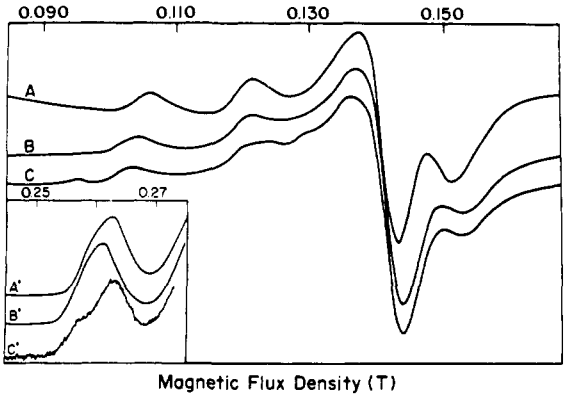

Figure 3. S-band EPR spectra of laccase derivatives recorded at room temperature in acetate buffer: A, ^{63}Cu-enriched laccase; B, ^{65}Cu-enriched laccase; C, ^{63}Cu-enriched T2D laccase that has been reconstituted with ^{65}Cu. Microwave frequency 3.92 GHz. Inset: Lowest flux density peaks in the corresponding X-band EPR spectra measured at 9.13 GHz and at −150 °C in acetate buffer. See text for a discussion of the shoulder in spectrum C'.

laccase along with the spectrum of a mixed-isotope sample prepared by reconstituting ^{63}Cu-enriched T2D laccase with the ^{65}Cu isotope. As predicted by the model, the room-temperature EPR spectrum of the mixed-isotope derivative shows that the type 2 copper site contains ^{65}Cu.

For completeness the $m_I = -3/2$ transitions of corresponding samples monitored at X-band frequency at −150 °C are presented in the inset in Figure 3. These data show that a different site is EPR active at low temperature. Signals 3C and 3C' were actually obtained with different solutions, but both samples were prepared in the same way. The extra feature on 3C' is assigned to splitting from ^{19}F due to a fluoride impurity which is bound to a fraction of the laccase molecules in solution.[17]

Even though the structural reorganization that has been identified occurs in frozen solution under nonphysiological conditions, the results are of considerable interest. In the first place, they indicate one type of reorganization that is feasible within the type 2/type 3 cluster. This is relevant because significant reorganization is likely to occur under active turnover, e.g., during the binding of reduced-oxygen intermediates or inhibitor anions.[10] In addition, this work adds credence to the recent proposal that a deprotonated water ligand can act as a bridge between the type 2 and the type 3 coppers in blue oxidases.[12] Finally, the results may explain why low-temperature EPR data implicate three nitrogens, rather than two, in the coordination sphere of the type 2 copper in tree laccase.[14]

Acknowledgment. This research was supported by Grants GM 22764 (D.R.M.) and RR-01008 (W.E.A.) from the National Institutes of Health.

Registry No. Cu, 7440-50-8; hydroxide, 14280-30-9; laccase, 80498-15-3.

(17) Simulations as well as the results in parts D and E of Figure 1 reveal that the splitting *cannot* be explained by invoking a mixture of ^{63}Cu and ^{65}Cu in the type 2 site.

Genetically Directed Syntheses of New Polymeric Materials. Expression of Artificial Genes Encoding Proteins with Repeating -(AlaGly)$_3$ProGluGly- Elements

Kevin P. McGrath,[†,‡] Maurille J. Fournier,[§,⊥] Thomas L. Mason,[§,⊥] and David A. Tirrell*[,†,§]

Contribution from the Departments of Polymer Science and Engineering and Biochemistry and Molecular and Cellular Biology Program, University of Massachusetts, Amherst, Massachusetts 01003. *Received April 11, 1991*

Abstract: The goals of this work were to develop reliable methods for the bacterial expression of artificial repetitive polypeptides and to examine the connection between primary sequence and solid-state conformation in such materials. DNA fragments encoding variable numbers of repeating (AlaGly)$_3$ProGluGly elements (**1**) were constructed via chemical synthesis and cloning of short oligonucleotides followed by self-ligation. The resulting multimer population was size fractionated and cloned into an adapter plasmid that provides methionine residues at positions flanking the repetitive sequences. DNA multimers modified in this way, and encoding 10, 18, 28, or 54 repeats of sequence **1**, were transferred to an expression vector that features a T7 phage promoter, and protein expression in *Escherichia coli* was monitored via incorporation of [^{3}H]glycine. Purification of the target proteins was accomplished by selective precipitation with ammonium sulfate and subsequently with dilute acetic acid, followed by ion-exchange chromatography on DEAE-Sephadex. Typical yields of purified protein were ca. 10 mg/L of fermentation medium. The structure of the longest chain length variant was confirmed by amino acid compositional analysis, N-terminal protein sequencing (through 58 residues), matrix-assisted laser desorption mass spectrometry, ^{1}H and ^{13}C NMR spectrometry, cyanogen bromide cleavage, and combustion analysis. This polymer can be cast into coherent, optically clear films and displays a reversible glass transition at 170 °C. Thermal decomposition starts at ca. 250 °C and results ultimately in the loss of 70% of sample weight. Evidence from differential scanning calorimetry, X-ray scattering, and Fourier transform infrared spectroscopy suggests that this polymer and its CNBr cleavage product form amorphous glasses at room temperature.

Introduction

The solid-state and solution properties of polymeric materials are determined by five critical architectural variables: molecular size, topology, composition, sequence, and stereochemistry. Of these, topology alone is subject to control in an absolute sense; conventional polymerization processes allow only statistical control of each of the remaining structural variables. As a result, the polymeric materials currently in use—even in fundamental studies—are not pure substances but instead are mixtures characterized by substantial molecular heterogeneity.

Continuing progress in the synthesis, cloning, and expression of artificial genes has provided a powerful method for the preparation of structurally homogeneous "synthetic" polypeptides. This

[†] Department of Polymer Science and Engineering.
[‡] Present address: Department of the Army, Natick Research, Development and Engineering Center, Natick, MA 01760.
[§] Department of Biochemistry.
[⊥] Molecular and Cellular Biology Program.

0002-7863/92/1514-727$03.00/0

development raises important questions about the role of molecular heterogeneity in determining the behavior of polymeric substances and offers the prospect of significant new opportunities in polymer materials science.

