Protein Structures

Total Chemical Synthesis and X-ray Crystal Structure of a Protein Diastereomer: [D-Gln35]Ubiquitin**

*Duhee Bang, George I. Makhatadze, Valentina Tereshko, Anthony A. Kossiakoff, and Stephen B. Kent**

Proteins found in nature contain polypeptide chains made up of L-amino acids and glycine. The sequence of amino acids in the polypeptide chain defines the folded tertiary structure of the protein molecule, and the protein owes its biological activity to that folded structure. Non-natural amino acids have been used to construct proteins that are not found in nature to investigate the role of amino acid chemical structure in the formation and stability of the folded protein molecule.[1] In particular, substitution of D-amino acid residues into the protein molecule has been explored for its effects on the binding of insulin to insulin receptors.[2] More recently, D-amino acid substitution has been proposed for enhancing protein stability,[3a] and the structure–function relationships in the K^+ ion channel selectivity filter have been explored by incorporation of a D-Ala residue in place of a Gly residue.[3b] However, despite their potential importance for investigating protein structure–function relationships, there is little detailed knowledge of how the incorporation of a D-amino acid residue affects the local and global conformations that define the folding, stability, and function of the protein molecule.

We set out to understand how a native protein would incorporate a D-amino acid residue into its overall architecture. To definitively explore the perturbation of the local and global conformation in a protein molecule that results from incorporation of D-amino acids, we synthesized and determined the X-ray crystal structure of the chemically engineered globular protein ubiquitin (76 amino acids) with a D-amino acid residue in place of the Gly35 residue in the C-cap region of an α helix. This residue was targeted because the conformation of the Gly35 residue ($\phi = 81°$, $\varphi = 5°$) in native ubiquitin[4] is only allowed for an L-amino acid in a left-handed α helix or for a D-amino acid residue. Herein, we report: 1) an efficient strategy for the total chemical synthesis of ubiquitin; 2) the direct observation of the conservation of amino acid configuration after protein desulfurization by Raney-nickel reduction (that is, L-Cys→L-Ala); 3) high-resolution crystal structures for native ubiquitin and for the ubiquitin protein diastereomer [D-Gln35]ubiquitin; and 4) a striking similarity between the molecular structures of native ubiquitin and the ubiquitin protein diastereomer. The significance of these results is also discussed.

Our research focuses on the investigation of protein folding and stability by using chemical techniques, so we set out to establish an efficient total chemical synthesis of the model protein ubiquitin and its various analogues.[5] We sought to use the recently developed one-pot ligation method[9] to covalently assemble three unprotected peptide segments. However, the ubiquitin molecule does not have the cysteine residues that are needed for native chemical ligation[10] (the amino acid sequence of ubiquitin is shown in Scheme 1a). We noted that human erythrocytic ubiquitin has two alanine residues, which are in positions 28 and 46 and so are suitably located for use as ligation sites (Swiss-Prot accession number P62988).[4] Thus, we adopted a protein-desulfurization strategy[11] that enables the use of native chemical ligation at Cys28 and cys46, after which the cysteine residues are converted into the native Ala28 and Ala46 residues. Our synthetic strategy is shown in Scheme 1b.

