

Biocatalytic Synthesis of Polymers of Precisely Defined Structures

Timothy J. Deming*, Vincent P. Conticello[†] and
David A. Tirrell

*Department of Polymer Science and Engineering
University of Massachusetts at Amherst
Amherst, MA 01003 USA*

* Department of Materials, University of California at Santa Barbara, Santa Barbara, CA

[†]Department of Chemistry, Emory University, Atlanta, GA

The fabrication of functional nanoscale devices requires the construction of complex architectures at length scales characteristic of atoms and molecules. Currently microlithography and micro-machining of macroscopic objects are the preferred methods for construction of small devices, but these methods are limited to the micron scale. An intriguing approach to nanoscale fabrication involves the association of individual molecular components into the desired architectures by supramolecular assembly. This process requires the precise specification of intermolecular interactions, which in turn requires precise control of molecular structure.

Organic polymers offer several advantages as materials for the construction of small-scale devices, including ease of synthesis and fabrication, well-delineated structure-property correlations, and thermal and mechanical stability. However the ability to precisely define polymer architecture is severely restricted for all but the most simple sequences. Macromolecular structure is defined in terms of four architectural variables: molecular size, composition, sequence, and stereochemistry. The degree to which these variables can be controlled depends on the method of polymer synthesis; conventional methods of polymer synthesis afford only statistical control of each. Consequently most synthetic polymers are not pure substances but instead comprise heterogeneous populations of molecular species, and significant advances in the solution and solid state properties of polymeric materials have been associated with increased degree of control over one or more of the architectural variables. For example, the stereoregular polymerization of α -olefins by the Ziegler-Natta procedure revolutionized polymer materials science by providing high melting, crystalline materials from simple, inexpensive building blocks.¹ In general, the synthesis of complex architectures is limited by the few available methods for the synthesis of precisely defined macromolecular architectures.

Polymers of uniform structure are synthesized by either of two techniques: iterative coupling of selectively activated monomers² or template-directed polymerization.³ The former process consists of stepwise assembly of the desired material via a repetitive sequence of intermolecular coupling and activation steps (Scheme I). Two reactive endgroups are coupled intermolecularly, and one of the remaining endgroups is selectively deprotected. The process is

repeated with addition of new reactants and the desired molecule is assembled sequentially. The serial nature of this process ensures that the microstructure of the polymer is determined by the identity of the individual reactants at each step. This process has found application in the assembly of complex structures both in solution and on solid supports. The former has been used, for example, in the synthesis of dendritic macromolecules,⁴ and the latter procedure forms the basis for the Merrifield synthesis of polypeptides.⁵ However, owing to the linear nature of the process, the degree of synthetic difficulty increases geometrically with the length (and thereby the complexity) of the sequence. This feature, coupled with the lack of an intrinsic proof-reading mechanism, limits this process to the synthesis of moderate length polymers, e.g., sixty residues for a polypeptide sequence.

In contrast to the iterative coupling procedure, template-directed synthesis can provide an intrinsic proof-reading capacity, and therefore a self-correction mechanism. In this procedure, a master template is used to specify the exact sequence of the target polymer, which is then assembled in a parallel process from the component monomers (Scheme II). The template can direct the synthesis of many copies of the target polymer by dissociation of the complementary polymer and repetition of the process. This procedure is currently limited in scope by the availability of suitably designed templates, which ideally must be uniform polymers themselves, and by formulation of appropriate mechanisms for the transfer of sequence information from the template to the reactive monomers. The best realized example of this process is protein biosynthesis, in which a DNA sequence serves as a template for the synthesis of the polypeptide chain through the intermediacy of a complementary mRNA sequence (Figure 1).

The sequence information specifying any given protein is encoded in a particular DNA sequence in the host organism. Under appropriate conditions, a copy of the coding sequence is transcribed as a messenger RNA (mRNA). The triplet codons of the mRNA, each specifying a particular amino acid residue, are then sequentially decoded on the ribosome by specifically amino-acylated transfer RNA molecules. This procedure serves as the mechanism for coupling the DNA sequence information to that of the polypeptide. Near-absolute specificity of all of the

aforementioned structural parameters, e.g., size, composition, sequence, and stereochemistry, is ensured for the nascent polypeptide not only by the template polymerization mechanism but also by simultaneous proof-reading steps that occur during the transcription/translation process. Thus polypeptides derived from genetic templates may be considered as model uniform polymers.

Polypeptides also differ from most synthetic polymers in that they adopt ordered three-dimensional structures in solution and in the solid state. Most naturally occurring proteins adopt globular structures consistent with their roles in substrate recognition and transformation, for example, as enzymes. However a number of naturally occurring proteins have fibrous structures and serve as structural components in *vivo*. These proteins are essentially polymeric in the classical sense, consisting of tandem repeats of oligomeric peptide sequences. Yet these proteins are synthesized by the same mechanism and exhibit the same uniformity of structure as their more complex globular counterparts. In addition, these polypeptides also adopt regular and persistent secondary structures in both the solution and solid states, consistent with the geometrical requirements of their oligopeptide repeats.

By utilizing the principles of protein structure and the concepts of material science as guides, non-natural protein-based materials can be designed that are capable of self-assembly into unique two- or three-dimensional shapes on the basis of their primary structures. The sequence of a polypeptide chain contains sufficient information to direct its specific assembly into a particular native conformation. The driving force for this process is the non-covalent, primarily hydrophobic interaction occurring between segments of the polypeptide chain. These interactions act cooperatively to direct the formation and association of secondary structure elements, e.g., α -helix, β -sheet, and reverse turn, into the native structure. The known preferences for the formation of specific secondary structures by particular amino acid sequences provide a pattern for the *de novo* design of protein materials. Novel combinations of secondary structure can be programmed into the polypeptide sequence to generate unique geometries in the folded protein. These dimensionally-defined protein materials may serve as subunits for the

hierarchical construction of nanoscale objects and devices. In addition, a diverse repertoire of functional groups may be incorporated into these materials, comprising the side chain functionalities of both natural and non-natural amino acid residues, providing geometrically-defined sites for further interaction between individual protein subunits.

Artificial protein sequences can be synthesized with absolute **uniformity** of structure by utilizing the techniques of recombinant DNA technology and bacterial protein expression (Figure 2).^{6,7} A fundamental oligopeptide repeat, designed for its ability to adopt a particular fold in the **final** material, is encoded into a double-stranded DNA sequence. The oligonucleotide repeat is generated by solid-state synthesis and then sequenced to verify its integrity. Ligation of the ends of the oligonucleotide in a specific head-to-tail fashion affords a population of DNA multimers that contain tandem repeats of the **coding** sequence. Fractionation generates the target length multimer (the artificial gene), which is inserted into a **plasmid**, an autonomously replicating, extrachromosomal circular genetic element. The **plasmid** also contains a gene that confers resistance to a particular antibiotic and thus provides a method for selection of transformed cells. The recombinant **plasmid** is then introduced into a bacterial strain (most typically *Escherichia coli*) that is capable of expression of the target protein. When cultured in the presence of an appropriate antibiotic, only bacteria that produce the products of the recombinant **plasmid** survive and grow. The transformed bacteria are grown to a particular cell density and target protein production is induced. The cells are harvested and lysed, and the target protein is then purified **from** the host proteins. Yields of up to grams of protein per liter of culture have been obtained in this manner.

Methods of this kind have now been used to prepare natural structural proteins as well as artificial proteins that have no direct parallel in nature. The remaining sections of this chapter summarize the state of the art of bacterial expression of protein-like polymers of potential utility in nanoscale fabrication of structures and devices.

EXPRESSION OF STRUCTURAL PROTEINS

Introduction

Natural structural proteins, e.g. silk, collagen, elastin, and keratin, are ubiquitous and essential to the proper function of all organisms. Natural evolutionary processes have yielded proteins that surpass the performance of man-made materials, e.g. mammalian elastin in the cardiovascular system that lasts half a century without loss of function, and spider webs composed of silk threads that are tougher than any synthetic fiber.⁶ There has been much recent interest in understanding and producing these natural polymers to make products for commercial applications. Purification of the biological materials from natural sources can be complicated or impossible in some cases. For example, adhesive proteins used by marine organisms to attach themselves to surfaces are irreversibly cross-linked upon secretion and cannot then be processed or reused.⁸ Isolation of structural proteins from biological sources (e.g. collagen) is practiced on a large scale commercially, yet the physical properties of those extracted materials are often less than optimal. Furthermore, problems in processing these materials combined with inconsistent quality and limited sources act to restrict commercial use of natural structural proteins to a few applications (e.g. photographic gelatin and cosmetic fillers).

As an alternative method to achieve production of large quantities of well-defined structural proteins, many researchers have turned to biocatalytic polymer synthesis using genetic engineering.^{6,7} Two approaches to the synthesis of natural polymers through this technique are (i) to clone the complementary DNA which encodes the protein of interest from the source organism and then express the resulting gene and (ii) to utilize or create a consensus repeat from highly repetitive proteins to construct a corresponding DNA sequence ("monomer") which is described in the previous section. This second procedure has been widely employed since it avoids the need to obtain cDNA clones and since it permits selection of codons to avoid unstable repetitive sequences and adaptation of codon usage to the requirements of the host organism.

6/15/95

Many natural structural proteins or artificial proteins designed to **mimic** natural materials have been successfully expressed in microorganisms and one such product has been commercialized.⁶

Silks

The silk fibroin of the common silk caterpillar *Bombyx mori* consists of extensive beta strands which hydrogen bond to form **sheets**.⁹ As such, the protein chains are very densely packed and are also highly oriented in silk fibers resulting in **high** tensile strength. *B. mori* silk fibroin primarily consists of tandem repeats of the sequence (GlyAlaGlyAlaGlySer-) and this has been the repeat sequence targeted for production of materials having the properties of silk fibers. Joseph Cappello and coworkers at Protein Polymer Technologies, Inc. (PPTI) have reported the expression in *E. coli* of high molecular weight (40-100 kDa) proteins that incorporate blocks of the *B. mori* silk hexapeptide repeat.⁶ These recombinant proteins were shown to be **crystalline** with structures **similar** to the beta sheet structures found in natural silk. Fibers prepared from these samples were found to contain beta sheet structures but no molecular orientation or alignment was seen for any of the samples. This is a limitation which must be overcome to attain and assess the ultimate physical properties of these silk-like polymers.