Polypeptides of regular, *periodic* structure are of particular interest in this regard. Naturally occurring structural proteins (e.g., silk, collagen, or elastin) consist in large part of highly repetitive sequences[1], and the extended (helical or beta) secondary structures characteristic of many repetitive polypeptides provide an intriguing set of building blocks for use in the design and construction of highly organized polymeric solids. This paper describes initial experiments directed toward the general goal of preparing architecturally well-defined polypeptides of predictable solid-state structures. To this end, we are exploring a wide variety of artificial proteins that feature periodically repetitive structural elements.

Artificial genes encoding large repetitive polypeptides have been expressed previously in bacterial hosts, but with mixed success.[2-5] Such experiments present special problems, because repetitive DNAs may be particularly prone to rearrangement[6] and because repetitive messenger RNAs may adopt folded structures that are translated inefficiently or degraded rapidly.

We report herein the synthesis and expression of a family of artificial genes that encode four chain length variants of polymers built from the nonapeptide repeating unit **1**. The design of **1**

$$-(\text{AlaGly})_3\text{ProGluGly}-$$
1

rested on several observations: (i) that poly(glycylalanine) adopts an antiparallel β-sheet arrangement in the solid,[7] (ii) that proline frequently initiates β turns in globular proteins,[8] and (iii) that proline and glutamic acid are poor β-sheet formers.[9] These observations led us to posit a folded-chain lamellar structure for polymers of **1**, analogous to the known cross-β conformation adopted by the egg stalk protein of *Chrysopa flava*.[10] In the expected structure, crystalline lamellae would be built from stacked, folded sheets, with the periodic proline and glutamic acid residues defining the lamellar thickness and decorating the lamellar surfaces. The expected thickness of ca. 30 Å is characteristic of cross-β structures,[10] and the use of an odd number of amino acids in the repeating sequence leads to an asymmetric (or "polar"[11]) chain arrangement in which alanyl methyl groups extend from only one face of the regularly folded sheet. Such an arrangement was sought in our design, since the generally accepted crystal structures of silk[12] and poly(glycylalanine)[7] are built from polar sheets with alanyl methyls in register.

We discuss below the efficient expression of several polymers of **1** in *Escherichia coli*, as well as isolation, purification, and structural analysis of these artificial proteins. We find that these materials do not, in fact, adopt the anticipated lamellar structure, and we offer a simple rationale for this result. The experiments reported herein provide a basis for rapid, efficient synthesis of artificial repetitive proteins and constitute a first step toward a general strategy for the design and control of structural organization in crystalline polymers.

Experimental Section

Materials. T4 DNA ligase, T4 polynucleotide kinase, and all restriction enzymes were obtained from New England Biolabs (Beverly, MA). Ribonuclease A and deoxyribonuclease I were from Sigma Chemical Co. Calf intestinal alkaline phosphatase was supplied by U.S. Biochemical Corp. (Cleveland, OH). *E. coli* strains DH5αF′ and HB101 were purchased from Bethesda Research Labs (Bethesda, MD). Plasmid p937.51 was a gift from Protein Polymer Technologies, Inc. (San Diego, CA). The expression plasmid pET3-b and *E. coli* strain BL21-(DE3)-pLysS were obtained from William Studier at the State University of New York, Stony Brook. [^{3}H]Glycine was supplied by Du Pont/NEN (Wilmington, DE).

General Methods. Bacterial growth media, DNA manipulations, and transformation conditions were as described in Sambrook, Frisch, and Maniatis.[13] Fourier transform infrared spectra were obtained on samples prepared as films on AgCl plates using an IBM System 9000 Fourier transform infrared spectrophotometer. ^{1}H NMR spectra were run on a Varian XL-300 spectrometer at an operating frequency of 300 MHz and ^{13}C NMR spectra on a Varian XL-200 instrument at an operating frequency of 50 MHz. Ultraviolet spectra were obtained in distilled water on a Hitachi U2000 spectrophotometer using quartz cuvettes with a path length of 1 cm. Polymer films were cast on Teflon from 1% (w/v) aqueous solutions. X-ray diffraction patterns were recorded in a Statton camera using a pinhole collimated, Ni-filtered Cu Kα X-ray source and a sample to film distance of 56 mm. Differential scanning calorimetry data were obtained on a Perkin-Elmer DSC 4 instrument at a heating rate of 20 °C/min. After loss of water on the first heating of the sample, subsequent heating runs were virtually identical.

Preparation of Synthetic DNA. Oligonucleotides were synthesized on a Biosearch Model 8700 DNA synthesizer using the phosphoramidite chemistry of McBride and Caruthers.[14] Syntheses were carried out on a 1.0-μmol scale. Crude oligonucleotides were purified by preparative gel electrophoresis, using 15% polyacrylamide gels containing 8.3 M urea. Purified oligonucleotides (500 pmol, 11.3 μg) were phosphorylated with T4 polynucleotide kinase, annealed at 80 °C, and cooled to room temperature over approximately 3 h. The annealed duplex was ethanol precipitated and dried in vacuo.

Cloning and Amplication of Synthetic DNA. The purified synthetic duplex (**2**, Figure 1) was inserted into pUC18[15] digested with *Eco*RI and *Bam*HI. The ligation mixture was used to transform *E. coli* strain DH5αF′ cells which were then grown on YT medium containing 200 μg/mL ampicillin, 25 μg/mL of the sugar analogue β-isopropylthiogalactoside (IPTG), and 40 ng/mL of the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). Plates were incubated at 37 °C overnight.

Recombinant plasmids containing the synthetic insert were identified by color assay and verified by sequencing of the double-stranded plasmid DNA. After isolation and purification from a 1-L growth, the recombinant plasmid was digested with *Ban*I and the 54-base-pair DNA monomer was purified by electrophoretic separation in 2% agarose and recovered by electroelution.

Polymerization of the DNA Monomer. Purified DNA monomer (1 μg) was self-ligated in a total volume of 30 μL using 6 Weiss units of T4 DNA ligase. The ligation products were separated by electrophoresis on a 1.5% agarose gel, and multimer fragments $\geq$10 monomers in length were recovered by electroelution.

Cloning of DNA Multimers. This population of multimers was inserted into the unique BanI site of p937.51,[16] and the resulting ligation mixture was used to transform *E. coli* strain HB101. Analysis of 60 transformants revealed that 22 contained synthetic inserts, ranging in size from 1 to 27 repeats of the monomer segment. Plasmids containing 5, 9, 14, and 27 repeats of the DNA monomer segment were chosen for further analysis.