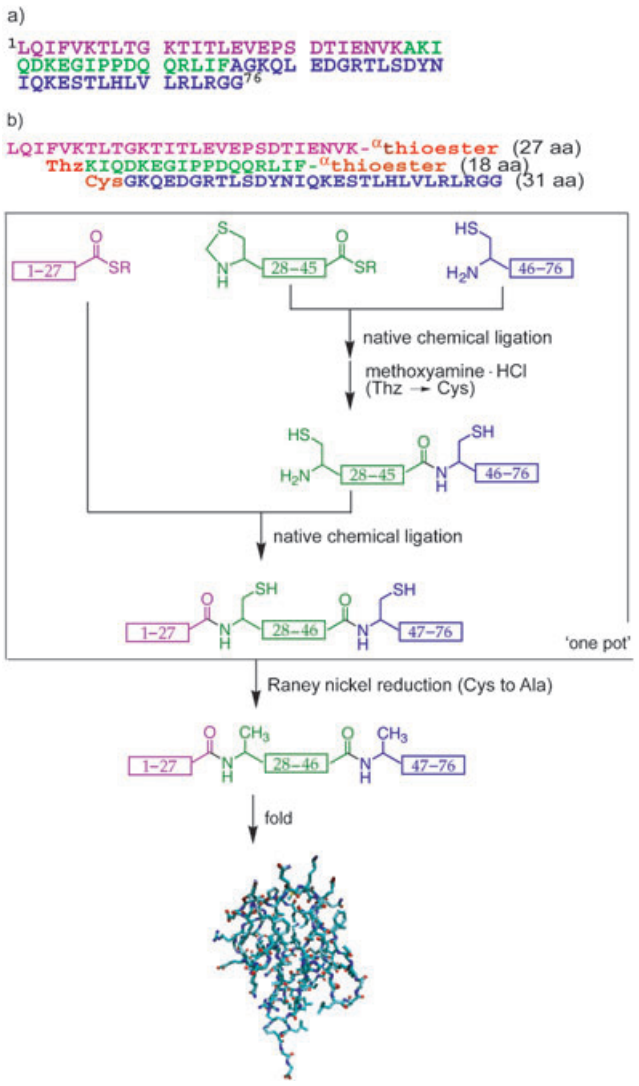
Data for the synthesis of native ubiquitin are shown in Figure 1. The C-terminal peptide and the peptide-$^{\alpha}$thioesters were prepared by solid-phase peptide synthesis by using manual stepwise Boc-chemistry (Boc = butoxycarbonyl) "in situ neutralization" protocols.[12] Several hundred milligrams of high-purity peptide were obtained by preparative reverse-phase HPLC from each 0.4-mmol-scale synthesis. The recovered peptide yields were approximately 30% for the (1–27)-$^{\alpha}$thioester and (Cys^{46}–76) peptides and approximately 45% for the (Thz^{28}–45)-$^{\alpha}$thioester (Thz = 1,3-thiazolidine-4-carboxo group) sequences, based on the starting aminoacyl resin. These three peptides were used as the substrates for the one-pot three-segment ligation on a tens-of-milligrams scale:

[*] D. Bang, V. Tereshko, Prof. A. A. Kossiakoff, Prof. S. B. Kent
Institute for Biophysical Dynamics
The University of Chicago
Chicago, IL 60637 (USA)
Fax: (+1) 773-702-0439
E-mail: skent@uchicago.edu
V. Tereshko, Prof. A. A. Kossiakoff, Prof. S. B. Kent
Department of Biochemistry and Molecular Biology
The University of Chicago
Chicago, IL 60637 (USA)
D. Bang, Prof. S. B. Kent
Department of Chemistry
The University of Chicago
Chicago, IL 60637 (USA)
Prof. G. I. Makhatadze
Department of Biochemistry and Molecular Biology
Pennsylvania State University
Hershey, PA 17033 (USA)

[**] D.B. was supported by funding from the NSF MRSEC program at the University of Chicago (grant DMR 0213745). We acknowledge funding from DOE GTL program grant no. DE-FG02–04ER63786 (to S.B.K) and NIH grant no. GM54537 (to G.I.M). Portions of this work were performed at the DND-CAT (sector 5-ID) and IMCA-CAT (sector 17-ID) of the Advanced Photon Source. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31–109-Eng-38. Use of the DND-CAT beamline is supported by the E.I. DuPont de Nemours, The Dow Chemical Company, the US National Science Foundation through grant no. DMR-9304725, and the State of Illinois through the Department of Commerce and the Board of Higher Education Grant IBHE HECA NWU 96. Use of the IMCA-CAT beamline at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with the Illinois Institute of Technology.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

  DOI: 10.1002/ange.200463040



Scheme 1. a) Target amino acid sequence for human erythrocytic [Met 1 Leu]ubiquitin. b) Peptide segments and strategy used for the total chemical synthesis of ubiquitin.