Sequences derived from elastin (ValProGlyValGly-) or from the cell-binding portion of fibronectin (-ArgGlyAspSer-) have also been incorporated into silk-like polymers.⁶ The result is that the elastin blocks lower the crystallinity of the polymers and enhance their solubility and processability. The incorporation of the fibronectin sequence into silk-like polymers creates a useful substrate material for cell culture. This polymer can be coated on plastic or glass surfaces and subjected to sterilization at high temperature and pressure without loss of activity. This material has been prepared in quantities greater than 500g and has been commercialized by PPTI as ProNectin F[®].⁶ Progress has also been made in the cloning and expression of fragments of the silk proteins **from** midges (*Chironomus tentans*) and spiders. Midge larvae spin insoluble fibers consisting in part of a family of silk proteins designated spI, which contain ca. 150 copies of an 82 residue core repeat. The spI are very large proteins, *ca.* 1MDa. Steven Case and his colleagues at

6/15/95

the University of Mississippi have prepared synthetic oligonucleotides which encode the 82 residue core repeat, and have reported successful expression of these oligonucleotides in *E. coli* under control of a bacteriophage T7 promoter.¹⁰ The protein has been purified to homogeneity and is being used to study the mechanisms of fiber assembly. Spider dragline silk from the golden orb weaver (*Nepila clavipes*) is produced as fibers which are stronger than those of high-strength synthetic aromatic polyamides such as Kevlar®.¹¹ Randolph Lewis and coworkers at the University of Wyoming have cloned genes encoding various spider silk proteins and attempts to express these proteins in *E. coli* are in progress.¹²

Collagen

Collagen makes up as much as one fourth of all the proteins in mammals. It is the major constituent of connective tissues and thus acts as the main load bearing, soft tissue component. Collagen fibers consist of bundles of triple helical proteins held together through extensive hydrogen bonding. The individual strands of the triple helices are constructed from many repeats of the tripeptide sequence (-GlyProXxx-) where Xxx can be any amino acid, but is usually proline or trans-4-hydroxyproline. The proline residues favor extended conformations of the polypeptide chains which are further stabilized by the intertwining of two additional protein strands. The glycine residues allow close packing of the three strands favoring formation of intrastrand hydrogen bonds and close packing of triple helical bundles.

Ina Goldberg and coworkers at Allied Signal, Inc. have prepared artificial genes which encode repetitive proteins (-GlyProPro-)_n analogous to natural collagen sequences.¹³ The synthetic genes have been cloned in *E. coli* using the thermally inducible promoter, λ pL. A 22 kDa polypeptide was successfully synthesized, but unless the heat-shock response of the *E. coli* host is altered, the protein is proteolytically degraded. The protein can be stably produced in mutants that have impaired protein degradation systems. Cappello and coworkers at PPTI have also created artificial genes based on human collagen for preparation of collagen-like polymers.⁶ In *E. coli*, they have

6/15/95

successfully expressed these genes to prepare polymers ranging in size from 40 to 70 kDa. The properties of these materials have not been extensively explored to date.

Elastin

The 72 kDa protein tropoelastin is a precursor of the protein elastin, which provides the elastic properties of mammalian arteries, lungs, skin, and other tissues.¹⁴ The tropoelastin sequence consists of tandem repetitions of a number of oligopeptide blocks which are four to nine amino acids in length. A synthetic analog of one of these blocks has been elaborated into a polymeric material which exhibits extraordinary elasticity. This material contains the block (-ValProGlyValGly-) which forms a type II beta reverse turn around the proline-glycine dipeptide.¹⁵ The amino acids not participating in the turn span the distance between turns, with glycine acting as a spacer of high flexibility. When the turns are linked together they wind into a spiral that functions as a molecular spring, resulting in a material that can be extended to over 300% of its resting length with no deformation.¹⁵

A tropoelastin cDNA has been cloned and expressed in *E. coli* by Joel Rosenbloom and his colleagues at the University of Pennsylvania.¹⁶ Since the free polypeptide is rapidly degraded in *E. coli*, it is necessary to express a fusion protein of tropoelastin linked at its N-terminus to an 81-amino acid fragment of influenza virus NS 1 protein. CNBr digestion can then be used to isolate the tropoelastin portion if a methionine codon is included at the junction of tropoelastin with the virus protein. (Tropoelastin contains no internal methionine residues.) The purified recombinant tropoelastin cross reacts with antibodies to elastin-like peptides and exhibits chemotactic activity toward fetal calf ligament fibroblasts.¹⁶ Recently McPherson, et al.¹⁷ have described the biosynthesis of an analogue of the core repeat of mammalian elastin, [Gly-(Val-Pro-Gly-Val-Gly)₁₉-Val-Gly-Pro-Gly], in *E. coli*.¹⁸ The target protein was isolated in high purity and exhibited spectroscopic properties identical to those of chemically synthesized samples.

Adhesive Proteins

Many marine organisms (e.g. barnacles, mussels, and tube worms) secrete adhesive proteins for surface attachment or for physical protection.⁸ These adhesives are remarkable in that they can be spread on a surface immersed in salt water and form a permanent bond within seconds. The bond will also withstand a wide range in temperature and large fluctuations in tidal currents. Since many man-made adhesives are ineffective in wet environments, the development of commercial adhesives based on these proteins is an attractive option. Although marine adhesive proteins from different organisms have different amino acid sequences, they all have a high percentage of DOPA (3,4-dihydroxyphenylalanine) residues which are present in repetitive sequences of 7 to 10 amino acids in length.⁸ The DOPA residues are created by post-translational hydroxylation of tyrosine in precursor proteins and are thought to be primarily responsible for the adhesive properties of the polymers. In *situ* enzymatic oxidation of the DOPA residues to quinones is thought to result in cross-linking of the materials and their adhesion to surfaces.¹⁸

Robert Strausberg and coworkers at Genex Corporation have isolated a cDNA clone from the marine mussel *Mytilus edulis* that consists of 20 repeats of (-AlaLysProSerTyrProProThrTyrLys-), the putative adhesive sequence. They have expressed this protein and several of its multimers, e.g., dimers, trimers and tetramers, in *Saccharomyces cerevisiae* by using an expression vector containing a promoter comprised of portions of the *S. cerevisiae* GAL1 and MF-a1 promoters.¹⁹ This expression system is efficient enough to produce the recombinant proteins in amounts corresponding to about 3-5% of the total cell protein. These proteins are all insoluble; the largest having a molecular weight of *ca.* 96 kDa. The purified precursor proteins do not show adhesive properties until they are treated with a bacterial tyrosinase, which converts tyrosine residues to DOPA.

A synthetic gene encoding an analog of the mussel adhesive protein has been constructed by Anthony Salemo and Ina Goldberg of Allied Signal. A DNA sequence encoding 20 copies of the decapeptide adhesive sequence was constructed by polymerization of a 30-base pair DNA monomer; the codons were chosen to correspond to *E. coli* use patterns and expression was controlled by the

6/15/95

bacteriophage T7 promoter.²⁰ The resulting recombinant protein accumulated in intracellular inclusion bodies and could be isolated in amounts approaching 60% of total cellular protein. When this material is mixed with mushroom tyrosinase, it forms an adhesive which can be spread on surfaces and used for cell attachment and growth. The adhesive has also been shown to be effective in forming a gas permeable seal when used as a cement for wet ocular tissue samples.²¹

Viral Proteins

Human adenoviruses are believed to attach to host cells via fibrillar protein spikes located on the surface of the viral capsid.²² The 200 Å long spikes consist of a repetitive 62 kDa protein organized into a trimeric array which is about 30 Å in diameter. The secondary structure of the protein has not yet been determined although triple helical and cross-beta models have been proposed. The potential for producing polymers which could be processed into high performance fibers through self-assembly makes the synthesis of analogues of these viral spike proteins an attractive prospect.

The 35 kDa fibrous protein from adenovirus serotype 3 has been expressed in *E. coli* by Corinne Albiges-Rizo and Jadwiga Chroboczek of the European Molecular Biology Laboratory. The product is a fusion protein containing at its N-terminus a short leader sequence derived from the cloning vector based on bacteriophage T7.²³ The recombinant protein was insoluble and appeared as a trimer when subjected to gel electrophoresis under denaturing conditions. Since the naturally occurring viral protein and the recombinant protein exhibit similar gel filtration behavior, Albiges-Rizo and Chroboczek have suggested that the native viral fiber also consists of a trimeric protein assembly.²³

John O'Brien and coworkers at E. I. du Pont de Nemours and Co. have expressed synthetic analogs of viral spike proteins in bacterial systems.²⁴ Oligonucleotide sequences encoding three synthetic analogs consisting of polypeptide repeats of 15 residues were prepared, multimerized and cloned in *E. coli*. The polypeptide repeat sequences were designed to contain reverse beta turns of either 4 or 5 residues and two short (three residue) beta strands. The resulting expressed proteins formed inclusion bodies and ranged in molecular weight from 20 to 100 kDa. All of the

6/15/95

proteins were soluble in hexafluoroisopropanol (HFIP). Most interesting is the observation that HFIP solutions of several of these polymers were birefringent, suggesting that liquid crystalline phases were formed. Fibers were drawn from the anisotropic solutions and were shown to have mechanical properties similar to those of textile fibers in commercial use.

Coiled-coil Proteins

A family of polymers analogous to the protein transcription factors GCN4, Fos and Jun has been expressed in bacterial systems by Kevin McGrath and David Kaplan of the U. S. Army Natick Research, Development and Engineering Center.²⁵ The transcription factors form coiled-coil structures consisting of two intertwined helices with a superhelical pitch of ca. 140 Å.²⁶ The amphiphilic helices assemble with the hydrophilic residues facing the solvent and helping to stabilize the assemblies by forming interchain electrostatic interactions; the hydrophobic residues are located on the inside of the coil. Heterodimerization of the protein chains is encouraged by the specific design of the artificial polypeptide sequences. Although these studies are at an early stage, the project has the potential for creating materials capable of self-assembly into complex structures and intermolecular recognition.