Construction of Bacterial Expression Vectors. Plasmids containing the various inserts were isolated from transformed cells grown to saturation in 100-mL cultures. Plasmid DNA was digested with *Bam*HI, and fragments containing the multimer inserts were separated by electrophoresis in 1% agarose and recovered by electroelution. Approximately

(1) Walton, A. G.; Blackwell, J. *Biopolymers*; Academic Press: New York, 1973; Chapter 10.
(2) McGrath, K. P.; Tirrell, D. A.; Kawai, M.; Mason, T. L.; Fournier, M. J. *Biotechnol. Prog.* **1990**, *6*, 188 and references therein.
(3) Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Biotechnol. Prog.* **1990**, *6*, 198.
(4) Goldberg, I.; Salerno, A. J.; Patterson, T.; Williams, J. I. *Gene* **1989**, *80*, 305.
(5) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213.
(6) Mahajan, S. K. *Genetic Recombination*; Kucherlapati, R., Smith, G. R., Eds.; American Society for Microbiology: Washington, DC, 1988; Chapter 3.
(7) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706. Anderson, J. M.; Chen, H. H.; Rippon, W. B.; Walton, A. G. *J. Mol. Biol.* **1972**, *67*, 459.
(8) Chou, P. Y.; Fasman, G. D. *J. Mol. Biol.* **1977**, *115*, 135.
(9) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211.
(10) Geddes, A. J.; Parker, K. D.; Atkins, E. D. T.; Beighton, E. *J. Mol. Biol.* **1968**, *32*, 343.
(11) This terminology appears to have been introduced by Takahashi and co-workers: Takahashi, Y.; Gehoh, M.; Yuzuriha, K. *J. Polym. Sci., Part B: Polym Phys.* **1991**, *29*, 889.
(12) Marsh, R. E.; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* **1955**, *16*, 1.
(13) Sambrook, J.; Frisch, E. F.; Maniatis, T. *Molecular Cloning, A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Press: Cold Spring Harbor, 1989.
(14) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.
(15) Yanisch-Perron, C.; Vieira, J.; Messing, J. *Gene* **1985**, *33*, 103.
(16) Ferrari, F. A.; Richardson, C.; Chambers, J.; Causey, S. C.; Pollock, T. J. U.S. Patent Appl. 927,258, 1986.

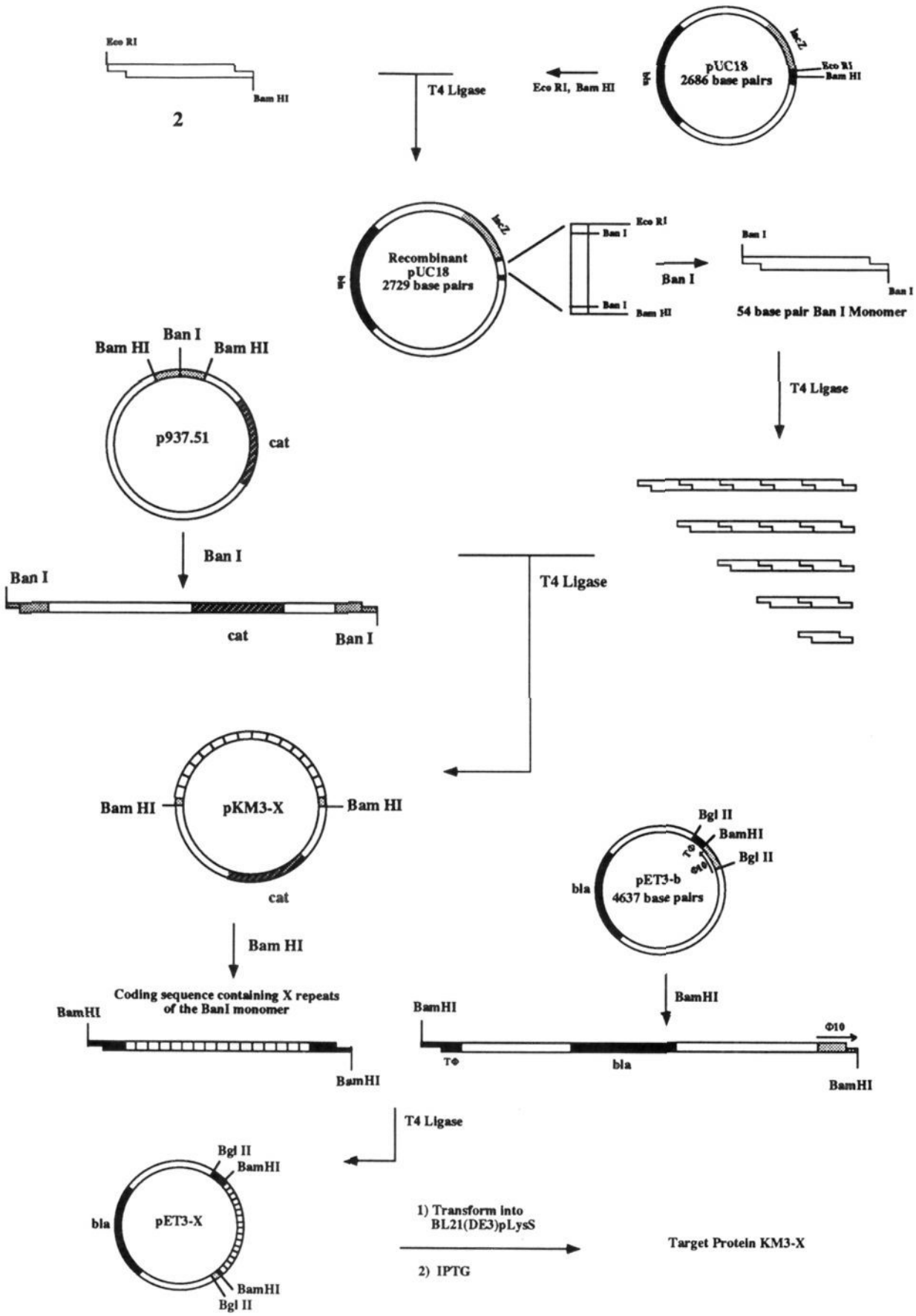

Figure 1. Strategy for cloning and expression of artificial genes encoding proteins with repeating $-(AlaGly)_3ProGluGly-$ elements. See text for details.

