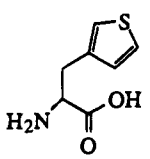
Biosynthesis of a Periodic Protein Containing 3-Thienylalanine: A Step Toward Genetically Engineered Conducting Polymers

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

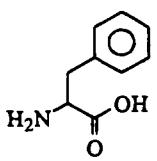
Departments of Biochemistry and Molecular Biology and Polymer Science and Engineering and Program in Molecular and Cellular Biology University of Massachusetts, Amherst, Massachusetts 01003

Received September 26, 1994

Protein engineering provides an approach to materials synthesis in which macromolecular architecture is subject to precise control. Proteins synthesized biologically from a genetic template by the protein synthesis machinery are monodisperse with respect to molecular weight, and they are stereochemically pure, as only L-amino acids are used. Chemical composition and sequence are controlled, as the covalent structure of the chain is encoded by the template messenger RNA. We reasoned that genetic engineering technology could be used to create conducting materials by incorporation of electroactive amino acids, which could subsequently be cross-linked (or grafted) to produce extended conjugated systems. The role of the protein backbone in such materials would be to control the spatial organization of the conducting layers or domains. We report here the successful synthesis of a recombinant protein containing 3-thienylalanine (3-TA, **1**) in place of phenylalanine (**2**).



1



2

The analog was incorporated at multiple positions in a periodic protein predicted to form lamellar crystals comprising regularly folded β sheets. The rationale for the choice of 3-TA is based on its similarity to the 3-alkylthiophenes. Poly(3-alkylthiophene)s are among the best conducting polymers, with conductivities after doping of about 2000 S cm^{-1}.[1] 3-TA may also be useful in electrodeposition of proteins onto electrodes, in the fabrication of enzyme-based sensory elements.

In vivo synthesis of proteins containing selenomethionine has been reported recently by several groups.[2,3] In these studies the host strain of *Escherichia coli* was able to use the analogue in place of methionine to support cell growth. Several analogues that do not support growth are also known to be utilized to some extent by the protein synthesis machinery.[4] In such cases it may be possible to produce proteins containing the analogue by first growing the recombinant cells in media containing the natural amino acid, inducing synthesis of the target protein, and adding the amino acid analogue shortly after induction. Our group has successfully produced a repetitive protein containing *p*-fluorophenylalanine in this manner.[5] We describe herein a method for the synthesis of a similar periodic protein containing 3-TA. This is the first demonstration of the use of this analogue in protein synthesis by *E. coli*.[6]

The target protein selected for this study is of sequence **3**. The construction of the gene encoding **3a** has been described previously.[5] This sequence is designed to adopt a lamellar

$$-[(\mathrm{GlyAla})_3\mathrm{GlyXxx}]-_{13}$$

3a: Xxx = Phe
b: Xxx = 3-TA

morphology with -GlyAla- dyads residing in the stem and Phe or 3-TA residues in the turn positions of the lamella. Evidence for such folded chain structures in similar polymers has been reported recently.[7,8]

Proteins **3a** and **3b** were produced in media containing phenylalanine and 3-thienylalanine, respectively (Figure 1). The target protein migrates at an apparent molecular weight of about 26 000, substantially higher than expected (7800). This anomalously low electrophoretic mobility is consistent with the behavior of several other proteins of this general design and is not due to genetic instability.[5,9,10] The small amount of protein observed in the absence of phenylalanine (Figure 1, lanes 1−3) may be due to incomplete removal of phenylalanine in the medium shift, or to production of phenylalanine by turnover of cellular proteins. These results indicate that 3-TA can be incorporated in place of phenylalanine in recombinant proteins produced in *E. coli*.