DE NOVO DESIGN AND SYNTHESIS OF WELL-DEFINED POLYPEPTIDES

Design Methodology

Previous research in this area initially had two major goals. The first was to evaluate the feasibility of using genetic engineering to create proteins with materials potential. Natural proteins known to have excellent materials properties were purposely set aside in favor of testing the potential of creating new proteins *ab initio*. These proteins would be designed from first principles utilizing structural elements found in natural proteins and their simpler analogs, namely beta strands, reverse turns and alpha helices. The second key objective was to assess the extent to which protein chain folding and supramolecular organization could, in fact, be controlled at the molecular level. Judicious

utilization of the compositional and sequential control available through biocatalytic synthesis should yield materials with precisely-defined structures.

Two classes of proteins have been featured in this work (Table 1). One consists of periodic chain-folding proteins designed to form crystalline lamellar solids. Nearly all flexible, stereoregular polymers crystallize in the form of lamellar aggregates in which the chain is oriented normal (or nearly normal) to the lamellar plane and folds regularly at the lamellar surface.²⁷ The thickness of such crystals is determined primarily by the kinetics of crystallization. Through proper sequence design, it has become possible to construct polymer chains from elements that are known to form antiparallel beta sheets and that allow control of the thickness and surface chemistry of lamellar polymer crystals. The second class of materials under study consists of proteins predicted to form alpha helices. Poly(α ,L-glutamic acid) (PLGA) and related polymers have played a central role in investigations of the physical chemistry and materials science of chain molecules.²⁸ Studies of this rod-like molecule are complicated, however, by the polydispersity of chain lengths obtained when this polymer is prepared synthetically. Our research efforts have concentrated on the production of monodisperse variants of PLGA and poly(α ,L-aspartic acid) to overcome these synthetic limitations.

More recent research in this field has branched into an additional area which serves extend the original goals. This new focus, and the subject of the next section, concerns the use of intact cellular protein synthesis machinery to incorporate *unnatural* amino acids into some of the artificial proteins described above. Unnatural amino acids (or amino acid analogs) allow the incorporation of unusual atoms and functional groups (e.g. fluorine and electroactive groups) into proteins, and extend the range of properties obtainable for protein materials.

Folding Beta Sheet Proteins

The formation of folded-chain lamellar crystals is common among polymers of regular chemical structure. For most such materials, however, the folded-chain structure is determined by the kinetics, rather than the thermodynamics, of the crystallization process, and the observed lamellar organization is metastable. The opportunity to design sequences of controlled periodicity

offers the prospect of thermodynamic control of regular chain folding. Our approach has been to attempt to exploit known sequence-dependent secondary structures (e.g., beta strands and reverse turns) to design folded-chain lamellar crystals of controlled thickness and surface functionality. As shown schematically in Figure 3, the thickness of the crystal is controlled by the length of the extended element (the beta strand), while the surface functionality is determined by the composition of the turn sequences.

Glycine-alanine (Gly-Ala) dyads are featured in the stem segments of our periodic proteins as this motif is known to form beta strands. GlyAla repeats are responsible for the beta sheets characteristic of the silk-II crystal morphology as well as the beta strands in type-I poly-AlaGly.²⁹ One of the design variables under study is the effect of the stem length on lamellar dimensions (*vide infra*). A variety of turn structures is also under study.

The first proteins to be constructed contained proline and glutamic acid in the turn sequences since proline is a well known turn-former and the bulky, polar glutamic acid residue should be excluded from the crystalline interior by both steric interactions and solvent-binding effects.³⁰ Crystalline solids have been obtained for several sequences of this type (Table 1), but the formation of regular lamellae is dependent on the size of the basic repeat unit (*vide infra*). A library of genes encoding several different two-amino acid turn sequences has been constructed, based on sequences reported for the type-I' and type-II' reverse turns found in natural proteins. The anticipated solid-state organization of these polymers would thus be characterized by regular folding of the chain at the turn sequences, with the glycylalanine beta sheets stacked in the lamellar interior and the turn sequences exposed at the surface. The first repetitive copolypeptides that we designed were multimers of either the nonapeptide [(GlyAla)₃GlyProGlu]⁷ 1 or the undecapeptide [(GlyAla)₄GlyProGlu],³¹ 2. Chain-length variants containing 10-54 repeats of the above sequences have been efficiently produced in *E. coli* using a bacteriophage T7-based expression system. Although several lines of evidence indicate that the expressed proteins have the correct sequence, such materials do not readily adopt regularly folded chain arrangements, but instead form amorphous glasses at room temperature.

In a second attempt at sequence-controlled crystallization, a gene encoding thirty-six repeats of the octapeptide [(GlyAla)₃GlyGlu] **3** was expressed in *E. coli* as a fusion protein containing short flanking sequences derived from the cloning and expression vectors.³² The target protein was confined to the soluble fraction of the whole cell lysate, and a simple procedure involving sequential pH adjustments with glacial acetic acid afforded a substantial enrichment in the supernatant. The product was isolated by ethanol precipitation at -10 C, and the periodic portion of the target protein was liberated from the flanking sequences by cleavage with cyanogen bromide. The predominance of the antiparallel beta sheet structure in the crystalline polypeptide is revealed in the key features of the infrared (IR), Raman, and cross polarization magic angle spinning nuclear magnetic resonance (CP/MAS NMR) spectra of samples crystallized from 70% formic acid.³² The IR spectrum of **3** exhibits amide I, II, and III vibrational modes at 1623, 1521, and 1229 cm⁻¹, respectively, characteristic of the beta sheet conformation, and the weak amide I component at 1698 cm⁻¹ indicates the regular alternation of chain direction that defines the antiparallel beta sheet architecture. The Raman spectrum of **3** exhibits the amide I band at 1664 cm⁻¹ and the splitting of the amide III band into two components at 1260 and 1228 cm⁻¹, characteristic of the antiparallel beta sheet conformation. Likewise, the CP/MAS ¹³C NMR spectrum of **3** showed chemical shifts analogous to those obtained for poly(L-AlaGly) in the beta form.³²

Wide-angle x-ray diffraction patterns of crystal mats of **3** were obtained with the x-ray beam parallel to the plane of the mat. The diffraction patterns exhibit discrete Bragg reflections consistent with a crystalline polymer and the reflections index on an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.060$ nm, and c (chain axis) = 0.695 nm. These unit cell parameters are commensurate with previously published x-ray diffraction results from various silk fibroins and synthetic polypeptides that exhibit similar crystalline structures.³³ In addition, well-defined meridional reflections were observed which correspond to a long period spacing of 36 Å. This result appears to be completely consistent with the expected lamellar thickness, since molecular modeling predicts a spacing of ca. 33 Å between the termini of glutamic acid side chains arrayed on opposite faces of crystals built up from

stacked, folded beta sheets of 3 (Figure 4). A long-period spacing of 36 Å would thus correspond to an interlamellar separation of ca. 3 Å, which appears plausible. Electron microscopy of **3** reveals individual lamellae ca. 1 μm in length, generally similar in appearance to crystals grown from conventional nylons.³² All of the experimental evidence collected on the solid-state structure of 3 (and other stem length variants of 3) strongly supports a crystalline antiparallel beta sheet architecture involving a chain-folded lamellar structure as the basic crystalline unit.

Other stem-length and turn sequence variants of 3 are also under investigation. One variant in particular is the sequence [(AlaGly)₃GluGly(GlyAla)₃GluGly], 4, which was synthesized to probe the chain folding pattern in the polymeric crystalline domains.³⁴ None of the diffraction analyses mentioned above directly addresses the nature of chain folding, yet successful assembly of the predicted lamellar structures depends on the ability of the chains to fold and crystallize in an adjacent reentry fashion. Polymers comprising 10 repeats of sequence 4 were expressed in *E. coli* under conditions where ¹³C-enriched glycine and alanine could be incorporated into the chains. Infrared spectra of blends of natural-abundance and ¹³C-enriched polymers of sequence 4 showed two bands in the polypeptide amide I region (1629 and 1596 cm⁻¹) assignable to hydrogen-bonded beta sheets of the isotopically distinct samples. It has been suggested that the frequency shift of the unperturbed amide I vibration is due to transition dipole coupling between adjacent carbonyls within the beta sheets. Hence, the presence of separate ¹³C and ¹²C derived amide I bands (instead of a composite band) led us to propose that a majority of the chains are folding in an adjacent reentry fashion as predicted?

Monodisperse Alpha Helical Proteins

Polymers that adopt a persistent rodlike structure in solution and in the solid state have been a topic of much current research in materials science.³⁵ These materials should form ordered solutions due to the shape anisotropy of the individual rods thereby facilitating their fabrication into oriented solid materials. Polypeptides that adopt a persistent α-helical secondary structure should display this behavior for structures that have a critical aspect ratio of rod length to diameter. Indeed, studies of poly(α,L-glutamic acid) (PLGA) and its ester derivatives have

played a central role in elucidating the physical chemistry and materials science of rodlike polymers.²⁸ PLGA itself has been used in fundamental studies of the helix-coil transition of polyelectrolytes,³⁶ and ester derivatives of PLGA most notably poly(γ -benzyl- α ,L-glutamate) (PBLG) form organic liquid crystalline solutions³⁷ and oriented monolayer films.³⁸ In addition, α -helical rods should have a large dipole moment associated with the peptide bond that is essentially oriented along the helical axis.³⁹ The dipoles of individual PBLG helices can be oriented in solution under the influence of an electric or a magnetic field affording ferroelectrically ordered mesophases and films.⁴⁰ These materials display properties associated with oriented acentric materials such as piezoelectric behavior and second order non-linear optical properties.