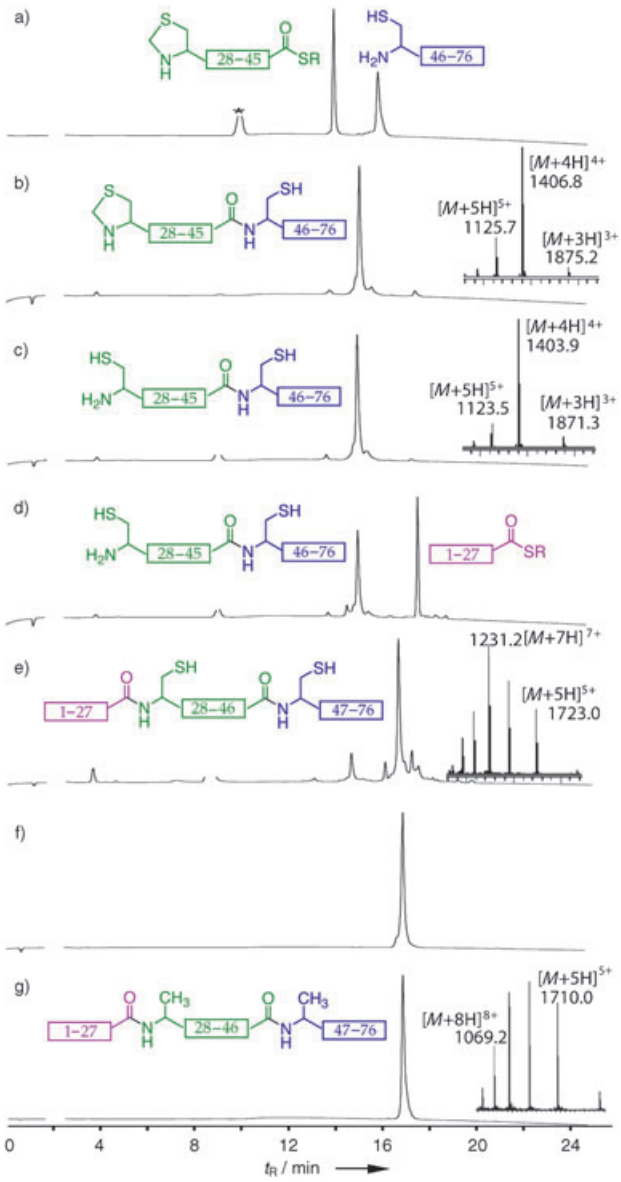


Figure 1. Data for the synthesis of native ubiquitin. Reactions were monitored by LC–MS. The UV profile at 214 nm is shown, together with ESIMS data (inset) corresponding to each major product. The chromatographic separations were performed using a linear gradient (10–50%) of buffer B in buffer A over 20 min (buffer A = 0.1% trifluoroacetic acid (TFA) in water; buffer B = 0.08% TFA in acetonitrile). a) $t = 0$ h, b) $t = 12$ h, c) $t = 16$ h, d) $t = 17$ h, e) $t = 40$ h, f) purified, and g) desulfurized and purified. Observed and calculated masses of the synthetic intermediates and final products were: b) observed mass (obs) 5623.1 ± 0.5 Da, calculated mass using average isotopes (calcd) 5623.4 Da; c) obs = 5611.6 ± 0.7 Da, calcd = 5611.4 Da; e) obs = 8610.2 ± 1.0 Da, calcd = 8610.9 Da; g) obs = 8546.0 ± 1.1 Da, calcd = 8546.8 Da.

the first ligation of the (Thz^{28}–45)-$^{\alpha}$thioester and (Cys^{46}–76) peptides was performed in phosphate buffer (100 mM, pH 6.8) containing guanidinium hydrochloride (6 M) and thiophenol (1% (v/v)) at a concentration of 3 mg mL^{-1} for each peptide (Figure 1a,b). Methoxyamine·hydrochloride (0.2 M) was added to convert the Thz peptide into the Cys peptide ligation product at pH 4.0 (Figure 1c). Readjustment of the solution to pH 6.8 was followed by addition of the third peptide segment, (1–27)-$^{\alpha}$thioester, to effect the second ligation (Figure 1d,e). The ligated full-length 76 residue polypeptide was purified by reverse-phase HPLC (Figure 1f) and lyophilized for the subsequent Raney-nickel desulfurization reaction.