100 ng of each fragment was mixed with 200 ng of *Bam*HI-digested pET3-b[17] and ligated as described above. A portion of each ligation mixture was used to transform *E. coli* strain HB101. The presence and orientation of inserts in the recombinant plasmids were determined by digestion with AvaI, which yields restriction fragments of sizes dependent on the direction of insertion. Recombinant plasmids containing insert in the correct orientation were used to transform *E. coli* strain BL21-(DE3)pLysS.[17]

Protein Expression. Cultures containing recombinant pET plasmids in BL21(DE3)pLysS were grown at 37 °C under antibiotic selection in 10 mL of M9 medium supplemented with 20 μg/mL of each amino acid except glycine. When the cell density reached ca. 5×10^8 cells/mL, [^{3}H]glycine (100 μCi) was added to each culture. Cultures were grown for another 10–15 min, IPTG was added to a final concentration of 0.4 mM, and incubations were continued for another 1 h.[18] Cells from each culture were lysed in 1% SDS and heated at 95 °C for 5 min. Lysates were resolved on 12% polyacrylamide gels by the method of Laemmli,[19] and the incorporation of label was analyzed by fluorography of the dried gel using Enlightening enhancer (Du Pont/NEN).

Large-Scale Expression and Purification of KM3-27. BL21(DE3)-pLysS transformed with pET3-27 (the recombinant plasmid containing 27 repeats of the DNA monomer) was grown to midexponential phase in a New Brunswick Scientific Microferm fermenter in 10 L of YT medium containing 200 mg/L ampicillin and 25 mg/L chloramphenicol at 37 °C, and expression was induced by the addition of 40 mL of 0.2 M IPTG in dimethylformamide. After 2 h the cells were isolated by centrifugation at 3000*g* for 10 min at 4 °C. Cells were resuspended in 400 mL of 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA, placed on ice, and sonicated using a Branson sonifier at 40 W for 15 min. Cellular debris was removed by centrifugation at 10000*g* for 20 min at 4 °C. The supernatant was collected and adjusted to 40% $(NH_4)_2SO_4$ saturation by the addition of 67 mL of 100% $(NH_4)_2SO_4$ solution per 100 mL of clarified lysate. After incubation of the solution overnight at 37 °C, the resulting precipitate was removed by centrifugation at 10000*g* for 20 min at 4 °C. The supernatant was collected, adjusted to 60% $(NH_4)_2SO_4$ saturation, and incubated overnight again at 37 °C. The precipitated proteins were collected by centrifugation at 10000*g* for 20 min at 4 °C, the supernatant was discarded, and the pellet was redissolved in 400 mL of distilled, deionized water. The solution was adjusted to 50 mM Tris-HCl (pH 8.0) and 10 mM $MgCl_2$, and 1 mg each of RNase A and DNase I was added. After incubation at 37 °C for 1 h, the solution was brought to pH 4.0 by the addition of dilute aqueous acetic acid and then incubated at 4 °C for 2 h. The precipitate was removed by centrifugation at 10000*g* for 20 min at 4 °C, and the supernatant was collected and dialyzed against constantly changing distilled water for 24 h. The dialyzate was collected and loaded onto a 10 × 120 mm column of DEAE-Sephadex A-25 equilibrated in 100 mM Tris-HCl (pH 8.0) and 1 mM EDTA. Proteins were eluted with a continuous gradient of NaCl in the equilibration buffer. Fractions eluting between 150 and 200 mM NaCl were pooled and dialyzed against distilled water for 2 days at room temperature. The protein was recovered by lyophilization and dried at 78 °C in vacuo to a constant weight. The yield of the protein, designated KM3-27, was 102 mg.

Cyanogen Bromide Cleavage of KM3-27. Cyanogen bromide cleavage was accomplished by the method of Smith.[20] Purified KM3-27 (15 mg) was dissolved in 4 mL of 88% formic acid. Distilled, deionized water (1 mL) was added to bring the final concentration of formic acid to 70%. Crystals of CNBr (approximately 30–50 mg) were added, and the reaction mixture was incubated at room temperature for times ranging from 12 to 48 h. The reaction was terminated by drying down the solution in vacuo. The pellet obtained was resuspended in 10 mL of distilled deionized water and placed into a Spectra-Por dialysis membrane with a molecular weight cutoff of 15000. The solution was dialyzed against constantly changing water for 4 days, and the cleaved protein was recovered by lyophilization. The amount of protein recovered was 8 mg (61%).

For the purposes of electrophoretic analysis of cleavage products, purified radiolabeled KM3-27 (ca. 300 μg) was dissolved in 1 mL of 70% formic acid. A single crystal of CNBr (approximately 4 mg) was added and the reaction mixture incubated at room temperature for 2 days. Every 12 h a 200-μL aliquot of the reaction mixture was removed, and the reaction therein was terminated by drying down the solution in vacuo. The resulting pellet was analyzed on a 10% polyacrylamide gel, essentially as described above. The gel was stained with Coomassie Brilliant Blue G-250, destained, then soaked for 1 h in Enlightening fluor, transferred onto Whatman 3MM filter paper, and dried at 60 °C for 4 h. The dried gel was exposed for 2 days at −80 °C.

Results and Discussion

Gene Construction. The strategy used to prepare artificial proteins of sequence **1** is outlined in Figure 1. The 70-base-pair DNA fragment **2** was synthesized by the phosphoramidite me-

```
          Stop Gly Ala Gly Ala Gly Ala Gly Pro Glu Gly Ala Gly Ala Gly Ala Gly Pro Glu Gly Ala
AAT TCG TAA GGT GCC GGC GCT GGT GCG GGC CCG GAA GGT GCA GGC GCT GGT GCG GGC CCG GAA GGT GCC G
      GC ATT CCA CGG CCG CGA CCA CGC CCG GGC CTT CCA CGT CCG CGA CCA CGC CCG GGC CTT CCA CGG CCTAG
EcoRI      BanI                                                                  ApaI      BanI  BamHI
```

2

thod,[14] phosphorylated, and cloned between the *Eco*RI and *Bam*HI sites of the cloning vector pUC18.[15] The design of fragment **2** addressed the following issues. First, the TAA stop codon immediately following the 5′ EcoRI site ensures disruption of the β-galactosidase α fragment encoded by pUC18 and allows strong blue/white screening for recombinant plasmids. Second, the coding sequence of **2** is flanked by two *Ban*I restriction sites which are used to isolate the 54-base-pair monomer after cloning and amplification. Use of cloned monomer fragments, rather than those obtained directly by chemical synthesis, allows the sequence to be confirmed before multimerization. The nonpalindromic ends generated by *Ban*I digestion suppress formation of inverted repeats upon polymerization of the DNA monomer. Last, the choice of codons for each amino acid reflects four considerations: (i) the pattern of codon use in *E. coli*;[19] (ii) avoidance of strict sequence periodicity within the monomeric DNA repeat; (iii) the need to eliminate all *Ban*I sites other than those flanking the coding sequence; (iv) incorporation of an *Apa*I site for use in subsequent screening of transformants for the presence of the synthetic insert. After cloning, fragment **2** was sequenced and amplified in *E. coli* strain DH5αF′. The monomer was then liberated by *Ban*I digestion and purified by agarose gel electrophoresis and electroelution.