Protein **3a** was isolated from *E. coli* strain AFpLysS/pET-13GF as a fusion to the carboxy terminus of a leader peptide encoded by the plasmid pET-3b. For synthesis of **3b**, the cells were grown in Luria broth and shifted to medium containing 3-TA 10 min after induction with isopropyl β-D-thiogalactopyranoside (IPTG). Overexpression of either **3a** or **3b** resulted in the formation of inclusion bodies. The protein aggregates were isolated from the cell lysate and the repetitive protein was cleaved from the fusion construct with cyanogen bromide. The target proteins were separated from the other peptides by stepwise neutralization with ammonium hydroxide. Proteins **3a** and **3b** were isolated and purified in overall yields of 20 and 15 mg/L, respectively, which correspond to approximately 10% of cellular protein at the cell densities used in these experiments.

The relative proportions of phenylalanine and 3-TA in the target proteins were assessed by UV spectroscopy. Phenylalanine has an absorption maximum at 256 nm whereas the maximum for 3-TA is at 233 nm. The presence of phenylalanine in **3a** is confirmed by an absorption maximum at 257 nm;

[†] Department of Biochemistry and Molecular Biology.
[‡] Program in Molecular and Cellular Biology.
[§] Department of Polymer Science and Engineering.

(1) Roncali, J. ; Yassar, A.; Garnier, F. *J. Chem. Soc., Chem. Commun.* **1988**, 581.
(2) Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. *EMBO J.* **1990**, *9*, 1665.
(3) Dougherty, M. J.; Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *Macromolecules* **1993**, *26*, 1779.
(4) Richmond, M. H. *Bacteriol. Rev.* **1962**, *26*, 398−420.
(5) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471.
(6) 2-Thienylalanine, a related analogue of phenylalanine, has previously been shown to be used by the *E. coli* protein synthesis machinery (Munir, R.; Cohen, G. N. *Biochim. Biophys. Acta* **1959**, *31*, 378. Gabius, H. J.; von der Haar, F.; Cramer, F. *Biochemistry* **1983**, *22*, 2331). However, the 2-substituted thiophene function is of limited interest with respect to conducting polymers, which are formed via oxidative polymerization through the 2- and 5-positions of the thiophene ring.
(7) Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427.
(8) Parkhe, A. D.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1993**, *26*, 6691.
(9) 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.
(10) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727.
(11) Miller, J. H. *Experiments in Molecular Genetics*; Cold Spring Harbor Laboratory: New York, 1972; p 419.
(12) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning—A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Laboratory Press: New York, 1989.