Desulfurization with Raney nickel has been applied to proteins to convert cysteine residues into alanine residues.[13] No change in the absolute configuration at the α carbon atoms of the cysteine/alanine residues was observed in model studies on acylated cysteine.[14,15] Yan and Dawson used Raney-nickel reduction subsequent to native chemical ligation at Cys to extend the use of chemical ligation to the synthesis of proteins that do not contain Cys.[11] One disdavantage of the Raney-nickel reduction is that desulfurization not only affects the cysteine residues but also results in the demethylthiolation of the methionine side chain (although at a slower rate).[11] The first residue of ubiquitin

is a methionine, so for synthetic convenience, we replaced Met 1 in ubiquitin with Leu 1 (Scheme 1 a).[16]

We prepared the Raney nickel as described:[11] Nickel acetate (600 mg) was dissolved in distilled water (15 mL) and then sodium borohydride (100 mg) was added slowly. The black nickel precipitate was collected by filtration and then washed extensively with distilled water. We carried out the conversion of Cys 28 and Cys 46 into Ala 28 and Ala 46 in the synthetic polypeptide by adding the freshly prepared nickel precipitate to the peptide solution (ca. 20 mg of purified 76 amino acid polypeptide at a concentration of 3 mg of lyophilized peptide per 1 mL of 0.1 M phosphate buffer containing 6 M guanidinium hydrochloride). We observed that oxidized cysteine residues retarded the rate of the reaction (the same observation has been made previously[13]). We added TCEP (tris(2-carboxyethyl)phosphane hydrochloride; 10 mM) to our reaction mixture to insure that the two cysteine residues were fully reduced. The final pH value of the reaction mixture was 6.0. The Raney-nickel reduction was complete in 12 hours and a mass decrease of 64 Da was observed, as expected for the loss of two sulfur atoms (Figure 1 g).

Essentially identical results were obtained for the total chemical synthesis of the [D-Gln 35]ubiquitin protein molecule (Figure 2). The folded structure of chemically synthesized ubiquitin was confirmed by dissolution of the lyophilized polypeptide under native conditions and the procurement of circular dichroism spectra, which were consistent with the native protein structure (data not shown). The folded protein was obtained at the high concentrations necessary for crystallization by dissolution of the the synthetic polypeptide (20 mg mL^{-1}) in phosphate buffer containing guanidinium hydrochloride (6 M) and was then dialyzed against distilled water.

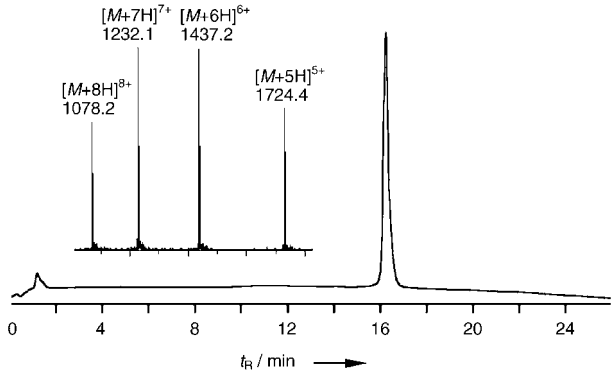


Figure 2. HPLC profile of purified [D-Gln 35]ubiquitin at 214 nm and ESIMS data (inset) corresponding to the product peak (observed mass = 8617.4 ± 0.5 Da; calculated mass = 8617.9 Da (average isotopes)).