Construction of genes encoding repeated –(AlaGly)$_3$ProGluGly– elements involved self-ligation of the DNA monomer, fractionation of the resulting multimers by electrophoresis in 1.5% agarose, and isolation of inserts of a range of sizes. Ligation typically afforded distributions of multimers with degrees of polymerization extending to more than 40, but with the center of the distribution between 10 and 15. This population was fractionated to remove multimers with degrees of polymerization of less than 10, and the resulting fraction was purified by electroelution. Although the separation of larger fragments from smaller ones was not absolute, this method proved useful in generating fragment populations with high degrees of polymerization.

This population of multimers was inserted into the unique *Ban*I site of p937.51, a small, high copy number cloning vector that carries an origin of replication derived from pBR322 and a gene encoding chloramphenicol acetyltransferase.[16] Transformants containing, inter alia, 5, 9, 14, and 27 repeats of the monomeric DNA fragment were identified by *Bam*HI digestion, and fragments containing the multimerized inserts were purified by electroelution.

Purified multimeric fragments were inserted into pET3-b, an expression vector developed by Studier and co-workers,[17] in which transcription is driven by the T7 phage RNA polymerase from the gene 10 promoter. The orientation of each insert was determined by *Ava*I digestion. The resulting recombinant plasmids encode variable numbers of repeats of the nonapeptide **1**, with the repetitive portion flanked in each case by N- and C-terminal extensions of 23 and 33 residues, respectively, as shown in sequence **3**.[22] These terminal regions, which are derived from p937.51 and

(17) Studier, F. W.; Rosenberg, A. H.; Dunn, J. J. *Methods Enzymol.* **1990**, 185, 60.

(18) Addition of the label 10–15 min before induction provides a convenient level of labeling of background proteins (cf. Figure 3). Addition of the label subsequent to induction leads to stronger labeling of the protein of interest but affords no real advantage in detection of the product.

(19) Laemmli, U. K. *Nature* **1970**, *227*, 680.

(20) Smith, B. J. *Methods in Molecular Biology, New Protein Techniques*; Humana: Clifton, NJ, 1988; p 70.

(21) Aota, S.; Gojobori, T.; Ishibashi, F.; Maruyama, T.; Ikemura, T. *Nucleic Acids Res.* **1988**, *16*, r315.

(22) Standard single-letter abbreviation for the amino acids may be found in many biochemical textbooks. See, for example: Mathews, K.; van Holde, K. E. *Biochemistry*; Benjamin/Cummings: Redwood City, 1990.

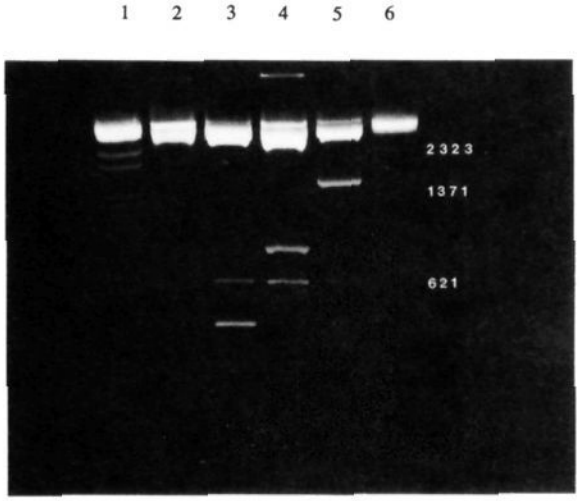

Figure 2. *Bam*HI digests of recombinant expression vectors in *E. coli* BL21(DE3)pLysS transformants: lanes 1 and 6, λ-BstEII molecular weight markers; lane 2, pET3-5; lane 3, pET3-9; lane 4, pET3-14; lane 5, pET3-27. Analysis was performed in 1.5% agarose; 2–3 μg of DNA was loaded per lane. The 621-base-pair band is derived from pLysS.

from pET3-b, can be removed efficiently by CNBr cleavage at the methionine residues flanking the sequence of interest (vide infra).

MASMTGGQQMGRDPMFKYSRDPM–
[GAGAGAGPEGAGAGAGPE]$_X$–
GARMHIRPGRYQLDPAANKARKEAELAAATAEQ 3

Protein Expression. The host used for protein expression was *E. coli* strain BL21-(DE3)pLysS.[17] In this strain, a gene encoding T7 RNA polymerase is incorporated into the bacterial chromosome under *lacUV5* control, and protein production is induced by addition of IPTG. The pLysS plasmid supplies T7 lysozyme, which inhibits T7 RNA polymerase and suppresses the basal level of protein expression.

Figure 2 shows the results of electrophoretic analysis of inserts recovered from BL21(DE3)-pLysS cells transformed with the recombinant pET vectors of interest. In each case (lanes 2–5), a single insert band was observed, in addition to a 621-base-pair band derived from pLysS. We have observed no evidence of genetic instability (e.g., as revealed by length polymorphism in the insert) in over 1 year of work with these sequences in the pET expression plasmids.

Protein expression was monitored by the in vivo incorporation of [^{3}H]glycine into proteins synthesized during exponential growth. Radiolabeling was dictated by the fact that the proteins of interest bind conventional stains weakly if at all. Figure 3 shows the results of electrophoretic analyis of whole-cell lysates prepared immediately before or 60 min after induction with IPTG. Prominent new protein bands are apparent after 60 min, whereas these bands are absent prior to induction. In each case, the new product migrates as a single band; no evidence is found for the accumulation of chain length variants that would arise from polymorphism in the insert or from premature translational termination. All of the target proteins were found in the soluble fraction of the cell lysate; none appeared to accumulate in inclusion bodies. Perhaps the high charge density on the chain precludes aggregation and precipitation.

