Structural Modification of a Periodic Polypeptide through Biosynthetic Replacement of Proline with Azetidine-2-carboxylic Acid

Timothy J. Deming,[†,§] Maurille J. Fournier,[‡] Thomas L. Mason,[‡] and David A. Tirrell*[,†]

Departments of Polymer Science and Engineering, and Biochemistry and Molecular Biology, University of Massachusetts at Amherst, Amherst, Massachusetts 01003

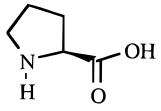
Received July 24, 1995; Revised Manuscript Received November 29, 1995[⊗]

ABSTRACT: Repetitive polypeptides comprising 16 repeats of the sequence −(AlaGly)$_3$ProGluGly− **(1a)** have been prepared from *Escherichia coli* as overexpressed recombinant proteins. Partial *in vivo* replacement of the proline (Pro) residues in sequence **1a** with L-azetidine-2-carboxylic acid (Aze) was achieved by expression of the target protein in medium containing Aze and lacking Pro. NMR and amino acid analysis for residual proline in the polymer indicated 25−40% replacement of Pro by Aze. While polymers of **1a** form conformationally disordered solids, incorporation of Aze allows this material to adopt a β-sheet structure in the solid state.

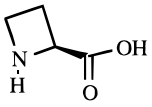
In an effort to prepare polymeric materials of precisely defined sequence, molecular weight, and composition, we have been exploring the use of recombinant DNA methods to produce copolypeptides consisting of periodic repeats.[1] In many cases, these polymers have been designed to contain secondary structural elements that direct (or allow) the chains to fold and pack into predetermined regular solid-state conformations. One such periodic polymer, containing multiple repeats of sequence **1a**, was designed to fold into lamellar crystals comprised of −(AlaGly)$_3$− β-strands and −ProGluGly− turn elements, but was found to be reluctant to crystallize.[2] The conformationally restricted proline (Pro, **2**) residue may prevent alignment and subsequent H-bonding of the β-strands either through improper orientation of chain trajectories or simply because of its steric bulk. If so, replacement of Pro with a smaller residue with different conformational properties might allow the polymer to order more effectively. For this reason we sought to replace Pro in sequence **1a** with its lower homolog, L-azetidine-2-carboxylic acid (Aze, **3**), to yield sequence **1b**.

-(AlaGly)$_3$XxxGluGly-

1a, Xxx = Pro
1b, Xxx = Aze



2



3

Aze has been shown to be incorporated *in vivo* and *in vitro* into the cellular proteins of *Escherichia coli*[3] as well as in chick embryos for studies on collagen biosynthesis.[4] Incorporation of this analog into proteins can result in marked effects on properties;[5a] for example, replacement of Pro in collagen with as little as 4% of Aze is sufficient to destabilize triple-helix bundle formation.[5b] These results have prompted several computational and experimental studies of Aze and its polymers and oligomers.[6] The calculations of Scheraga and co-workers[6a] suggest that peptides containing Aze are more flexible than the corresponding sequences containing Pro and that the propensity for β-bend formation is higher for GlyAze (as in **1b**) than for GlyPro (as in **1a**) dipeptides. Both features would be expected to relieve the constraints that frustrate ordering processes in polymers of **1a** and to facilitate the assembly of β-sheets in polymers of **1b**.

In order to obtain high levels of analog incorporation, it is beneficial to utilize a host strain which cannot produce Pro, i.e., a Pro auxotroph,[7] so that the cells and media can be depleted of the natural amino acid. Since Aze does not support the growth of *E. coli*,[8] it is necessary to grow cells first in medium supplemented with Pro and then to induce expression of the target protein in medium containing Aze.