0002-7863/95/1517-0536$09.00/0

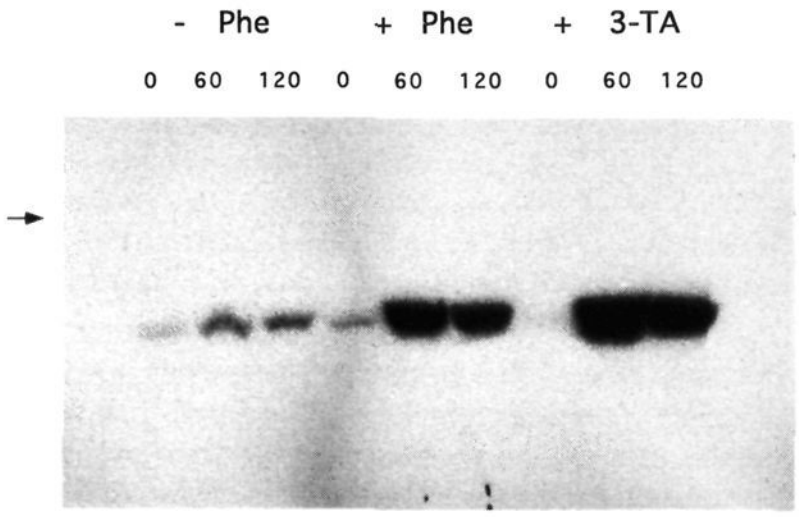


Figure 1. Autoradiogram showing patterns of protein synthesis. The phenylalanine auxotroph AFpLysS and the plasmid encoding the target protein have been described.[5] Cells were cultured in M9 medium[11] containing the 20 natural L-amino acids at 20 μg/mL and supplemented with 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, ampicillin at 200 μg/mL, and chloramphenicol at 40 μg/mL. Synthesis of the target protein was initiated by addition of 0.4 mM isopropyl β-D-thiogalactopyranoside (IPTG). Ten minutes later, the cells were harvested by centrifugation at 8000*g* for 10 min and suspended in an equal volume of medium containing M9 salts.[11] The cells were divided into three equal parts and repelleted by centrifugation. One aliquot was resuspended in M9 minimal medium containing all 20 natural L-amino acids, a second was resuspended in M9 medium containing 19 natural amino acids with phenylalanine omitted, and the third aliquot was resuspended in M9 medium containing 19 natural amino acids (minus phenylalanine) and 3-L-thienylalanine (50 μg/mL). All three mixes contained 200 μg/mL rifampicin, 0.4 mM IPTG, and 100 μCi of [^{3}H]glycine. Cells were incubated at 37 °C, and samples were withdrawn at 0, 60, and 120 min. The cells were collected by centrifugation, and cell lysates corresponding to OD_{600} = 0.2 unit were fractionated on a 12% polyacrylamide gel;[11] proteins were stained with Coomassie brilliant blue. The gel was destained and soaked in Enlightening enhancer (DuPont/NEN) for 30 min, dried, and exposed to X-ray film to detect the radiolabeled proteins. Lanes 1−3 contain target protein from cells shifted to medium lacking phenylalanine; lanes 4−6 are from cells shifted to medium containing phenylalanine; and lanes 7−9 represent cells shifted to medium containing 3-thienylalanine. Samples were withdrawn just after the medium shift (lanes 1, 4, 7), 60 min after the shift (lanes 2, 5, 8), and 120 min after the shift (lanes 3, 6, 9). Estimates of the sizes of **3a** and **3b** are based on extrapolation from the 29 kDa (arrow) and 48 kDa (not shown) standard markers.

Table 1. Amino Acid Compositions of Proteins **3a** and **3b**

		mol % (obsd)	
amino acid	mol % (theor)	3a	3b
glycine	50.0	49.4	47.0
alanine	37.7	35.6	34.1
phenylalanine	12.3[a]	11.7	2.2
3-TA	12.3[b]		10.8

[a] Expected for **3a.** [b] Expected for **3b.**

this peak is absent from the spectrum of **3b** (Figure 2). The presence of a shoulder at 233 nm on the major amide absorption peak argues that 3-TA is present in protein **3b** (data not shown). Quantitative analysis of the UV spectrum of **3b** places the extent of substitution at about 90%.

Direct evidence of incorporation of 3-TA was obtained from amino acid compositional analyses of **3a** and **3b** (Table 1). These analyses indicate that both proteins are at least 90% pure, and the data obtained for **3b** indicate that the extent of substitution is about 80%.

The incorporation of 3-TA was also verified by NMR spectroscopy. The aromatic proton regions of the spectra obtained from the free amino acids and from proteins **3a** and **3b** are shown in Figure 3. The spectrum of phenylalanine shows two sets of signals, both of which are observed in protein **3a**. Three signals are evident in the spectrum of 3-TA, corresponding

