Design and Preparation of β-Sheet Forming Repetitive and Block-Copolymerized Polypeptides

Seiichiro Higashiya,† Natalya I. Topilina,† Silvana C. Ngo,† Dmitri Zagorevskii,‡ and John T. Welch*,†

Department of Chemistry, University at Albany, SUNY, 1400 Washington Avenue, Albany, New York 12222, and Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180

Received November 17, 2006; Revised Manuscript Received January 31, 2007

The design and rapid construction of libraries of genes coding β-sheet forming repetitive and block-copolymerized polypeptides bearing various *C*- and *N*-terminal sequences are described. The design was based on the assembly of DNA cassettes coding for the $(GA)_3GX$ amino acid sequence where the (GAGAGA) sequences would constitute the β-strand units of a larger β-sheet assembly. The edges of this β-sheet would be functionalized by the turn-inducing amino acids (GX). The polypeptides were expressed in *Escherichia coli* using conventional vectors and were purified by Ni-nitriloacetic acid (NTA) chromatography. The correlation of polymer structure with molecular weight was investigated by gel electrophoresis and mass spectrometry. The monomer sequences and post-translational chemical modifications were found to influence the mobility of the polypeptides over the full range of polypeptide molecular weights while the electrophoretic mobility of lower molecular weight polypeptides was more susceptible to *C*- and *N*-termini polypeptide modifications.

Introduction

The preparation of genetically engineered protein-based polymers, where the polypeptides can be prepared with precise control of the sequence, molecular weight, and stereochemistry, is an exciting and very promising area of polymer chemistry.[1] A unique aspect of genetically engineered polymeric materials is that the preparation of the biochemical polymer template, synthesis of the coding DNA sequence, is required only once. Subsequent biosynthetic processes obviate the need for difficult or complex synthetic manipulations. In particular, modern cloning techniques facilitate the systematic substitution of unique or regular sites in very large polymers.

Rational design of polypeptide-based polymers that fold via established rubrics to form specific secondary structures enables the programming of those mechanical or physical properties[1,2] into the sequence that may be desirable for drug delivery[1e,3] or protein purification.[4] The construction of β-sheet forming polypeptides with utility as constructs for the study of β-sheet folding,[5] aggregation, or amyloid formation[5a,6] is possible. Ultimately, the self-assembly of such molecular building blocks may also find applications in modern nanotechnology[1a−c,7] where precise dimensional control and selective functionalization are crucial.

The most common strategy for the biological synthesis of repetitive polypeptides requires the chemical synthesis of the DNA coding sequence for the construction of a double-stranded DNA cassette, subsequent enzymatic concatenation of the cassettes, and expression of the polypeptide in suitable cellular system.[8] The most difficult step in this strategy is the rapid, precise, and flexible concatenation or block-copolymerization of double-stranded DNA cassettes. Successful methods for these processes would enable the sequential engineering of genetic blocks and therefore unique block polypeptides.

Concatenation has been used to create self-ligated DNA with a variable number of monomeric tandem repeat sequences. For effective single-step processes, cyclization of the growing oligomer must be obviated by the method of McPerson et al.[4a] where special pendent DNA sequences are incorporated in the ligation reaction mixture to serve as "chain terminators" thereby promoting formation of polymers with the desired molecular weight. Often with this method, the yield of high molecular weight concatemers is relatively low. Other approaches to the synthesis of high molecular weight concatemers are based on the repetitive ligation of pairs of DNA molecules which results in a stepwise, geometric, increase in the DNA molecular weight.[9] A slight modification of this strategy coupled with sequentional ligation of the obtained concatemers into a vector can be employed for preparation of elastin-like block-copolymers.[1e] After hydrophilic and hydrophobic blocks in separate DNA cassettes were encoded, each cassette was concatenated independently, and the concatenated blocks were joined. When introduced via sequential ligation, the desired block oligomers are formed yet the necessary head-to-tail structure is maintained in the copolymers. These approaches guarantee the correct fusion order but require extensive cloning.

In this article, a strategy for the rapid creation of repetitive polypeptide libraries is reported. Constructed from small DNA cassettes comprised of approximately 24 nucleotides coding for eight amino acids in a sequence designed to form a single β-strand and a single turn, these libraries can be combined to form de novo designed antiparallel β-sheet structures. Planned studies of the fundamental properties of β-sheet structure will employ these polypeptide libraries for the systematic investigation of protein folding and misfolding, tertiary peptide interactions, and amyloidogenesis. For this reason, a cloning strategy that allowed the rapid and facile creation of the necessary libraries was required. Where McPerson et al.'s approach[4a] was

* Author to whom correspondence should be addressed. E-mail: jwelch@uamail.albany.edu.

† University at Albany.