We performed several syntheses of both the native ubiquitin and the [D-Gln 35]ubiquitin diastereomer to produce enough protein material for the screening of crystallization conditions and for the optimization of the growth of high-quality crystals. An average yield of recovered product of 36 % from the overall preparation of the full-length polypeptide containing Cys 28 and Cys 46 was obtained using the protocols described above (Figure 1 a–f). The yields of recovered product from the desulfurization reaction were 75 % on average (Figure 1 f–g). Ten- to twenty-milligram quantities of wild-type ubiquitin and its diastereomer were produced in each synthesis. The average synthetic yields were 27 % based on the peptides used. Therefore, excellent yields of high-quality protein constructs were obtained from the synthetic strategy reported herein, and furthermore we could apply this multisegment strategy to the ready synthesis of analogues. Thus, we could reuse the peptides (1–27)-$^{\alpha}$thioester and (Cys^{46}–76) and it was only necessary to remake the short [Gly 35D-Gln](Thz^{28}–45)-$^{\alpha}$thioester for the synthesis of the diastereomer analogue protein.

At first, we considered a protein NMR experiment to investigate the local and global structure of the ubiquitin diastereomer. However, we concluded that high-resolution crystallography would be better suited to explore even small configuration changes caused by the incorporation of D-Gln 35 into the protein molecule and to definitively show the local and global structural changes of a protein diastereomer compared to the wild-type protein molecule. Ubiquitin is reported to be difficult to crystallize.[4,17–19] Attempts to form ubiquitin crystals under the conditions used in the first successful crystallization[17] have since failed.[4,17–19] Instead, most of the crystals of sufficient quality for X-ray crystallographic studies have been produced by seeding with samples of the original crystal.

We carried out extensive screening of crystallization conditions for ubiquitin, but without seeding we were unable to obtain crystals. However, we were aware that solid-state NMR spectroscopy researchers were using ubiquitin nanocrystals to study protein structures.[20] The conditions reported for the production of ubiquitin nanocrystals used cadmium salts not usually included in commercially available crystal screening kits. By using cadmium salts in combination with the crystal screening kits, we were able to grow two different forms of ubiquitin crystals under different conditions. Subsequent rounds of optimization led to crystals of both the wild-type ubiquitin and its diastereomer that were of sufficient quality for X-ray crystallographic studies. Crystal formation could be reproducibly obtained by mixing ubiquitin solution (2 μL, 20 mg mL^{-1}) and $CdCl_2$ solution (2 μL, 50 mM), and poly(ethylene glycol) 2000 monomethyl ether (PEG MME; 20 %, *w*/*v*) in bis–tris buffer (bis(2-hydroxyethyl)iminotris(hydroxymethyl)methane; 0.1 M, pH 6.5).

X-ray diffraction data were collected using synchrotron radiation at the Advanced Photon Source at the Argonne National Laboratory. Cubic ($P4_332$) and orthorhombic ($P2_12_12_1$) space groups were identified for both the wild-type and diastereomer crystals. In particular, the orthorhombic crystals were highly ordered and diffracted well to 1.3 Å (a complete data set was obtained to only 1.5 Å because of the geometric limitations of the CCD detector). The structures of the native ubiquitin and the diastereomer protein were solved by the molecular replacement method, in which the ubiquitin structure PDB accession number 1UBQ was used as the test model.[21] Subsequent refinement procedures were carried out

using the CNS and refmac refinement packages (see the Supporting Information for data collection and refinement statistics for both native ubiquitin and [D-Gln35]ubiquitin).

The synthetic native ubiquitin protein had the same molecular structure as that previously reported.[4] The stereochemical configuration of the D-Gln35 residue was directly confirmed by the high-resolution electron-density map of the ubiquitin diastereomer, in which the side chain of D-Gln35 could be seen pointing in the opposite direction to that expected for the side chain of an L-Gln residue (Figure 3). Thus, the ubiquitin diastereomer can incorporate D-Gln35 into its overall fold without perturbing the global architecture of the protein molecule (Figure 4). Both Gly35 and D-Gln35 were in a left-handed conformation ($\phi = 80°$ for Gly35 and $\phi = 77°$ for Gln35; see the Supporting Information).