Proteins containing sequences **1a** and **1b** were synthesized as described previously,[2] using *E. coli* host strain BL21(DE3)pLysS pET3-8 Pro$^-$ in growth media containing either Pro or Aze (Figure 1). The target protein migrates at a molecular weight of about 38 000, which is significantly higher than the predicted value (16 900). This anomalous migration behavior is consistent with previous observations and is not a result of genetic instability.[1,2] From the autoradiogram, it can be seen that protein accumulates to significantly higher levels in Aze medium than in the proline-free control, consistent with incorporation of Aze into the polypeptide.

Isolable quantities of polymers of **1a** were prepared by growth of the host strain in rich medium (2× YT)[9] followed by expression of the target sequence by addition of isopropyl β-D-thiogalactopyranoside (IPTG). The polymer was expressed as a fusion protein of overall sequence, MASMTGGQQMGRDPMFKYSRDPMG[AG-AGAGPEG]$_{16}$ARMHIRPGRYQLDPAANKARKEAELA-AATAEQ; the segments flanking the repetitive portion of the chain were encoded by the pET3b and p937.51 vectors used for cloning and expression.[2] For the synthesis of polymers of **1b**, cells were grown in rich medium and then shifted to medium containing Aze followed by induction with IPTG. The recombinant

§ Current address: Department of Materials, University of California Santa Barbara, Santa Barbara, CA 93106.
† Department of Polymer Science and Engineering.
‡ Department of Biochemistry and Molecular Biology.
⊗ Abstract published in *Advance ACS Abstracts,* February 1, 1996.

0024-9297/96/2229-1442$12.00/0

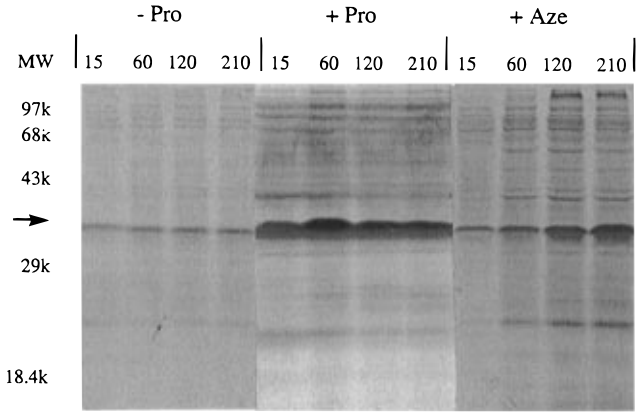


Figure 1. SDS-PAGE analysis of whole cell lysates visualized by [^{35}S]Met labeling. A single colony of BL21(DE3)pLysS pET3-8 Pro$^-$ was used to inoculate M9AA minimal medium (30 mL of M9 medium[9] containing the 20 natural amino acids at 20 μg/mL and supplemented with 0.1 mM $CaCl_2$, 1.0 mM $MgSO_4$, 25 μg/mL chloramphenicol, and 200 μg/mL ampicillin) which was shaken at 37 °C until OD(600 nm) = 0.75. IPTG (final concentration = 0.4 mM) was added to the mixture and shaking continued for 10 min. Cells were isolated by centrifugation of two-thirds of the mixture (20 mL), and the pellet was washed with M9AA minimal medium without proline. The cells were resuspended in M9AA minimal medium without proline (20 mL). To the remaining one-third of original culture (10 mL) was added a solution of [^{35}S]Met (20 μL, 300 μCi), and to the sample of washed culture (20 mL) was also added [^{35}S]Met (40 μL, 600 μCi). The 20 mL culture was divided into two equal portions. To one portion was added Aze (200 μg of the L-isomer as a 1 mg/mL aqueous solution) and to the other portion nothing was added. All three samples were shaken at 37 °C for 3.5 h and aliquots (1.5 mL) were removed at the time intervals of 15, 60, 120, and 210 min after addition of the analog. Cells were isolated from each aliquot by centrifugation and the supernatants were discarded. Cells were suspended in M9AA minimal medium without proline (1.0 mL), centrifuged again, and lysed, and the proteins were fractionated on a 12% polyacrylamide gel. The protein fractions were electrophoresed at 12 mA constant current overnight and then stained with Coomassie Brilliant Blue G-250. After destaining with 45% MeOH/10% HOAc, the gels were dried and then exposed to X-ray film for 2 days at −80 °C. Lanes 1−4 contain target protein from cells shifted to medium lacking proline; lanes 5−8 contain target protein from cells incubated in medium containing Pro; and lanes 9−12 contain target protein from cells shifted to medium containing Aze. The target protein is indicated by an arrow. Molecular weight standards are listed to the left of the autoradiogram: the apparent molecular weights of **1a** and **1b** were estimated from the size standards.