The traditional synthetic route to PLGA and its esters involves the ring-opening polymerization of *N*-carboxy- α -amino acid anhydrides. This technique affords a heterogeneous population of chains that have relatively broad distributions of molecular sizes. This heterogeneity of molecular lengths complicates the assessment of the critical structural criteria for the observation of liquid crystallinity in rodlike polymers, which remains an open question. The utility of these materials in the design and construction of complex macromolecular architectures also is limited by this structural heterogeneity. The incorporation of more complex sequences of residues in a rational manner is also clearly precluded via this synthetic approach due to the inability to accurately specify the positions of individual residues.

Uniform length α -helical rods of a PLGA analogue [GluAsp(Glu₁₇Asp)₄GluGlu], 5, have been prepared via a genetically directed synthesis.⁴¹ The periodic aspartic acid residues provide recognition and cleavage sites for the restriction enzyme *Bbs*I, which is used to liberate the coding sequence for head-to-tail oligomerization of the oligonucleotides. Aspartic acid was chosen as the second residue because of its similarity to glutamic acid, which is expected to reduce to a minimum any perturbation of the chemical and physical properties of the chain. The gene for 5 was expressed as a fusion protein to the affinity label glutathione-S-transferase, which enabled a convenient separation from the bacterial host proteins by chromatography. The target polypeptide was liberated from the fusion protein by either enzymatic proteolysis or cyanogen

6/15/95

bromide cleavage. The PLGA analogue **5** displays chemical and spectroscopic properties consistent with its expected structure, in particular, a pH dependence of its circular dichroism spectrum indicative of a helix to coil transition upon deprotonation at alkaline pH. In addition the electrophoretic behavior of **5** compares favorably with commercial samples of narrow polydispersity (polydispersity index ≈ 1.2), chemically synthesized PLGA. The latter samples appear as broad smears while the former appears as a single, tightly focussed band, thus illustrating the power of the biosynthetic strategy for the control of macromolecular architecture.

Monodisperse **5** has been cleanly converted to its benzyl ester derivative **6** by alkylation of the carboxylate groups with phenyldiazomethane. GPC analysis of the product indicates that the uniformity of **5** is retained during the benzylation. Spectroscopic analysis of **6** is consistent with chemically synthesized samples of PBLG. The uniform polyglutamates should generate unique liquid crystalline phases because the polymer's uniform chain length should allow it to stack into a phase with both orientational and positional order. In contrast, the polydisperse polymer lacks positional order and forms a more limited range of liquid crystalline structures. Solutions of **6** in benzyl alcohol form optically birefringent phases characteristic of lyotropic liquid crystalline solutions of rodlike polymers. The effect of the uniformity of the polypeptide on the degree of order displayed by the liquid crystalline phases is currently under investigation.

INCORPORATION OF ARTIFICIAL AMINO ACIDS INTO PROTEINS

Methodology

While biological synthesis offers superior control over polymer chain architecture, including purity of sequence, size and stereochemistry, its versatility is normally limited to the twenty amino acids that appear in natural proteins. Thus, an important aim for the future will be to expand the number of amino acids that can be utilized in protein biosynthesis. First efforts in this direction would logically start with amino acid analogs known to be incorporated during translation.⁴² The unnatural amino acids (or amino acid analogs) employed are those which are in many ways similar to a

6/15/95

natural amino acid, yet differ slightly in structure (thienylalanine vs. phenylalanine), functionality (hydroxyproline vs. proline), or constituent atoms (selenomethionine vs. methionine) (Tables 2 and 3).

The biosynthesis of proteins containing non-natural amino acids has been approached in three different ways. One approach involves attaching a desired substituent to a suppressor tRNA that decodes a nonsense codon in an appropriate location in the gene of interest.⁴³ This approach allows the usual restrictions on enzymatic charging of tRNA to be by-passed; it also permits the non-natural monomer to be introduced into a specific site. However, only sub-milligram quantities of protein are likely to be obtained, since chemical acylation is experimentally difficult and the tRNA is not recycled. Furthermore, since the efficiency of suppression generally is 50% or less, multi-site replacement (for example, in a repetitive protein) is impractical. In a second method, the amino acid analogue is added to a cell-free protein synthesis system.⁴⁴ Only substrates that can be enzymatically activated and charged can be used, but the method avoids the difficulties of low suppression efficiency and chemical synthesis of the aminoacyl tRNA. The third approach, which provides the potential for synthesizing proteins containing many modified sites, is *in vivo* incorporation of non-natural amino acids.

It has been known for many years that certain analogs of the natural amino acids can be incorporated efficiently *in vivo* into bacterial proteins.⁴³ Success of this process is entirely dependent on the ability of the protein synthesis machinery of the cell to accept the analog in place of the natural amino acid. The artificial amino acid must first enter the cell; however, because the relevant transport processes are usually not very specific, transport is not normally limiting to analog incorporation.⁴⁵ Much more important are the aminoacyl-tRNA synthetases, which are responsible for recognition of the 20 natural amino acids and attachment ("charging") to their respective tRNAs.⁴⁶ The non-natural analogs must be similar enough to their natural counterparts to be recognized and activated by these enzymes. In addition, a certain lack of specificity is desirable in the aminoacyl-tRNA synthetase enzyme for the amino acid of interest: the less specific the enzyme, the greater the range of analogs which can be successfully incorporated.

Since the synthetase enzymes are nearly always much more effective in activating the natural amino acid than the analog, it is often crucial to decrease the amount of natural amino acid available to the cells. This requirement is met through the use of host strains which are amino acid auxotrophs, i.e., cells which require a certain amino acid (the one of interest for analog incorporation) in order to grow. Thus, our methodology for incorporating **non-natural** amino acids into artificial proteins begins with the preparation of a suitable *E. coli* **auxotroph**⁴⁷ followed by transformation of this strain with an expression vector containing the gene encoding the target protein of interest with the *natural* amino acid encoded in the positions where the analogs are to be inserted. If the amino acid analog supports cell growth, the cells can then be grown in medium lacking the natural amino acid, but supplemented with the analog of interest. Expression of the target protein is then induced and the protein is isolated by standard methods. If the analog is toxic to the cells or does not support growth, an alternate procedure is employed; the cells are grown in medium containing the natural amino acid, they are then pelleted, washed, and resuspended in medium containing the analog and **lacking** the natural amino acid, and protein expression is induced. This procedure is necessary to grow large numbers of cells for production of sufficient quantities of target proteins.

Incorporation of Selenomethionine

In a first attempt to generate artificial proteins containing non-natural amino acids, we produced a beta sheet forming protein containing **selenomethionine**.⁴⁸ The protein was expressed in *E. coli* from plasmid DNA that encoded a lamellar protein of sequence [(GlyAla)₃GlyMet]₁₁, **7**. Selenomethionine was selected for two reasons: first, it is known to support growth of *E. coli* methionine auxotrophs in the absence of externally supplied **methionine**;³⁹ second, the presence of selenomethionine endows product proteins with useful sites for chemical modification and a ⁷⁷Se NMR probe useful for mapping local folding properties. In addition, the selenium containing side chains can be utilized selectively for addition of structurally or functionally important new substituents or for site-specific linkage to other materials.

The target protein used in our study contains methionines in each putative beta-turn segment, offering the potential of selective placement of selenomethionines on the surfaces of a resulting lamellar material. This prediction is supported (but certainly not conclusively demonstrated) by x-ray scattering data obtained with the methionine containing product, showing that this material assumes a crystalline morphology.⁴⁸ An *E. coli* expression strain was first converted to a strict methionine auxotroph and shown to grow in medium in which methionine was replaced with selenomethionine; although the growth rate was slower with the analog, similar saturating cell densities were observed.⁴⁸ Electrophoretic analysis showed that the target protein was produced in cells grown on selenomethionine and that no inhibitory effects on protein production were observed. The level of selenomethionine substitution was estimated indirectly using a competition assay where the loss of radiolabeled methionine in product protein was correlated with the ratio of methionine and selenomethionine in the growth medium. The results were consistent with perfect or near-perfect replacement of methionine with selenomethionine. Plans for future work include characterization of the crystalline morphology of the analog containing material and chemical modification studies.

Fluorine containing polypeptides

Another area of interest is the incorporation of fluorinated amino acid analogs. Natural polymers lack halogens, yet many synthetic polymers which incorporate halogens have proven to be extremely useful. In particular, fluorine containing polymers display the useful characteristics of excellent solvent resistance, good hydrolytic stability, low surface energy, and low coefficient of friction.⁵⁰ We reasoned that it would be advantageous to transfer some of these properties to polypeptides to create unique materials which might prove useful for the fabrication of novel membranes, low-friction surface treatments, and biomedical materials. Furthermore, since fluorine is only slightly larger than hydrogen it is not surprising that many fluorinated amino acids are recognized and utilized by the bacterial protein synthesis machinery.⁴²

The analogs chosen for the *in vivo* synthesis of fluorinated polypeptides in our laboratory have included para-fluorophenylalanine (*pff*) and 5,5,5-trifluoroleucine (TFL), since both had been shown previously to be incorporated into *E. coli* proteins.⁴² The strategy was similar to that used for selenomethionine, with the exception that the host auxotroph strains were first grown in the presence of the natural amino acids and then protein synthesis was induced in medium containing the analog. This step was necessary because these analogs did not support growth of the *E. coli* strains used for protein expression. The artificial proteins into which the analogs were inserted were, as with selenomethionine, lamellar proteins of the type [(GlyAla)₃GlyXxx]_n, **8**, where Xxx was encoded as Phe for *pff* (n = 13) and Leu for TFL (n = 12).