The apparent molecular weights of these proteins as estimated by electrophoresis in the presence of SDS are significantly larger than expected. Figure 4 compares the observed mobilities of the four proteins of interest with those of the molecular weight marker proteins used to calibrate the gel. In every case, the mobility of the artificial protein is anomalously low, though the effect is reduced at short chain lengths. Because the chain lengths are in fact as expected (vide infra), this phenomenon must reflect an unusual sequence-dependent reduction in electrophoretic mobility. We suggest that these highly acidic artificial proteins bind the anionic surfactant SDS only weakly (and therefore suffer a reduced effective charge in the electrophoretic separation) though we cannot rule out geometric effects on mobility.

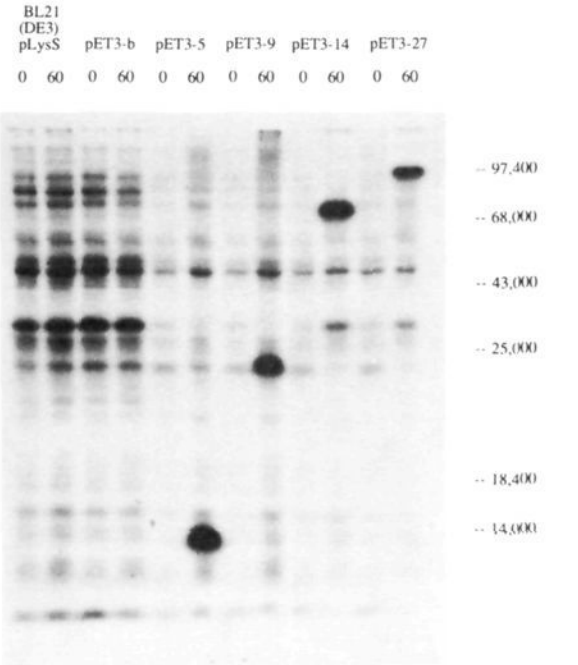

Figure 3. In vivo labeling of proteins containing repeated $(AlaGly)_3ProGluGly$ elements. Time points indicated (in minutes) are relative to IPTG addition. Analysis performed in 12% polyacrylamide at 15-mA constant current. Molecular weight markers are indicated at the right.

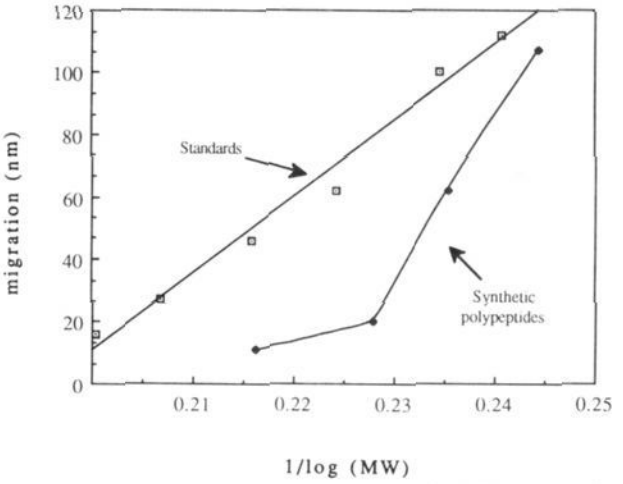

Figure 4. Plot of mobility versus the reciprocal of log MW for the series of polypeptides containing repeated $(AlaGly)_3ProGluGly$ elements. Analaysis performed in 10% polyacrylamide at 15-mA constant current. Standards: α-lactalbumin, β-lactoglobulin, carbonic anhydrase, albumin (egg), bovine serum albumin, phosphorylase B (rabbit muscle).

Protein Purification and Analysis. Characterization efforts were focused on the polymer containing 54 repeats of the nonapeptide sequence **1**, which was derived from an artificial gene comprising 27 repeats of the DNA monomer and designated KM3-27. The protein was isolated from 10-L batch fermentations by a series of ammonium sulfate and acid precipitations followed by ion-exchange chromatography. A brief nuclease treatment was required to eliminate contamination by nucleic acids, as shown by monitoring of the absorbance of polymer solutions at 260 nm. Typical yields after purification were ca. 10 mg of protein/L of culture.

The primary structure of KM3-27 was verified by N-terminal sequencing, which confirmed the expected sequence through the first 58 residues (the experimental limit). The N-terminal me-

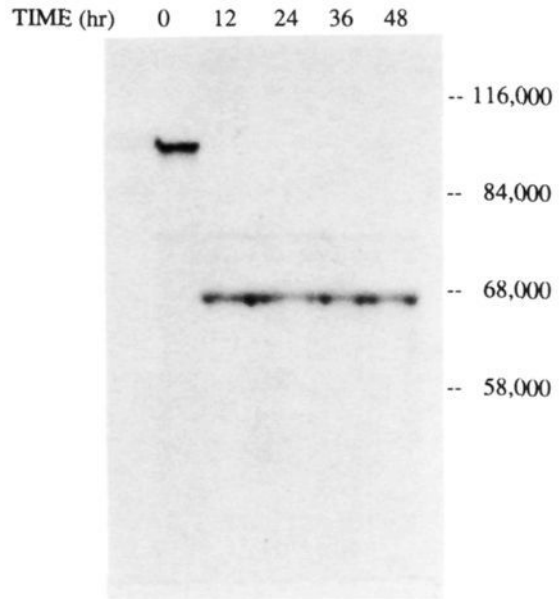

Figure 5. CNBr cleavage analysis of protein KM3-27. Radiolabeled protein was treated with CNBr for the times indicated, and the digested protein was resolved by electrophoresis in 10% polyacrylamide at 15-mA constant current. Molecular weight markers indicated at the right.

thionine residue was not detected and is apparently removed in vivo. Amino acid compositional analysis showed the product to consist of 98% glycine, alanine, proline, and glutamic acid vs the expected total of 94% for these four residues. By matrix-assisted laser desorption mass spectrometry[23] (courtesy of Professor Brian Chait and Dr. Ronald Beavis of Rockefeller University), a molecular ion was observed at m/z 42 089, in excellent agreement with the expected value of 42 123. ^1H and ^{13}C NMR spectra were fully consistent with sequence **3**. Anal. Calcd for **3** plus 13% water: C, 43.2; H, 6.7; N, 16.7. Found: C, 43.2; H, 6.0; N, 15.9. The presence of 13% water in the polymer was confirmed by thermogravimetric analysis (vide infra).