proteins partitioned into the soluble portion of the cell lysate and were isolated by lowering the pH of the lysate to pH = 4.0 and then precipitating the target protein with increasing concentrations of $(NH_4)_2SO_4$. Protein containing **1b** was found to precipitate in the range of 0−40% $(NH_4)_2SO_4$ saturation at ambient temperature, while the polymer of **1a** was found to precipitate in the range of 40−60% $(NH_4)_2SO_4$ saturation. Removal of the fusion segments with CNBr gave the target polypeptides, which were purified in overall yields of 30 mg/L (Pro medium) and 5 mg/L (Aze medium), which represent approximately 20 and 3% of total cellular protein, respectively, at the cell densities used in these experiments.

The relative amounts of Pro and Aze in the polypeptides were first determined using amino acid compositional analyses (Table 1). Aze is degraded under the conditions used to hydrolyze and analyze peptides (6N HCl or CH_3SO_3H) and so cannot be measured directly using this technique.[10] However, diminution in the level

Table 1. Amino Acid Compositions of Proteins Prepared in Pro and Aze Media

amino acid	mol % (theor)	mol % (obsd) Pro	mol % (obsd) Aze[b]
glycine	34.5	35.3	33.8
alanine	29.0	27.9	28.1
glutamic acid	9.5	10.6	10.9
proline	10.0[a]	10.1	7.4

[a] Expected for **1a**. [b] Aze decomposes under conditions for peptide cleavage and could not be determined directly.

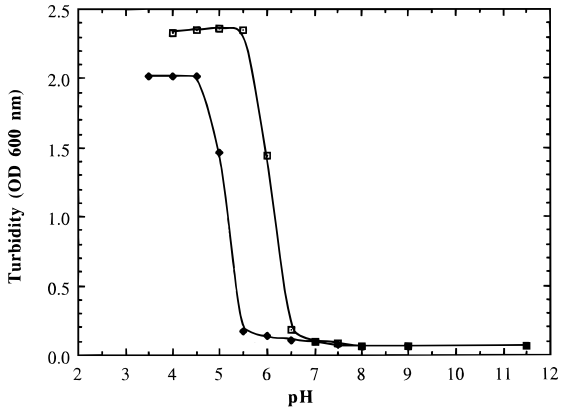


Figure 2. Cloud point determination for polymers of **1a** and **1b** in distilled water at ambient temperature. Black diamonds represent data for the protein containing **1a**, and open squares represent data for the protein containing **1b**. Polymers were dissolved in H_2O (each was 0.5 mM in 1.0 mL solution) which had been adjusted to pH = 11.5 with NaOH. Turbidity of the solutions was monitored at 600 nm while the pH of both samples was lowered incrementally through addition of 1.0 N HCl. Samples were stirred during addition of the HCl and allowed to equilibrate for 1 min before each measurement.

of Pro in these experiments implies the presence of Aze in its place, and from the data in Table 1, we estimate that the extent of substitution with Aze is approximately 25%. The extent of analog substitution can also be estimated from ^{1}H NMR spectra of the purified polymers. Integration of Pro resonances (3.62 ppm) versus Ala and Gly resonances (1.32 and 4.20 ppm, respectively) placed the amount of analog incorporation at 40%. Electrospray ionization mass spectrometry gave the mass of the polymer containing Aze as 11 125 after CNBr cleavage, compared with an expected mass of 11 122, assuming 33% Aze substitution and pairing of 7 Na^+ ions with the glutamate side chains.[11]