Proteins containing *pff*⁵¹ or TFL⁵² were expressed and purified through standard techniques. The extent of analog incorporation was found to be very high for both materials (>95% for *pff* and 88% for TFL) as determined by ¹H NMR spectra and amino acid analyses. Both materials were crystallizable and were found to adopt antiparallel beta sheet structures similar to those obtained for other proteins of this type (*vide supra*). Thus, it is expected that the fluorinated residues, since they occupy turn positions, will be exposed at the surfaces of the polymer crystals and should affect the bulk material properties of the samples. Advancing contact angle measurements with hexadecane for the TFL polypeptides gave increasing contact angles with increasing analog content relative to the natural residues. The presence of fluorine in the samples was found to decrease the surface energy as determined by decreased wettability toward hydrocarbon liquids.⁵²

Electroactive Substituents

The scope of functional groups that can be incorporated into the protein backbone by genetic engineering has recently been extended to residues with electroactive substituents, such as 3-thienylalanine (3-TA).⁵³ 3-Alkylthiophenes, which are structurally similar to 3-TA, can be oxidatively polymerized through the 2- and 5-positions of the ring to produce polymers with extended p conjugation. Poly(3-alkylthiophene)s can be electrochemically doped to form

conducting polymers with conductivities (2000 S cm^{-1}) that are among the highest recorded for organic materials.⁵⁴ Copolymerization of the 3-TA residues with 3-alkylthiophenes can provide a method for the crosslinking (or grafting) of polypeptides onto electrodes via a layer of conducting polymer that may be useful for the fabrication of enzyme-based sensory elements.

Prior to this work, 3-thienylalanine had not been shown to be utilized by the protein synthesis machinery of *E. coli*, although the isomeric 2-TA had been incorporated into native bacterial proteins in place of phenylalanine. Therefore, the target protein selected for the incorporation assay was the sequence, $[(\text{GlyAla})_3\text{GlyPhe}]_{13}$ **9**, which contains a unique Phe residue in each repeat, which can serve as an accepting site for 3-TA. The chain-folded structure of the polypeptide could potentially control the spatial orientation of the electroactive substituents and facilitate conversion of these groups to an oriented conducting polymer. In vivo radiolabelling experiments demonstrated that the analogue-containing target protein was expressed at levels similar to the natural variant **9** in a Phe auxotroph. The extent of substitution, quantitated by UV and ^1H NMR spectroscopy, and amino acid analysis, is about 80%, suggesting that 3-TA is efficiently utilized by the *E. coli* protein synthesis machinery under these conditions. Preliminary electrochemical investigations indicate that the 3-TA-containing polypeptide can be oxidatively copolymerized with 3-methylthiophene yielding conductive films.

Structural Modifications of Polypeptides

It was mentioned earlier that a key to successful in vivo analog incorporation is lack of specificity in the aminoacyl tRNA synthetases. The prolyl-tRNA synthetase has been reported to be the least specific of all.⁵⁵ Not only does this mean that a wide variety of proline analogs should be incorporated into proteins, but proline, the only imino acid, is also a key residue for structural modification studies. The lack of a hydrogen atom on nitrogen to serve as a hydrogen bond donor, coupled with the steric bulk of the prolyl ring, in almost all cases excludes proline from involvement in α -helical and β -sheet structures.⁵⁶ Proline therefore usually occupies the key positions of helix

starter or helix terminator, and plays an active role in protein folding. It has also been reported that **proline** can act to bend the axis of a helix by *ca.* 30°. Finally, although it is a hydrophobic residue, **proline** is usually exposed at the hydrophilic surfaces of proteins by virtue of its localization in turns. This is an ideal location if reactive analogs are to be accessible for post-translational modification.

It was thought that analogs of **proline** might cause significant perturbations in the solid-state structure of polymers of the sequence [(GlyAla)₃GlyProGlu] **1** (*vide supra*). Although these polypeptides crystallize with difficulty, it was proposed that substitution of an analog for **proline** might facilitate crystallization of the AlaGly repeats into a lamellar β -sheet architecture. Of key interest was incorporation of **homologs** of **proline**, where a simple change in the size of the ring structure would cause a significant change in the conformation of the residue and thus allow chain folding. As precedent for this, it has been reported that incorporation of L-azetidine-2-carboxylic acid in place of **proline** at levels as low as 4% is sufficient to prevent triple helix formation in collagen.⁵⁷

Many analogs of **proline** have been reported and some have been demonstrated to be incorporated *in vivo* into cellular proteins.⁵⁸ These results are briefly summarized in Table 3. The bulk of studies on analogs of **proline** have focused on collagen synthesis,⁵⁹ although a few reports exist on the behavior of these analogs in bacteria.⁶⁰ The **proline** analogs used in our studies began with some of those which had been previously incorporated into proteins, namely: L-azetidine-2-carboxylic acid (Azc),⁵⁷ L- γ -thiaproline (Thp),⁶¹ and 3,4-dehydroproline (Dhp)⁶². As can be seen in Table 3, Azc, Thp, and Dhp have all been incorporated into proteins in *E. coli* cells *in vivo*.

A **proline** auxotroph of *E. coli* containing a gene encoding sixteen repeats of **1** was used for the expression of proteins containing the **analogs**.⁶³ Proteins containing varying degrees of Azc, Thp, and Dhp were expressed and purified. Dhp appeared to be utilized most efficiently with incorporation levels approaching 100%. Azc was incorporated to a lesser extent (*ca.* 40%) and Thp was incorporated to a level of around 10%. The physical properties of the Dhp variant were found to be very similar to those of the **proline** form: both formed amorphous solids and were very soluble in water. **While** the material containing Dhp is amorphous, it is an intriguing material which contains olefinic functionality which can serve as a site of specific reactivity not present in natural proteins.

Both the Thp and Azc variants were found to be water insoluble, and the Azc material fractionated differently from the parent material during purification. The decreased solubility of these polypeptides is an indication that they may be crystalline; a hypothesis supported for the Azc material by preliminary structural studies. These proteins illustrate the use of small specific variations in monomer structure to control and vary the overall conformation of the chain.

HYBRID ARTIFICIAL PROTEINS

Design Criteria

The coupling of natural and artificial domains to produce hybrid artificial proteins offers the prospect of new classes of **functionalized** nanoscale architectures. In preliminary experiments directed toward these objectives, enzymes have been linked at the genetic level to crystalline artificial proteins to promote the assembly of catalytic arrays and sensors (Figure 5). In designing materials of this type many issues relating to the composition of the natural and artificial domains must be addressed. Choice of the artificial domain in our initial experiments was simple: we sought a polypeptide which would present functional groups useful in surface assembly and attachment. Such functional groups might include the ammonium groups of lysine (which can attach to anionic surfaces), the sulfhydryl groups of cysteine (which can bind to gold surfaces), and the carboxylate groups of glutamate and aspartate (which can be attached to cationic surfaces). A polypeptide which fulfilled all of these requirements was the repeating octapeptide **3** which is expressed in high yield in *E. coli*.

The enzyme comprising the natural domain of the hybrid protein must satisfy a more complex set of requirements. The enzyme should function as a single-subunit, so that post-translational association of enzyme subunits, common in many enzymes, is not essential. It must express well in appropriate cellular hosts, and it must tolerate modification (i.e., fusion to the artificial domain) without loss of function. Our first candidate was a single chain phosphotriesterase from *Pseudomonas diminuta* which hydrolyzes many phosphorus based pesticides and nerve agents.⁶⁴

This enzyme should be useful in detection and decontamination of these toxins, it has been cloned and expressed in *E. coli*, and it was available in plasmid-encoded form at the outset of our experiments.⁶⁵

Results

A gene encoding both the 27 kDa artificial protein of sequence 3 and the 39 kDa phosphotriesterase domain was constructed and expressed in *E. coli* using a bacteriophage T7-based expression system.⁶⁶ The enzyme itself was also expressed in *E. coli* and was found to form inclusion bodies which segregate into the insoluble portion of the cell lysate. The fusion protein was found to segregate into the soluble portion of the cell lysate which should help to reduce problems involved in renaturing the enzymatic portion of the protein. Fractionation experiments on an ion-exchange column showed that the fusion protein could be effectively separated from the free enzyme based on the excess acidic sites present in the artificial domain. Furthermore, these sites were sufficient to bind the fusion protein to the column under conditions where the free enzyme is not bound, and preliminary results suggest that enzymatic activity is retained. Current work is aimed at immobilizing the fusion protein onto optical glass fibers to create biosensor devices.

References

1. Pino, P.; Moretti, G. *Polymer* **1987**,*28*, 683.
2. Tour, J.M. *Trends in Polymer Science* **1994**,*2*, 332.
3. Tan, Y.Y. *Prog. Polym. Sci.* **1994**,*19*, 561.
4. Fréchet, J.M.J. *Science* 1994,*263*, 1710.
5. Merrifield, R.B. *Science* **1986**,*232*, 241.
6. Cappello, J.; Ferrari, F. in *Plastics from Microbes: Microbial Synthesis of Polymers and Polymer Precursors*; Mobley, D.P., Ed.; Hanser/Gardner Publications: Munich, 1994; pp 35-92.

6/15/95

7. McGrath, K.P.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *J. Am. Chem. Soc.* **1992**, *114*, 727.
8. Waite, J.H. *Biol. Rev.* **1983**, *58*, 209.
9. Tsujimoto, Y.; Suzuki, Y. *Cell* **1979**, *18*, 591.
10. Case, S.T.; Powers, J.; Hamilton, R.; Burton, M.J. In *Silk Polymers*; Kaplan, D.; Adams, W.W.; Farmer, B.; Viney, C., Eds.; ACS Symposium Series 544; American Chemical Society: Washington, DC, 1994; pp 81-90.
11. Gosline, J.M.; DeMont, M.E.; Denny, M.W. *Endeavor* **1986**, *10*, 37.
12. (a) Hinman, M.B.; Lewis, R.V. *J. Biol. Chem.* 1992, *10*, 1. (b) Xu, M.; Lewis, R.V. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 7120.
13. Goldberg, I.; Salerno, A.J. In *Materials Synthesis Utilizing Biological Processes*; Reike, P.C.; Calvert, P.D.; Alper, M., Eds.; Mater. Res. Soc. Proc. 174; Materials Research Society: Pittsburg, PA, 1990, pp 229-236.
14. Sandberg, L.B.; Soskel, N.T.; Leslie, J.B. *N. Engl. J. Med.* 1981, *304*, 566.
15. Urry, D.W. *J. Prot. Chem.* 1988, *7*, 1.
16. Indik, Z.; Abrams, W.R.; Kucich, U.; Gibson, C.W.; Mecham, R.P.; Rosenbloom, J. *Arch. Biochem. Biophys.* 1990, *280*, 80.
17. McPherson, D.T.; Morrow, C.; Minehan, D.S., Wu, J.; Hunter, E.; Urry, D.W. *Biotechnol. Prog.* 1992, *8*, 347.
18. (a) Waite, J.H.; Tanzer, M.L. *Science* 1981, *212*, 1038. (b) Waite, J.H.; Housely, T.J.; Tanzer, M.L. *Biochemistry* 1985, *24*, 5010. (c) Jensen, R.A.; Morse, D.E. *J. Comp. Physiol. B* **1988**, *158*, 317.
19. Filipula, D.R.; Lee, S.-M.; Link, R.P.; Strausberg, S.L.; Strausberg, R.L. *Biotechnol. Prog.* **1990**, *6* 171.