CNBr Cleavage of KM3-27. Further confirmation of the structure of KM3-27 was obtained by cleavage of the polymer with cyanogen bromide. Electrophoretic analysis of the CNBr cleavage products in Figure 5 shows that a single predominant product was produced in reaction times of 12–48 h. None of the initial sample was resistant to digestion, and there was no evidence of multiple cleavage fragments. As before, the molecular weight of the product as determined by electrophoresis is higher than expected, but mass spectrometry shows a well-defined molecular ion at m/z 36 450, within 10 amu of the expected mass. Although they were not apparent on gels, some minor fragments of lower m/z appeared in the mass spectrum of the sample digested for 48 h and probably represent products of slow cleavage reactions within the periodic portion of sequence **3**. Amino acid analysis showed the CNBr cleavage product to be comprised exclusively of glycine, alanine, proline, and glutamic acid, and N-terminal sequencing confirmed removal of the N-terminal fusion fragment.

Solid-State Properties. KM3-27 can be cast into optically clear, coherent films from formic acid or from water. Differential scanning calorimetry shows loss of water upon first heating and a reversible glass transition at 170 °C upon subsequent heatings. No evidence of melting or crystallization is obtained by this method. Thermogravimetric analysis reveals an initial loss of 13% water by ca. 100 °C, followed by a relatively rapid decomposition above 250 °C that results in ultimate loss of 70% of the weight of the sample.

Figure 6 shows a typical wide-angle X-ray scattering pattern for a film of KM3-27 cast from water. The pattern consists of diffuse halos and offers no evidence for the β-sheet structure

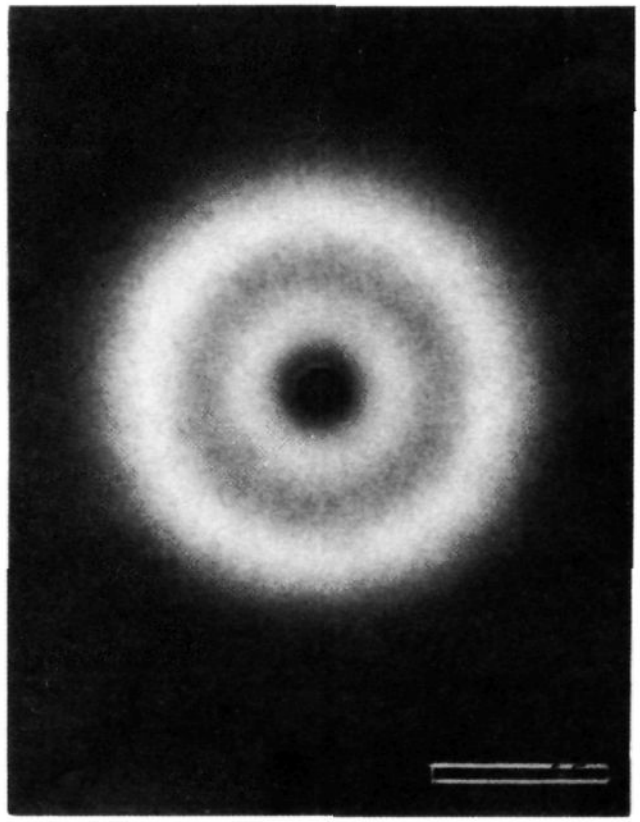

Figure 6. Wide-angle X-ray scattering from film of KM3-27 cast from water. The sample to film distance was 52.6 mm; the bar represents 20 mm on the original exposed film.

characteristic of poly(glycylalanine)[7] or *Bombyx mori* silk fibroin.[12] Fourier transform infrared spectra also indicate an absence of the β conformation, in that the amide I and amide II bands appear at 1653 and 1540 cm^{-1}, respectively, rather than at ca. 1630 and 1525 cm^{-1}, as expected for the β form.[24] Similar results were obtained on the CNBr cleavage fragment of KM3-27 and on samples of these polymers prepared in lyophilized form or as films cast from formic acid.

The rationale for the design of sequence **1** was outlined in the Introduction and assigns clear conformational roles to each of the constituent amino acids. An additional feature of the design was the use of a nine amino acid repeat, which places alanyl residues on neighboring β strands in register in the folded structure, such that the alanyl methyl groups all extend from the same face of the sheet. Adjacent sheets would then be expected to pack as they do in poly(glycylalanine),[7] with glycine–glycine contacts alternating with alanine–alanine contacts in the stacked sheet array.

It now appears likely that it is the use of an odd number of amino acids in the repeat that has frustrated the development of regular secondary structure in polymers built from sequence **1**. Examination of molecular models of the folded chain shows that the array of hydrogen bonds required to stabilize the β conformation cannot be maintained over the full extent of the sheet. Furthermore, the chain trajectory at the termini of the β strands is such that normal β turns can be accommodated at only one—not both—of the edges of a sheet built from repeating units consisting of odd numbers of amino acids. We propose that these features of sequence **1** destabilize the folded conformation to such an extent that the entropically favored glassy state of the solid is preferred at room temperature. In support of this view, we note that polymers of the octapeptide sequence **4**, recently prepared in our laboratory, have shown evidence for a folded β structure in preliminary experiments. A full description of the preparation and analysis of polymers of sequence **4** will be forthcoming.

(23) Beavis, R. C.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6873.

(24) Moore, W. H.; Krimm, S. *Biopolymers* **1976**, *15*, 2465.

$$-(AlaGly)_3GluGly-$$
$$4$$

Finally, we note a recent report in which Takahashi and coworkers challenge the classical "polar" sheet model for *B. mori* silk fibroin and propose instead an apolar antiparallel structure with alanyl methyls extending from both sheet faces. Clearly, the availability of appropriately engineered polypeptides will facilitate the development of an understanding of the factors that control the conformational and crystallization behavior of polymeric solids, both natural and synthetic.

Conclusions

Genes encoding four different chain length variants of polymers of the nonapeptide sequence **1** have been constructed and expressed in *E. coli*, and the corresponding proteins have been isolated and characterized. There was no evidence of genetic instability of the synthetic DNA in the translation system used. Mass spectrometry confirms the size of the largest protein, which was obtained in a yield of approximately 100 mg from a 10-L fermentation. Structural analysis of powders and films prepared from 54 repeat variants of sequence **1** indicates that the polymers form amorphous glasses. These results demonstrate the biological feasibility of generating useful quantities of repetitive artificial proteins of this general class and illuminate further the issues to be addressed in the design of solid-state structure in polymers.