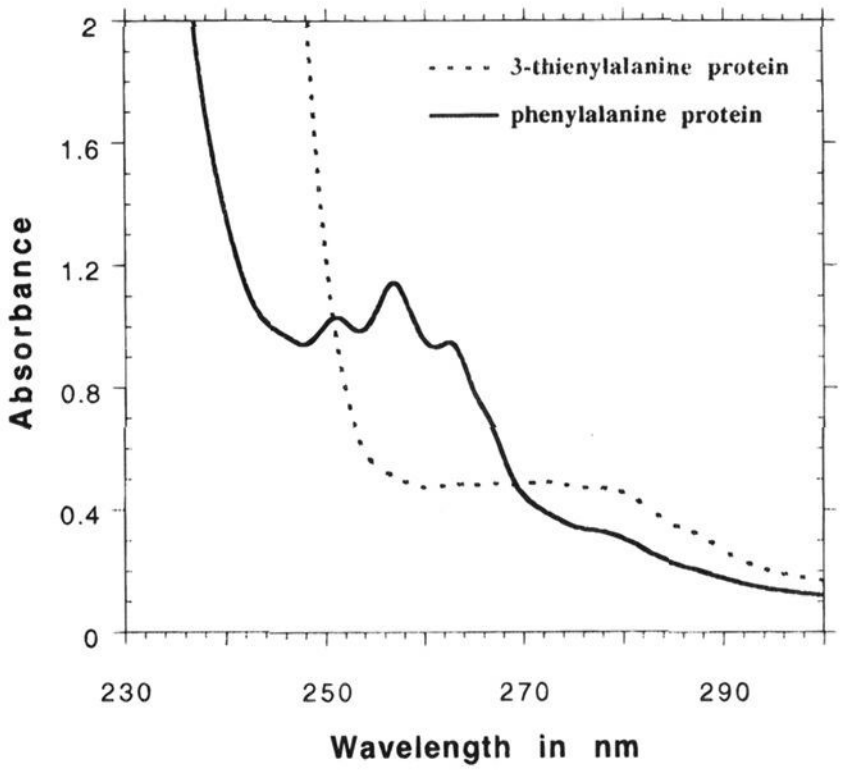


Figure 2. UV spectra of proteins **3a** and **3b** containing phenylalanine and 3-thienylalanine, respectively. The spectra were obtained with proteins dissolved in hexafluoro-2-propanol-d_2 at a concentration of 3.5 mg/mL.

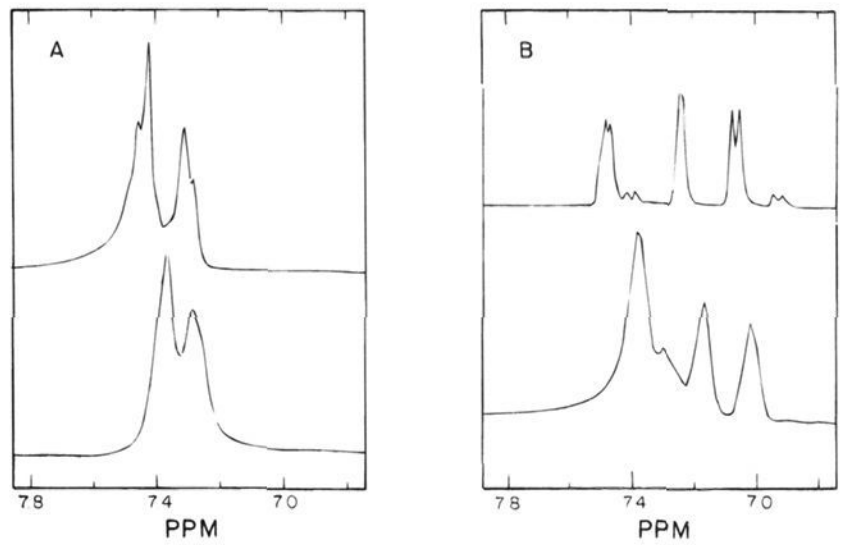


Figure 3. 3. NMR spectra of phenylalanine (top A), phenylalanine protein **3a** (bottom A), 3-thienylalanine (top B), and 3-thienylalanine protein **3b** (bottom B). All spectra were obtained in hexafluoro-2-propanol-d_2 as solvent.

to the three protons in the monosubstituted thiophene ring. Similar signals are observed in the spectrum of **3b**, with an increase in the line width and a small upfield shift analogous to that observed for **3a**. The increased intensity of the low-field signal in the spectrum of **3b** is ascribed to phenylalanine. Careful integration of the spectrum indicates an extent of substitution of about 80%, in good agreement with the amino acid compositional results. There is no evidence in these spectra that the 3-TA residues in **3b** have undergone any modification.

These results demonstrate an efficient strategy for *in vivo* incorporation of the amino acid analogue 3-TA into recombinant protein materials in place of phenylalanine. This is the first report of the incorporation of 3-TA into a recombinant protein product. Exploration of the electrochemical properties of protein **3b** are underway.

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. NMR spectra were recorded in the University of Massachusetts NMR Laboratory, which is supported by the NSF Materials Research Laboratory at the University.

JA943170I