6/15/95

20. Salemo, A.J.; Goldberg, I. *Appl. Microbiol. Biotechnol.* **1993**, *58*, 209.
21. Benedict, C.V.; Picciano, P.T. In *Adhesives From Renewable Resources*; Hemingway, R.W.; Conners, A.H.; Branham, S.J., Eds.; ACS Symposium Series 385; American Chemical Society: Washington, DC, 1989; pp 465-483.
22. Valentine, R.C.; Peseira, H.G. *J. Mol.Biol.* **1965**, *13*, 13.
23. Albiges-rizo, L.; Chroboczek, J. *J. Mol.Biol.* 1990,212,247.
24. O'Brien, J.P.; Hoess, R.H.; Gardner, K.H.; Lock, R.L.; Wasserman, Z.R.; Weber, P.C.; Salemme, F.R. In *Silk Polymers*; Kaplan, D.; Adams, W.W.; Farmer, B.; Viney, C., Eds.; ACS Symposium Series 544; American Chemical Society: Washington, DC, 1994; pp 104-117.
25. McGrath, K.P.; Kaplan, D.L. In *Biomolecular Materials*; Viney, C.; Case, S.T.; Waite, J.H., Eds.; Mater. Res. Soc. Proc. 292; Materials Research Society: Pittsburg, PA, 1993, pp 83-92.
26. O'Shea, E.K.; Kim, P.S. *Science* 1991,254,539.
27. Bassett, D.C. *Principles of Polymer Morphology*; Cambridge University Press: Cambridge, 1981.
28. Block, H. *Poly(g-benzyl-L-glutamate) and Other Glutamic Acid Containing Polymers*; Gordon and Breach: New York, NY, 1983.
29. Fraser, R.D.B.; McRae, T.P. *Conformations of Fibrous Proteins*; Academic: New York, 1973.
30. Chou, P.Y.; Fasman, G.D. *Biochemistry* **1974**, *13*,211.
31. Creel, H.S.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *Macromolecules* **1991**, *24*, 1213.
32. Krejchi, M.T.; Atkins, E.D.T.; Waddon, A.J.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *Science* 1994,265,1427.

6/15/95

33. Warwicker, J.O. *J. Mol. Biol.* **1960**, *2*, 350.
34. Parkhe, A.J.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *Macromolecules* **1993**, *26*, 6691.
35. Finkelman, H. *Angew. Chem. Int. Ed. Engl.* **1987**, *26*, 816.
36. Doty, P. *J. Polymer Sci.* **1957**, *23*, 851.
37. Horton, J.C.; Donald, A.M.; Hill, A. *Nature* 1990, 346; 44.
38. McMaster, T.C.; Carr, H.J.; Miles, M.J.; Cairns, P.; Morris, V.J. *Macromolecules* **1991**, *24*, 1428.
39. Hol, W.G.J. *Nature* **1978**, *273*, 443.
40. (a) Go, Y.; Ejiri, S.; Fukuda, E. *Biochim. Biophys. Acta* 1969, 175, 454. (b) Iizuka, E. *Biochim. Biophys. Acta* 1969, 175, 456.
41. Zhang, G.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *Macromolecules* 1992, 25, 3601.
42. (a) Wilson, M.J.; Hatfield, D.L.; *Biochim. Biophys. Acta* 1984, 781, 205. (b) Richmond, M.H. *Bacteriol. Rev.* 1962, 26, 398. (c) Hortin, G.; Boime, I. *In Methods in Enzymology*; Fleischer, S.; Fleischer, B., Eds.; Academic: New York, 1983; Vol. 96, pp 777-784.
43. (a) Noren, C.J.; Anthony-Cahill, S.J.; Griffith, M.C.; Schultz, P.G. *Science* 1989, 244, 182. (b) Bain, J.D.; Glabe, C.G.; Dix, T.A.; Chamberlin, A.R. *J. Am. Chem. Soc.* 1989, 111, 8013. (c) Mendel, D.; Ellman, J.A.; Schultz, P.G. *J. Am. Chem. Soc.* 1991, 113, 2758. (d) Mendel, D.; Ellman, J.A.; Chang, Z.; Veenstra, D.L.; Kollman, P.A.; Schultz, P.G. *Science* 1992, 256, 1798.
44. Zubay, G. *Ann. Rev. Gen.* **1973**, *7*, 267.
45. Saks, M.E.; Sampson, J.R.; Abelson, J.N. *Science* 1994, 263, 191.
46. Schimmel, P. *Ann. Rev. Biochem.* **1987**, *56*, 125.

6/15/95

47. Miller, J.H. *Experiments in Molecular Genetics*; Cold Spring Harbor: Cold Spring Harbor, NY, 1972.
48. Dougherty, M.J.; Kothakota, S.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *Macromolecules* **1993**, *26*, 1779.
49. Tuve, T.; Williams, H. *J. Am. Chem. Soc.* 1957, *79*, 5830.
50. Gangal, S.V. *In Encyclopedia of Polymer Science and Engineering*, 2nd ed.; Wiley-Interscience: New York, 1990, Vol. 16, pp 577-648.
51. Yoshikawa, E.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *Macromolecules* **1994**, *27*, 5471.
52. Kothakota, S.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. Manuscript in preparation.
53. Kothakota, S.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *J. Am. Chem. Soc.* **1995**, *117*, 536.
54. Roncali, J.; Yassar, A.; Garnier, F. *J. Chem. Soc. Chem. Commun.* **1988**, 581.
55. Papas, T.S.; Mehler, A.H. *J. Biol. Chem.* **1970**, *245*, 1588.
56. Fasman, G.D. *Prediction of Protein Structure and the Principles of Protein Conformation*; Plenum: New York, 1989, pp 48-53.
57. (a) Grant, M.M.; Brown, A.S.; Corwin, L.M.; Troxler, R.F.; Franzblau, C. *Biochim. Biophys. Acta* **1975**, *404*, 180. (b) Fowden, L.; Richmond, M.H. *Biochim. Biophys. Acta* 1963, *71*, 459. (c) Takeuchi, T.; Prockop, D.J. *Biochim. Biophys. Acta* 1969, *175*, 142.
58. Mauger, A.B.; Witkop, B. *Chem. Rev.* 1966, *66*, 47.
59. (a) Rosenbloom, J.; Prockop, D.J. *J. Biol. Chem.* 1970, *245*, 3361. (b) Uitto, V.-J.; Uitto, J.; Kao, W.W.-Y., Prockop, D.J. *Arch. Biochim. Biophys.* 1977, *185*, 214. (c) Jimenez, S.; Rosenbloom, J. *Arch. Biochim. Biophys.* 1974, *163*, 459. (d) Uitto, J.; Prockop, D.J. *Biochim. Biophys. Acta* 1974, *336*, 234. (e) Gottlieb, A.A.; Fujita, Y.; Udenfriend, S.; Witkop, B. *Biochemistry* 1965, *4*, 2507. (f) Cleland, R.; Olson, A.C. *Biochemistry* 1968, *7*, 1745. (g) Rosenbloom, J.; Prockop, D.J. *J. Biol. Chem.* 1971, *246*, 1549.

6/15/95

60. (a) Busiello, V.; Di Girolamo, M.; Cini, C.; De Marco, C. *Biochim. Biophys. Acta* 1979, 564, 311. (b) De Marco, C.; Busiello, V.; Di Girolamo, M.; Cavallini, D. *Biochim. Biophys. Acta* 1977, 478, 156.

61. Busiello, V.; Di Girolamo, M.; Cini, C.; De Marco, C. *Biochim. Biophys. Acta* 1980, 606, 347.

62. (a) Fowden, L.; Neale, S.; Tristram, H. *Nature* 1963, 199, 35. (b) Smith, L.C.; Ravel, J.M.; Skinner, C.G.; Shive, W. *Arch. Biochim. Biophys.* **1962**, 99, 60.

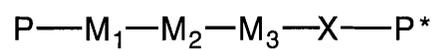
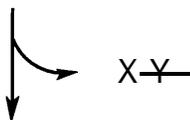
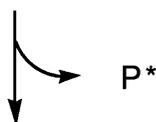
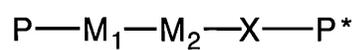
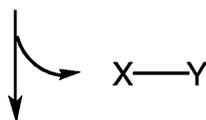
63. Deming, T.J.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *ACS Polym. Mat. Sci. Engr.* 1994, 71, 673.