Acknowledgment. We thank Dr. Joseph Cappello of Protein Polymer Technologies, Inc., for the gift of p937.51 and for fruitful discussions regarding the gene construction described herein. Professor Brian Chait and Dr. Ronald Beavis of Rockefeller University provided mass spectra and insight into their interpretation. Dr. Harald Modler recorded the scattering pattern in Figure 6, Mr. Jeffrey Kollodge assisted in thermal analysis, and Professor E. D. T. Atkins of the University of Bristol has made many useful contributions to our thoughts regarding solid-state protein structure. This work was supported in part by the Center for University of Massachusetts-Industry Research on Polymers and in part by a grant from the Polymers and Genetics Programs of the National Science Foundation (NSF DMR 8914359). NMR and infrared spectra were recorded in the spectroscopy facility of the NSF Materials Research Laboratory at the University of Massachusetts.

Registry No. 1, 137334-66-8.

Cannizzaro-Based O_2-Dependent Cleavage of DNA by Quinocarcin

Robert M. Williams,*,† Tomasz Glinka,† Mark E. Flanagan,† Renee Gallegos,† Hazel Coffman,† and Deihua Pei‡

Contribution from the Departments of Chemistry, Colorado State University, Fort Collins, Colorado 80523, and University of California, Berkeley, California 94720.
Received May 28, 1991

Abstract: A novel mechanism for the reduction of molecular oxygen that results in the O_2-dependent cleavage of both single-stranded and double-stranded DNA by quinocarcin (**1**) is presented. The results are discussed in the context of a redox self-disproportionation (Cannizzaro-type) of the oxazolidine moiety of quinocarcin which produces superoxide.

It is now widely recognized[1] that a variety of clinically significant antitumor antibiotics can mediate oxygen-dependent cleavage of the ribose–phosphate backbone of cellular DNA and RNA. An enormous structural array of interesting natural products and semisynthetic and totally synthetic substances mediate oxidative strand scission of nucleic acids through three main families[2] of reactions: (1) metal-mediated activation of O_2 ultimately producing hydroxyl radical or other reactive oxygen species;[3-14] (2) non-metal-dependent generation of reactive carbon radicals[15] that mediate C–H abstraction from the deoxyribose backbone (the resulting deoxyribosyl radical subsequently reacts with molecular oxygen culminating in strand scission), and (3) photolytic production of hydroxyl radical,[16] which does not require metal participation for the DNA cleavage event. A rich array of chemistry can be found in the metal-dependent family of DNA damaging agents. Many readily oxidizable organic substances are capable of reducing molecular oxygen, resulting in the production of superoxide such as semiquinone radical anions, thiols,[17] and ascorbate, among others. Superoxide is well-documented[6,7,18,19] to be capable of mediating DNA strand breakage via dismutation to hydrogen peroxide and reduction of adventitious metals such as Fe(III) to Fe(II) (Haber–Weiss cycling) culminating in the reduction of hydrogen peroxide by Fe(II), generating the highly reactive hydroxyl radical (Fenton reaction).

The capacity of many antitumor antibiotics to cause oxidative damage to DNA in cancerous tissues is typically inseparable from the nonspecific damage inflicted on healthy cells by these reduced oxygen species and is widely recognized to be associated with the

† Colorado State University.
‡ University of California.

(1) Fisher, J.; Aristoff, P. A. *Prog. Drug Res.* **1988**, *32*, 411. Remers, W. A. *J. Nat. Prod.* **1985**, *48*, 173.
(2) Radiolytic decay is an important fourth family but is initiated by a nuclear event rather than a chemical event.
(3) Stubbe, J.; Kozarich, J. W. *Chem. Rev.* **1987**, *87*, 1107. Rabow, L. E.; Stubbe, J.; Kozarich, J. W. *J. Am. Chem. Soc.* **1990**, *112*, 3106. Rabow, L. E.; McGall, G. H.; Stubbe, J.; Kozarich, J. W. *J. Am. Chem. Soc.* **1990**, *112*, 3203.
(4) Natrajan, A.; Hecht, S. M.; van der Marel, G. A.; van Boom, J. H. *J. Am. Chem. Soc.* **1990**, *112*, 4532. Hecht, S. M. *Fed. Proc.* **1986**, *45*, 2784.
(5) Sausville, E. A.; Peisach, J.; Horwitz, S. B. *Biochemistry* **1978**, *17*, 2740.
(6) Graf, E.; Mahoney, J. R.; Bryant, R. G.; Eaton, J. W. *J. Biol. Chem.* **1984**, *259*, 3620. Graf, E.; Empson, K. L.; Eaton, J. W. *J. Biol. Chem.* **1987**, *262*, 11647. Ursini, F.; Mariorino, M.; Hochstein, P.; Ernster, L. *Free Radical Biol. Med.* **1989**, *6*, 31. Halliwell, B.; Gutteridge, J. M. C.; Aruoma, O. I. *Anal. Biochem.* **1987**, *165*, 215. Walling, C. *Acc. Chem. Res.* **1975**, *8*, 125.
(7) Sutton, H. C.; Winterbourn, C. C. *Free Radical Biol. Med.* **1989**, *6*, 53.
(8) Tullius, T. D.; Dombroski, B. A.; Churchill, M. E. A.; Kam, L. *Methods Enzymol.* **1987**, *155*, 537. Tullius, T. D. *TIBS* **1987**, *12*, 297. Tullius, T. D.; Dombroski, B. A. *Science* **1985**, *230*, 679. Latham, J. A.; Cech, T. R. *Science* **1989**, *245*, 276.
(9) Hertzberg, R. P.; Dervan, P. B. *Biochemistry* **1984**, *23*, 3934. Dervan, P. B. *Science* **1986**, *232*, 464. Dreyer, G. B.; Dervan, P. B. *Proc. Natl. Acad. Sci., U.S.A.* **1985**, *82*, 968. Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645. Chu, B. C. F.; Orgel, L. E. *Proc. Natl. Acad. Sci., U.S.A.* **1985**, *82*, 968. Vary, C. P. H.; Vournakis, J. N. *Proc. Natl. Acad. Sci., U.S.A.* **1984**, *81*, 6978. Kean, J. M.; White, S. A.; Draper, D. E. *Biochemistry* **1985**, *28*, 1054. Tanner, N. K.; Cech, T. R. *Nucleic Acids Res.* **1985**, *13*, 7759. Lin, S. B.; Blake, K. R.; Miller, P. S.; Ts'o, P. O. P. *Biochemistry* **1989**, *28*, 1054.

0002-7863/92/1514-733$03.00/0