Comparison of the wild-type and diastereomer structures showed a striking similarity in both the local and global conformations (Figure 4). The root-mean-square difference (rmsd) value for the comparison of the Cα atoms at corresponding positions in the two protein structures was 0.39 Å, and the rmsd value for all corresponding main-chain atoms in the local region (from residue 31 to 37) near the D-Gln35 substitution was 0.24 Å (see the Supporting Information). These values are within the experimental variation expected for two proteins of identical structure at the resolution of the analysis. This striking similarity 1) validates the replacement of natural amino acid residues occupying left-handed conformations in native protein molecules by D-amino acids and 2) provides experimental support for the assumption that replacement of a Gly residue occupying a left-handed conformation by a D-amino acid does not alter the local and global conformations of the protein molecule.

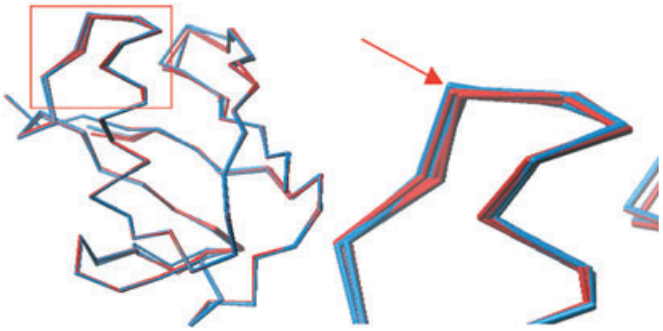


Figure 4. Superposition of the molecular structures of the three native and three [D-Gln35]ubiquitin protein molecules found in the unit cell for the two different proteins. Left: the backbone of Cα-atom traces of the two proteins are shown superimposed; [D-Gln35]ubiquitin (blue); native ubiquitin (red). Right: close up of nine residues near D-Gln35/Gly35 (arrow).

The conservation of the L configuration of Ala28 and Ala46 after the desulfurization reaction could be directly observed in the electron-density map of the synthetic proteins (Figure 5). The conservation of the L configuration during the Raney-nickel reduction further validates the synthetic strategy that Yan and Dawson devised to prepare cysteine-free proteins by using native chemical ligation followed by a desulfurization reaction.[11]

In summary, we have developed an efficient route to cysteine-free ubiquitin molecules by combining one-pot three-peptide-segment native chemical ligation[9] and protein desulfurization[11] for chemical protein synthesis to produce large quantities of high-quality protein materials for X-ray crystallographic studies. We have also verified a striking