64. Dumas, D.P.; Rauschel, F.M. *J. Biol. Chem.* 1990, 265, 21498.

65. McDaniel, C.S.; Harper, L.L.; Wild, J.R. *J. Bacteriol.* **1988**, 170, 2306.

66. Dong, W.; Fournier, M.J.; Mason, T.L.; Tirrell, D.A. *Polym Prepr.* **1994**, 35(2), 419.

Scheme I. Iterative Step Growth



Scheme II. Template Polymerization

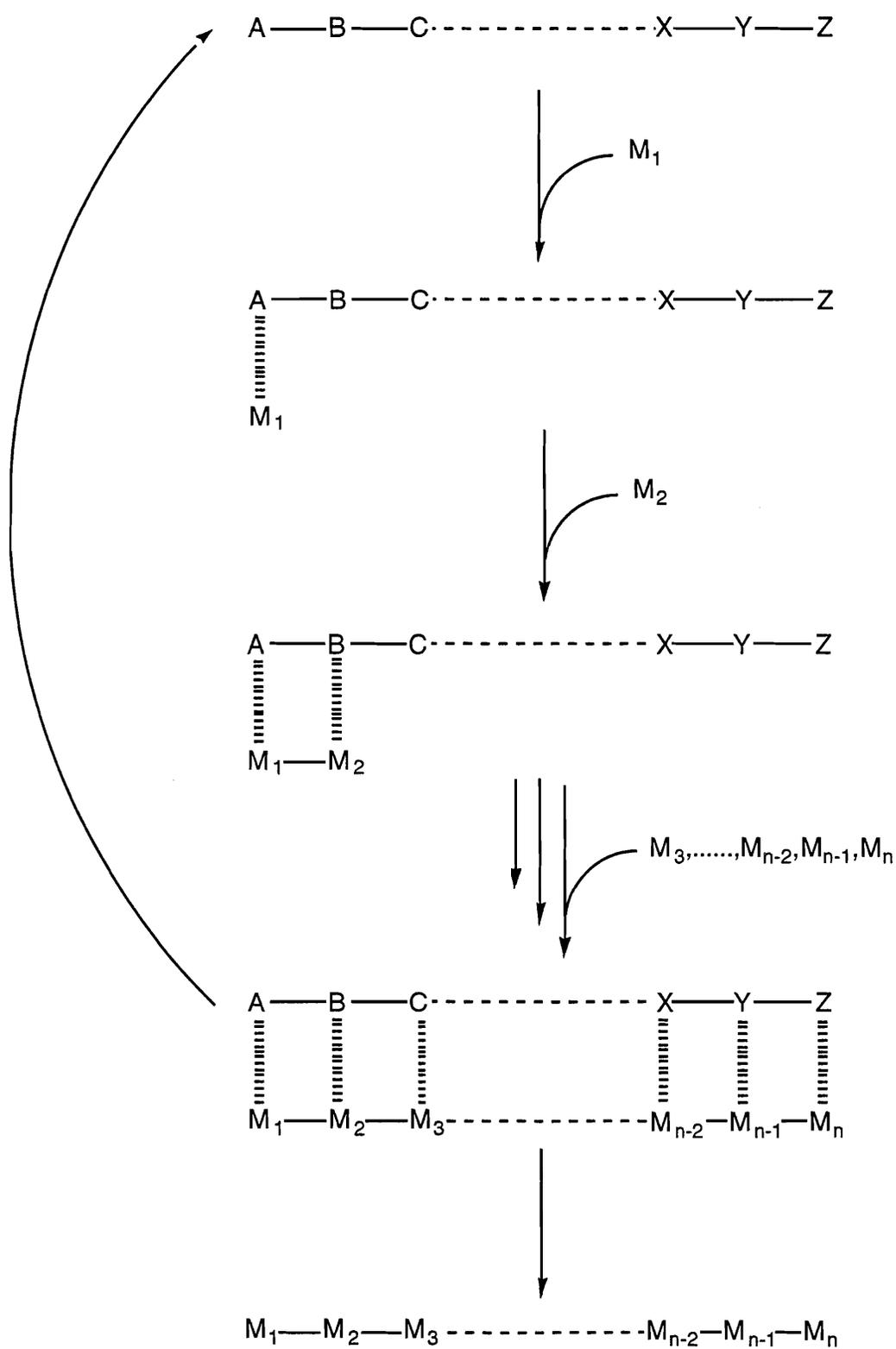


Table 1. Repeating Units of Periodic Polypeptides Currently Under Study

| <i>Repeating Unit</i> | <i>Interest</i> | <i>Status^a</i> |
|--|---|---|
| (GlyAla), GlyProGlu | Sequence-dependent chain folding | A (x = 3, 4, 5, 6, ∞) |
| (GlyAla) ₃ XY | Turn requirement for folded-chain lamellae | A (XY = GlyLeu, GlySer, GlyVal, GlyMet, SerGly, GlyAsp, GlyTyr, GlyAsn, GlyPhe, GlyThr) B (XY = 10 additional pairs) |
| (GlyAla), GlyGlu | Correlation of chemical and spatial periodicity | A (x = 3-6) |
| (GlyAla) ₃ GlyZ ^b (GlyAla), GlyZGlu ^c | Incorporation of non-natural amino acids | A |
| (GlyAla) ₃ GlyGlu(AlaGly) ₃ GlyGlu on crystal structure and morphology | Effect of stem sequence polarity | A |
| (GlyAla) ₃ GlyGlu(GlyAla) ₃ GlyVal (GlyAla) ₃ GlyGlu(GlyAla) ₃ GlyMet | Differentiation of lamellar surface functionality | B B |
| Glu ₁₇ Asp | Monodisperse poly(glutamic acid) and mesogenic poly(γ-benzyl-L-glutamate) | A |

^a A: Protein expressed and isolated;
B: Coding sequence prepared and cloned.

^b Z = L-selenomethionine, *p*-fluorophenylalanine, 5,5,5-trifluoroleucine, 3-thienylalanine.

^c Z = azetidine-2-carboxylic acid, thiazolidine-4-carboxylic acid, 3,4-dehydroproline.

Table 2. Unnatural Amino Acid Analogs Incorporated into [(GlyAla)₃-XxxGly] .

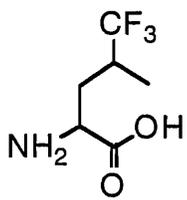
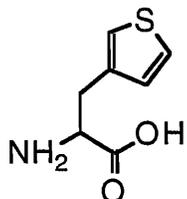
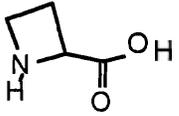
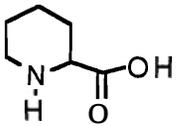
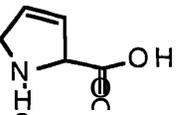
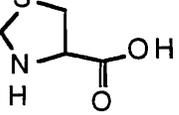
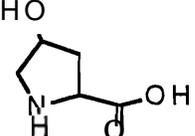
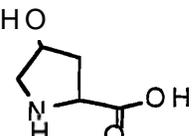
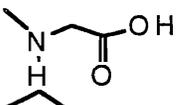
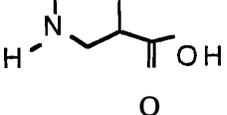
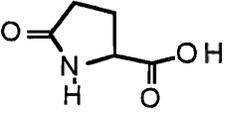
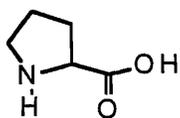
| Code | Structure | Name | Previous <i>E. coli</i> Incorporation |
|-------|---|---------------------------|---------------------------------------|
| SeMet |  | selenomethionine | yes |
| p-FP |  | p-fluorophenylalanine | yes |
| TFL |  | 5,5,5-trifluoroisoleucine | yes |
| 3-TA |  | 3-thienylalanine | no |

Table 3. Analogs of Proline Under Study

| Code | Structure | Name | Previous <i>E. coli</i> ' Incorporation |
|-------|---|--------------------------------|--|
| Azc |  | azetidine-2-carboxylic acid | yes |
| Pip |  | piperidine-2-carboxylic acid | no |
| Dhp |  | 3,4-dehydroproline | yes |
| Thp |  | thiazolidine-4-carboxylic acid | yes |
| c-Hyp |  | cis-4-hydroxyproline | yes |
| t-Hyp |  | trans-4-hydroxyproline | no |
| Sar |  | sarcosine | no |
| Nip |  | piperidine-3-carboxylic acid | no |
| Pyr |  | pyrrolidone-5-carboxylic acid | no |



Pro

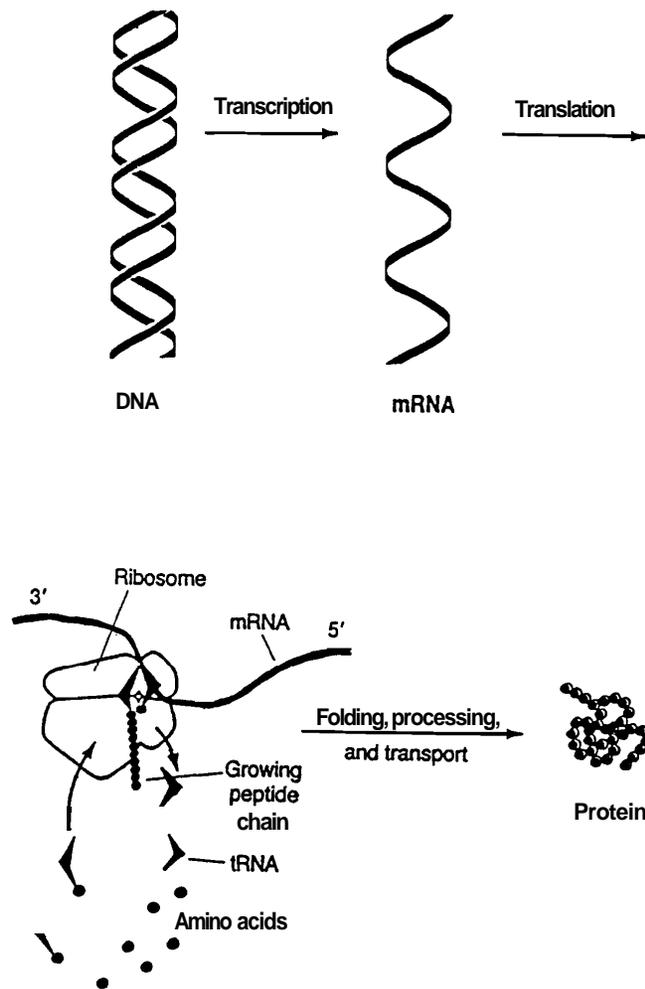


Figure 1. Schematic representation of protein biosynthesis from a DNA template. Reproduced with permission from King, J. *Chem. Eng. News* 1989, April 10, 32.