We noted earlier that the proteins containing sequences **1a** and **1b** were observed to precipitate at different concentrations of $(NH_4)_2SO_4$. Further examination revealed differences in pH-dependent solubility as well. Partial acid titration of the glutamate side chains results in precipitation of both polymers (Figure 2); however, partial replacement of Pro by Aze shifts the cloud point by a full unit of pH (from pH 5.5 for polymers of **1a** to pH 6.5 for chains containing **1b**). We propose that transformation of **1a** to **1b** results in changes in the conformational properties of the polymer, as it seems unlikely that the cloud point shift arises from changes in the intrinsic hydrophobic character of the chain units.

To examine the conformational properties of the polymers in the solid state, samples were dissolved in 70% formic acid and stirred under MeOH vapor for several days until they gelled (7 days for both). The gelled samples were washed with MeOH, dried under vacuum, and analyzed by FTIR spectroscopy (Figure 3).

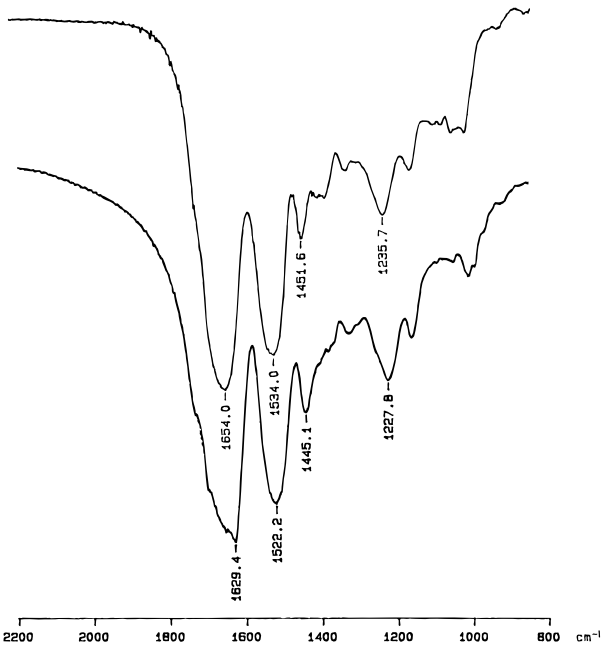


Figure 3. FTIR spectra of the polypeptides containing Pro (top) and Aze (bottom), as KBr pellets.

The amide IR absorptions are sensitive to polypeptide conformation; in particular, the amide I and amide II absorptions should appear at 1630 and 1525 cm^{-1}, respectively, for a polypeptide in the β-form, while those for a disordered polypeptide appear at 1658 and 1535 cm^{-1}. [12] The polymer of **1a** appears to be devoid of regular secondary structure (amide I, 1654 cm^{-1}; amide II, 1534 cm^{-1}), after attempted crystallization, while the amide vibrations at 1629 and 1522 cm^{-1} confirm the β-sheet structure of the polymer containing **1b** (Figure 3).

These results demonstrate that even partial substitution of Aze for Pro in sequence **1a** is sufficient to allow the chain to adopt the β-conformation under conditions where the proline variant is disordered. This observation is consistent with the prediction of Scheraga and co-workers[6a] that Aze increases chain flexibility and the probability of β-bend formation, but elucidation of the causes underlying the conformational ordering of polymers containing **1b** must await further experiments. We find the observed changes remarkable given that 25−40% substitution of Pro by Aze represents modification of only 4−6 (of a total of 148) residues in the chain. Analog substitutions such as this provide a general strategy for the analysis of structure−property relations in both natural and artificial protein materials.