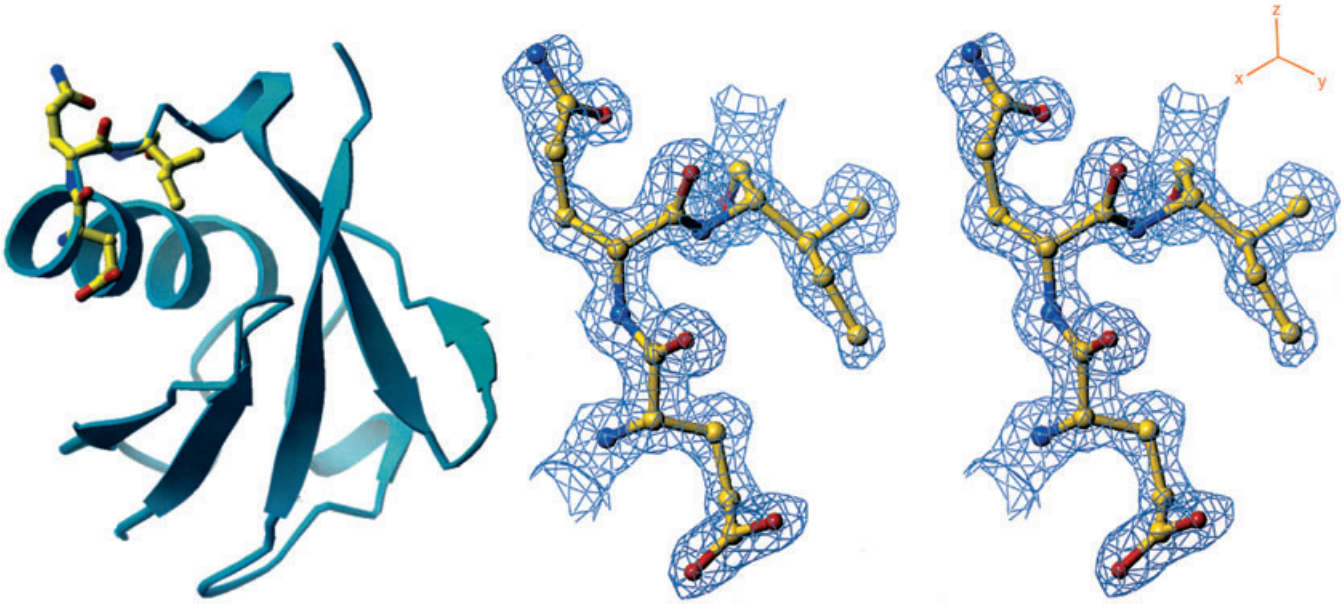


Figure 3. Left: structure of the [D-Gln35]ubiquitin diastereomer; the sequence Glu34–D-Gln35–Ile36 is highlighted in yellow. Right: stereoview of the local conformation near the D-Gln35 residue; the residues Glu34–D-Gln35–Ile36 are shown fitted to the $2F_o - F_c$ electron-density map contoured at 1σ. (It was not possible to unambiguously differentiate the oxygen and nitrogen atoms in the carboxamide side chain at a resolution of 1.5 Å; therefore, the assignment shown is arbitrary.)

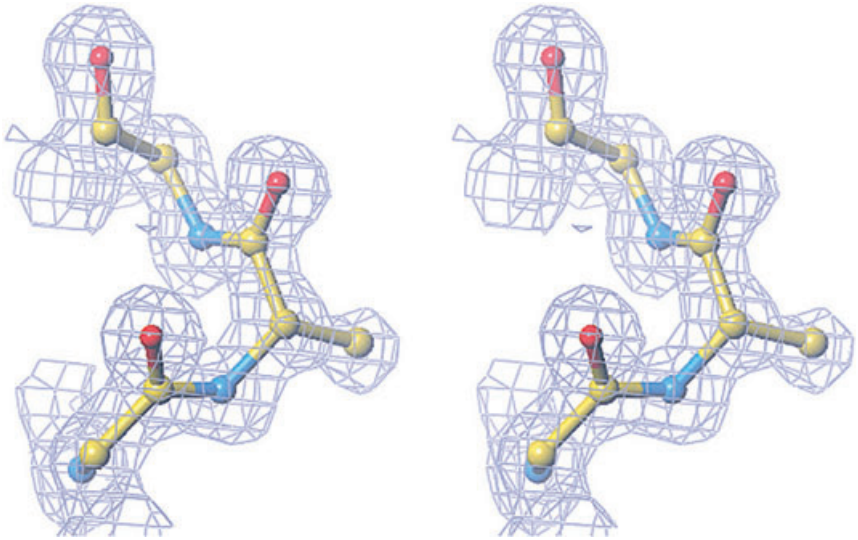


Figure 5. Stereoview showing Ala 46 fitted to the $2F_o-F_c$ electron-density map contoured at 1σ. The conservation of configuration in the desulfurization reaction (L-Cys→L-Ala) is apparent. Occupancy and thermal factor refinement indicated that the Cβ atoms showed full occupancy for the L conformation of the alanine side chain, within experimental uncertainty (<10%), with an accompanying low thermal factor ($B=3$ Å^2).

conservation of molecular structure between the folded conformations of a protein diastereomer and the corresponding wild-type protein by using high-resolution X-ray crystallography. This study provides experimental support that replacement of a native Gly residue that occupies a left-handed conformation by a D-amino acid does not significantly perturb the local or global conformations of a protein molecule. We are currently using the ubiquitin synthesis and crystallographic methods reported herein, in combination with other modern biophysical techniques, to systematically dissect the molecular basis of stability in this model protein.