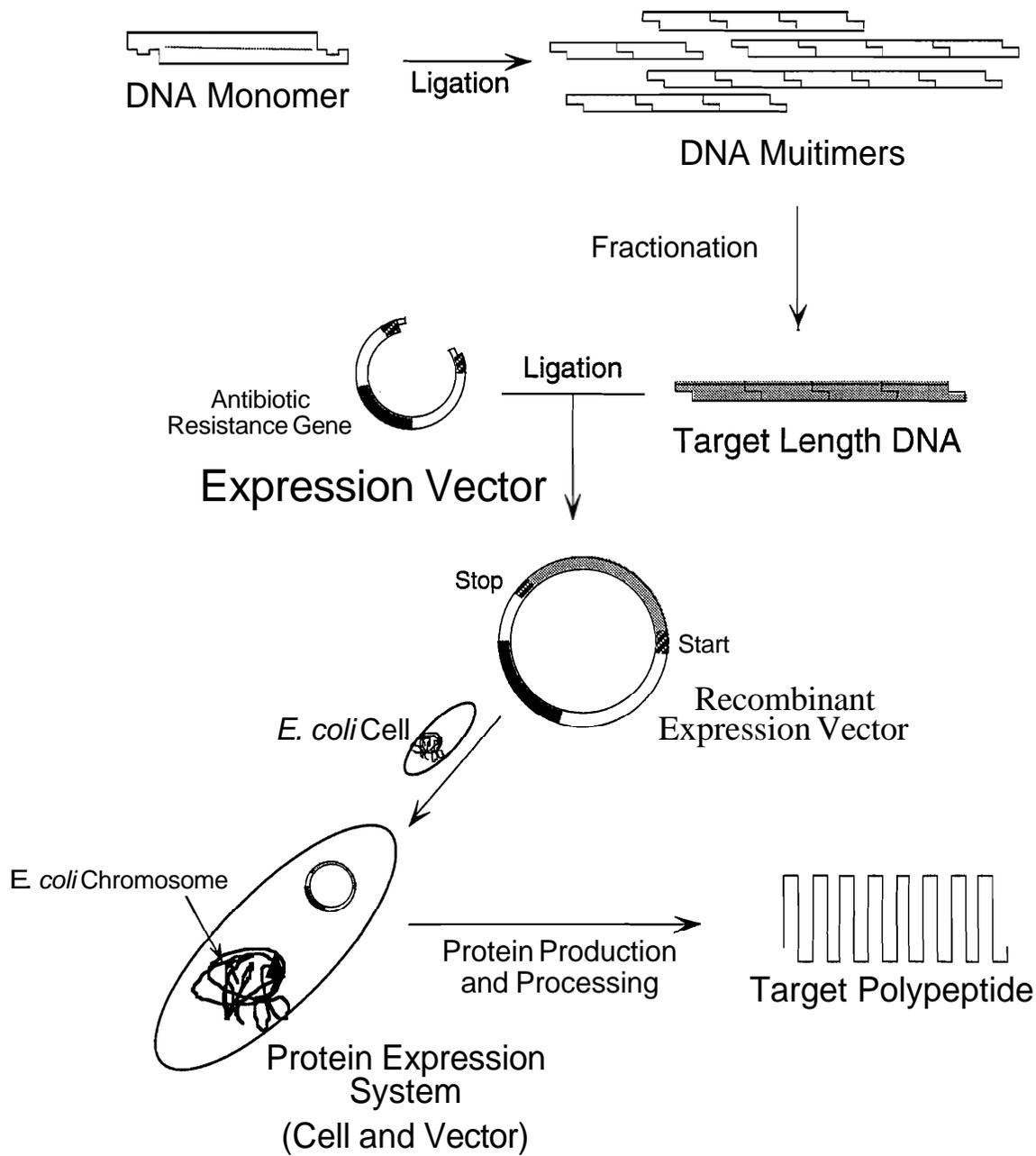


Figure 2. Synthesis, cloning, and expression of artificial genes encoding rRepetitive polypeptides.

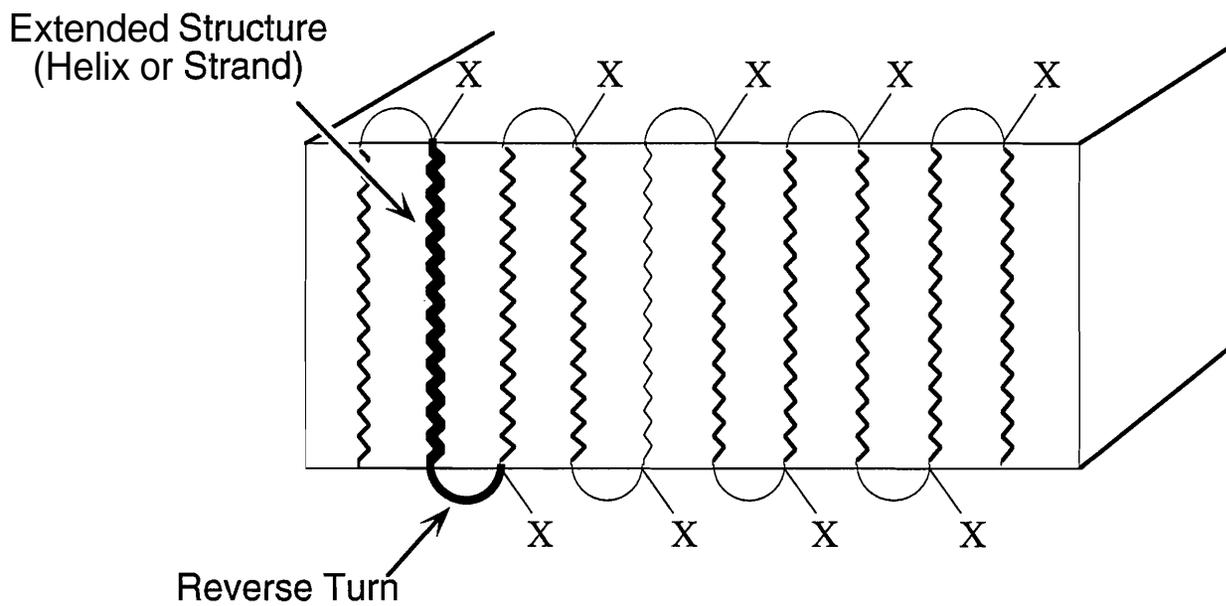


Figure 3. Schematic representation of a folded chain lamellar crystal of a periodic polypeptide built up from alternating extended and folded secondary structural elements.

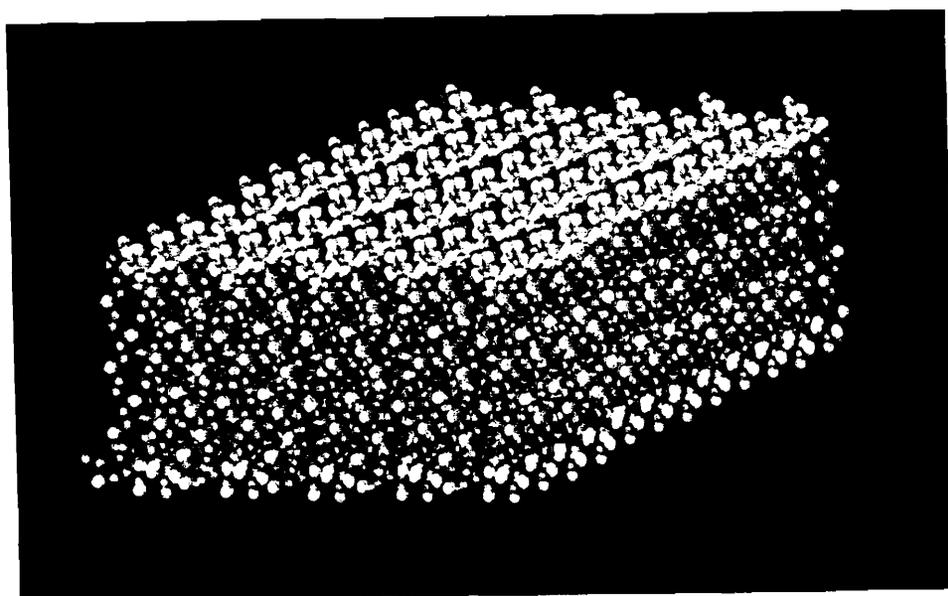


Figure 4. Molecular graphics representation of a folded-chain lamella crystal formed from a repeating polymer of sequence **3**. The blue and red "core" of the aggregate comprises the repeating alanylglycine β -sheet portion of the chain; the green surfaces present the reactive side chains of the periodic glutamic acid residues. The essential elements of the structure are supported by the results of spectroscopic and scattering studies. Reprinted with permission from reference 32.

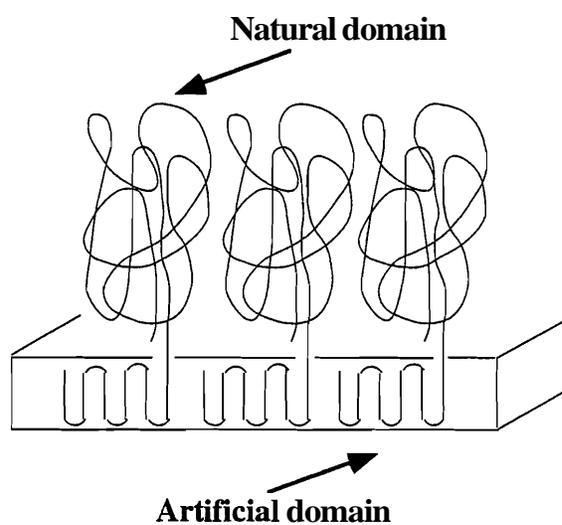


Figure 5. Schematic representation of a new class of reactive polymers, hybrid artificial proteins, which combine the catalytic or recognition properties of natural proteins with the materials properties of artificial proteins. In the authors' laboratories, hybrid artificial proteins are being evaluated as elements of enzyme-based sensors and reactors.