski, R. E., Chen, D. M., Glickson, J. D. & Goldstein, G. (1977) *Biochim. Biophys. Acta* **494,** 126–130.
24. Cary, P. D., King, D. S., Crane-Robinson, C., Bradbury, E. M., Rabbani, A., Goodwin, G. H. & Johns, E. W. (1980) *Eur. J. Biochem.* **112,** 577–580.
25. Jenson, J., Goldstein, G. & Breslow, E. (1980) *Biochim. Biophys. Acta* **624,** 378–385.
26. Cook, W. J., Suddath, F. L., Bugg, C. E. & Goldstein, G. (1979) *J. Mol. Biol.* **130,** 353–355.
27. North, A. C. T., Phillips, D. C. & Mathews, F. S. (1968) *Acta Crystallogr. Sect. A* **24,** 351–359.
28. Tickle, I. J. (1975) *Acta Crystallogr. Sect. B* **31,** 329–331.
29. Lehman, M. S. & Larsen, F. K. (1974) *Acta Crystallogr. Sect. A* **30,** 580–582.
30. Blundell, T. L. & Johnson, L. N. (1976) *Protein Crystallography*, eds. Horecker, B., Kaplan, N. O., Marmur, J. & Scheraga, H. A. (Academic, New York).
31. Jones, T. A. (1979) *J. Appl. Crystallogr.* **11,** 268–272.
32. Johnson, C. K. (1965) *ORTEP, A Fortran Thermal Ellipsoid Plot Program for Crystal Structure Illustrations*, Report ORNL-3794, revised (Oak Ridge National Laboratory, Oak Ridge, TN).
33. Lesk, A. M. & Hardman, K. D. (1982) *Science* **216,** 539–540.