Received: December 23, 2004
Published online: April 18, 2005

Keywords: chemical ligation · diastereomers · protein structures · protein synthesis · X-ray diffraction

[1] a) A. Wlodawer, M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, S. B. H. Kent, *Science* **1989**, *245*, 616–621; b) M. Baca. S. B. H. Kent, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11638–11642; c) M. Baca, S. B. H. Kent, *Tetrahedron* **2000**, *56*, 9503–9513; d) G. Simmons, P. R. Clapham, L. Picard, R. E. Offord, M. M. Rosenkilde, T. W. Schwartz, R. Buser, T. N. Wells, A. E. Proudfoot, *Science* **1997**, *276*, 276–279; e) Z. Zhang, J. Gildersleeve, Y. Y. Yang, R. Xu, J. A. Loo, X. Uryu, C. H. Wong, P. G. Schultz, *Science* **2004**, *303*, 371–373.

[2] M. Kobayashi, S. Ohgaku, M. Iwasaki, H. Maegawa, Y. Shigeta, K. Inouye, *Biochem. Biophys. Res. Commun.* **1982**, *107*, 329–336.

[3] a) B. Anil, B. Song, Y. Tang, R. D. P. Raleigh, *J. Am. Chem. Soc.* **2004**, *126*, 13194–13195; b) F. I. Valiyaveetil, M. Sekedat, R. MacKinnon, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17045–17049.

[4] S. Vijay-Kumar, C. E. Bugg, W. J. Cook, *J. Mol. Biol.* **1987**, *194*, 531–544.

[5] The synthesis of crystalline ubiquitin by stepwise solid-phase methods has been previously reported; see Refs. [6]–[8].

[6] R. Ramage, O. Ogunjobi, J. Green, *Tetrahedron Lett.* **1989**, *30*, 2149–2152.

[7] O. Ogunjobi, R. Ramage, *Biochem. Soc. Trans.* **1990**, *18*, 1322–1323.

[8] R. Ramage, J. Green, T. W. Muir, O. M. Ogunjobi, S. Love, K. Shaw, *Biochem. J.* **1994**, *299*, 151–158.

[9] D. Bang, S. B. Kent, *Angew. Chem.* **2004**, *116*, 2588–2592; *Angew. Chem. Int. Ed.* **2004**, *43*, 2534–2538.

[10] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776–779.

[11] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526–533.

[12] M. Schnolzer, P. Alewood, A. Jones, D. Alewood, S. B. Kent, *Int. J. Pept. Protein Res.* **1992**, *40*, 180–193.

[13] M. T. Perlstein, M. Z. Atassi, S. H. Cheng, *Biochim. Biophys. Acta* **1971**, *236*, 174–182.

[14] G. S. Fonken, R. Mozingo, *J. Am. Chem. Soc.* **1947**, *69*, 1212–1213.

[15] R. Mozingo, D. E. Wolf, S. A. Harris, K. Fokers, *J. Am. Chem. Soc.* **1943**, *65*, 1013–1016.

[16] Leu1 ubiquitin has full biological activity: C. G. Benitez-Cardoza, K. Stott, M. Hirshberg, H. M. Went, D. N. Woolfson, S. E. Jackson, *Biochemistry* **2004**, *43*, 5195–5203.

[17] S. Vijay-Kumar, C. E. Bugg, K. D. Wilkinson, W. J. Cook, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 3582–3585.

[18] S. Vijay-Kumar, C. E. Bugg, K. D. Wilkinson, R. D. Vierstra, P. M. Hatfield, W. J. Cook, *J. Biol. Chem.* **1987**, *262*, 6396–6399.

[19] D. Alexeev, S. M. Bury, M. A. Turner, O. M. Ogunjobi, T. W. Muir, R. Ramage, L. Sawyer, *Biochem. J.* **1994**, *299*, 159–163.

[20] R. W. Martin, K. W. Zilm, *J. Magn. Reson.* **2003**, *165*, 162–174.

[21] The PDB accession numbers for the crystal structures are: 1YIW for wild type ubiquitin and 1J1 for [D-Gln35]ubiquitin.