‡ Rensselaer Polytechnic Institute.

10.1021/bm061098y CCC: $37.00
Published on Web 03/28/2007

suitable for assembly of different building blocks and for preparation of homorepetitive units, our strategy makes possible concatenation of two different cassettes which can then be oligomerized in a second-level concatenation step to form a new binary-derived DNA construct. By straightforward extension, further combination of various binary or single DNA cassettes allows easy construction of more complex yet still oligomeric DNA sequences. Specific adaptive DNA sequences that introduce the appropriate endonuclease restriction sites are uniquely utilized in the terminal concatenation steps. Other adaptive DNA sequences are employed to introduce specialized fusion constructs at both the *N*- and *C*-termini of the peptide chain. These approaches enable retention of the basic polypeptide core, yet quickly and precisely enable variations, to include such important parameters as polypeptide length (the number of repeats), turn substituents, strand architecture, specifically located charge, and hydrophobicity or hydrophilicity.

Materials and Methods

Restriction endonucleases and modifying enzymes were purchased from New England Biolabs (Beverly, MA) unless otherwise mentioned. Inoue ultra-competent and electrocompetent cells of XL1-Blue (Stratagene, La Jolla, CA) and DH5αF′ (Invitrogen, Carlsbad, CA), were prepared, transformed, and selected on LB agar plates containing ampicillin (100 μg/mL) with blue-white selection for pUC18 clones or kanamycin sulfate (50 μg/mL) for pET-28 clones according to literature methods.[10] Competent cells of BLR(DE3) and BLR(DE3)-pLysS were purchased from Novagen, Inc. (Madison, WI). BLR(DE3)-pLysSRARE was prepared from BLR(DE3) transformed by pLysSRARE isolated from Rosetta(DE3)pLysS (Novagen, Inc., Madison, WI), and the standard competent cells were prepared according to literature methods (glycerol was added at 15% to prevent the autolysis upon thawing).[10] Plasmids and DNA fragments separated by agarose gel electrophoresis were purified using QIAprep Spin Miniprep Kit and QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA), respectively.

Repetitive polypeptides are represented first by the *N*-terminal sequence and then the repeat unit and finally the *C*-terminal sequence. The number preceding the repeat sequence refers to the total number of amino acid residues in a single monomer unit, and the suffix value indicates the number of repeats. For example, 32YEHK7 and 16YF12 represent 7 repeats of 32-amino acid unit $(GA)_3GY(GA)_3GE(GA)_3GH(GA)_3GK$ and 12 repeats of 16-amino acid unit $(GA)_3GY(GA)_3GF$, respectively. To simplify, the same compound number was used for both the constructed expression vectors and the repetitive polypeptide of that vector product unless specified.

Synthetic DNA for Monomers and Adaptive Sequences. Sequences of synthetic DNA for coding units and adaptive DNA are listed in Figure 1. Each DNA coding pair **1** (8F), **2** (16YF), **3** (8Y), **4** (8E), **5** (8H), and **6** (8K) was mixed at 20 mM and the 5′ ends were phosphorylated for 30 min to 1 h at 37 °C using T4 polynucleotide kinase (New England Biolabs, Beverly, MA) and then were annealed with simultaneous deactivation of T4 polynucleotide kinase. The 5′ end of each of the oligonucleotides units **7b**, **8a**, **9b**, **10a**, **11b**, **12b**, **13a**, and **14b** was phosphorylated for 30 min to 1 h at 37 °C, and then the kinase was deactivated at 80 °C for 20 min. To each solution was added the respective complimentary oligonucleotides **7a**, **8b**, **9a**, **10b**, **11a**, **12a**, **13b**, and **14a** at a final concentration of 20 mM. The mixtures were then annealed to give **7** (A1-1), **8** (A1-2), **9** (A2-1), **10** (A2-2), **11** (A3-1-H6), **12** (NdeI-A3-1-H6), **13** (A3-2-H6), and **14** (A3-1-C2), respectively. The resulting internally phosphorylated adaptive sequences were appropriately mixed to give **15** (A1, **7** + **8**), **16** (A2, **9** + **10**), **17** (H6-H6, **11** + **13**), **17′** (NdeI-H6-H6, **12** + **13**), and **18** (C2-H6, **14** + **13**) at the final concentration of 1 mM (see phosphorylated sites represented as "p" in Figure 1).