Acknowledgment. This work was supported by Contract No. DAAHO4-93-G-0217 from the U.S. Army Research Office and by a grant (DMR-8914359) from the Polymers and Genetics Programs of the National Science Foundation. T.J.D. acknowledges the National Institutes of Health for a National Research Service Award postdoctoral training fellowship. NMR spectra were recorded in the University of Massachusetts NMR Facility, which is supported in part by the NSF Materials Research Science and Engineering Center at the University. We thank Brian Price and Wendy Petka for assistance with electrophoretic analyses.

References and Notes

(1) (a) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471−5475. (b) Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 536−537. (c) Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427−1432. (d) Parkhe, A. D.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1993**, *26*, 6691−6692. (e) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213−1214. (f) Beavis, R. C.; Chait, B. T.; Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 7584−7585.

(2) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727−733.

(3) (a) Grant, M. M.; Brown, A. S.; Corwin, L. M.; Troxler, R. F.; Franzblau, C. *Biochim. Biophys. Acta* **1975**, *404*, 180−187. (b) Fowden, L.; Richmond, M. H. *Biochim. Biophys. Acta* **1963**, *71*, 459−461. (c) Papas, T. S.; Mehler, A. H. *J. Biol. Chem.* **1970**, *245*, 1588−1595.

(4) (a) Uitto, J.; Prockop, D. J. *Biochim. Biophys. Acta* **1974**, *336*, 234−251. (b) Takeuchi, T.; Prockop, D. J. *Biochim. Biophys. Acta* **1969**, *175*, 142−155.

(5) (a) Mauger, A. B.; Witkop, B. *Chem. Rev.* **1966**, *66*, 47−86. (b) Lane, J. M.; Dehm, P.; Prockop, D. J. *Biochim. Biophys. Acta* **1971**, *236*, 517−527.

(6) (a) Zagari, A.; Nemethy, G.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 951−959. (b) Zagari, A.; Nemethy, G.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 961−966. (c) Zagari, A.; Nemethy, G.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 967−974. (d) Zagari, A.; Palmer, K. A.; Gibson, K. D.; Nemethy, G.; Scheraga, H. A. *Biopolymers* **1994**, *34*, 51−60. (e) Tsai, F. H.; Overberger, C. G.; Zand, R. *Biopolymers* **1990**, *30*, 1039−1049. (f) Tsai, F. H.; Overberger, C. G.; Zand, R. *J. Polym. Sci., A: Polym. Chem.* **1992**, *30*, 551−559. (g) Blessing, R. H.; Smith, G. D. *Acta Crystallogr. B* **1982**, *38*, 1203−1207. (h) Boni, R.; Verdini, A. S. *Macromolecules* **1973**, *6*, 517−520. (i) Boni, R.; DiBlasi, R.; Verdini, A. S. *Macromolecules* **1975**, *8*, 140−145. (j) Boni, R.; DiBlasi, R.; Farina, A.; Verdini, A. S. *Biopolymers* **1976**, *15*, 1233−1242.

(7) (a) Wilson, M. J.; Hatfield, D. L. *Biochim. Biophys. Acta* **1984**, *781*, 205−215. (b) Richmond, M. H. *Bacteriol. Rev.* **1962**, *26*, 398−420. (c) Hortin, G.; Boime, I. *Methods Enzymol.* **1983**, *96*, 777−784.

(8) Peterson, P. J.; Fowden, L. *Nature* **1963**, *200*, 148−151.

(9) Miller, J. H. *Experiments in Molecular Genetics*; Cold Spring Harbor Laboratory Press: Plainview, NY, 1972; p 419.

(10) Fowden, L. *Biochem. J.* **1956**, *64*, 323−332.

(11) The observed mass of the Pro form of the polymer was 11 065, in excellent agreement with the expected mass of 11 065, given binding of one Na^+ counterion.

(12) Bandekar, J.; Krimm, S. In *Advances in Protein Chemistry*; Anfinsen, C. B., Edsall, J. T., Richards, F. M., Eds.; Academic Press: Orlando, FL, 1986; Vol. 38, pp 183−364.

MA9510698