Preparation of Recipient Vectors. pUC18 (10 μg, Bayou Biolabs, Harahan, LA) and pET-28a (1 μg, Novagen, Inc., Madison, WI) were digested by *Bam*HI and *Eco*RI, and the desired fragments were separated by agarose gel electrophoresis, purified using QIAquick Gel Extraction Kit and eluted with TE (pH 8.0) 100 μL to give **19a** (pUC18/*Bam*HI-*Eco*RI) and 40 μL to give **19b** (pET-28a/*Bam*HI-*Eco*RI). One of pET-28a-c vectors (1 μg, Novagen, Inc., Madison, WI) was digested with *Nco*I and *Bam*HI and then was purified by the same way to give **19c** (pET-28/*Nco*I-*Bam*HI). pET-9a (1 μg, Novagen, Inc., Madison, WI) was digested with *Nde*I and *Bam*HI to give **19d** (pET-9a/*Nde*I-*Bam*HI).

First-Generation Repetitive Polypeptides: A2-F_{13} and A2-$(YE)_{12}$. *(a) Concatenation of 1 (8F) and 2 (16F).* The oligonucleotide solutions 1 (8F) and 2 (16F) (1∼2 μL each) were mixed with the adaptive DNA solution 15 in a ratio of 10∼100:1 (0.2∼2 μL), PEG8000 (final concentration at 4%), and H_2O (up to 9 μL) and were annealed at 45 °C for 5 min. Then, T4 DNA ligase (1 μL) and the supplied ligation buffer (1 μL) were added, and then the system was allowed to incubate at 4 °C overnight or longer. The resulting population of oligomers bearing the adaptive sequence at both termini was subjected to electrophoresis on agarose gel. Oligomerized mixtures with the desired molecular weight were excised from the gel, were purified, and were precipitated with EtOH in a 0.6-mL microcentrifuge tube.

To the each tube was added vector solution **19a** (pUC18/*Bam*HI-*Eco*RI) (2 μL), PEG8000 (final concentration at 4%), T4 DNA kinase (0.25 μL), the supplied kination buffer (0.5 μL), and H_2O (up to 5 μL), and then the mixture was incubated at 37 °C for 30 min. T4 DNA ligase (0.25 μL) was added at room temperature, and the mixture was incubated at 4 °C overnight or longer. The ligation mixture was used to transform XL1-Blue or DH5αF' cells. Plasmids of selected colonies were isolated and after *Bam*HI digestion were examined for the appropriate fragment sizes. Plasmids harboring 13 repeats of 8F (**20**, 8F13) and 4 repeats of 16YF (**21**, 16YF4) were selected and used for further oligomerization.

(b) Construction of Expression Vectors for 8F and 16YF Repetitive Polypeptides. **20** (8F13) and **21** (16YF4) were digested by *Bsa*I, and fragments of the oligomerized units were isolated, purified, and recursively oligomerized with adaptive sequence **16** as described. The desired oligomer was isolated, purified, and ligated with recipient vector **19b** (pET-28a/*Bam*HI-*Eco*RI), and the resultant ligation mixtures were used to transform *Escherichia coli*. Plasmids of selected colonies were isolated and, after *Nco*I-*Bam*HI digestion, the fragment sizes were determined. Plasmids harboring 13 repeats of 8F (**22**, pET-28a/A2-8F13) and 12 repeats of 16YF (**23**, pET-28a/A2-16YF12) were identified. A schematic representation of cloning is depicted in Figure 2.

Second-Generation Polypeptide Libraries. Construction of Expression Vectors of Repetitive Polypeptides Libraries: 27 (H6-(32YEHK)n-H6), 28 (C2-(32YEHK)n-H6), 29 (H6-(32HKYE)n-H6), 30 (C2-(32HKYE)n-H6), 32 (A2-32YEHKn), 36 (H6-16YEn-H6), 37b (H6-(YE8HK8)n-H6), 38 (H6-(HK8YE8)n-H6), 39 (H6-32KYEYn-H6), 40 (C2-32KYEYn-H6), and 41 (C2-32YKYEn-H6). *(a) Assembly of **3** (8Y), **4** (8E), **5** (8H), and **6** (8K) for Construction of Repeating Units **24a** (16YE), **25a** (16HK), 16YK.* Initially, **3** (8Y) and **4** (8E) were copolymerized with adaptive sequence **15** and were cloned to **19a** using the method described for construction of expression vectors 8F and 16YF. Plasmids harboring dimerized **24a** (16YE), **24b** (16YY), **24c** (16EY), and **24d** (16EE) were identified. Similarly, plasmids harboring **25a** (16HK) were prepared from **5** (8H) and **6** (8K) along with **25b** (16HH), **25c** (16KH), and **25d** (16KK).

*(b) Assembly of **24a** (16YE), **25a** (16HK) for Construction of Repeating Units **26a** (32YEHK), **26b** (32HKYE).* An equal number of plasmids harboring **24a** (16YE) and **25a** (16HK) was digested by *Bsa*I, and the resulting mixture of 16YE and 16HK units was purified and ligated with adaptive sequence **15**. The desired fragments of adapted dimers were isolated, purified, and cloned to **19a** (pUC18/*Bam*HI-*Eco*RI). Plasmids harboring **26a** (32YEHK) were identified along with **26b** (16YE2 or 32YEYE), **26c** (32HKYE), and **26d** (16HK2 or 32HKHK).

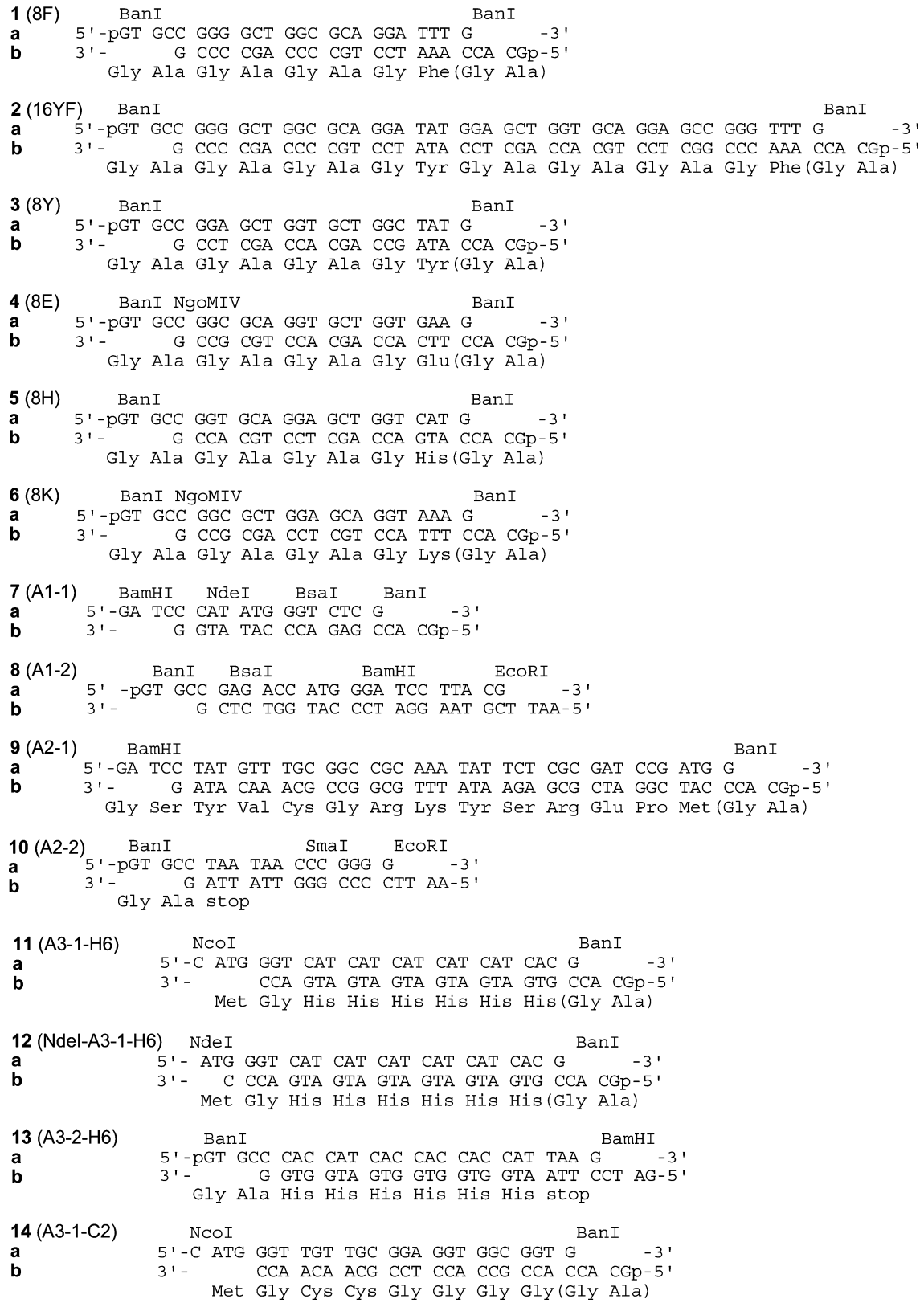

Figure 1. DNA design and sequences.

*(c) Concatenation of **26a** (32YEHK), **26b** (32HKYE) and Construction of the Expression Vectors with Genes Coding Low Molecular Weight Polypeptides.* **26a** (32YEHK) and **26c** (32HKYE) were digested by *Bsa*I, and the resulting 32YEHK and 32HKYE fragments were purified, oligomerized with adapters **17** and **18**, respectively, and were cloned to **19c** (pET-28/*Nco*I-*Bam*HI) to give a library of expression vectors harboring **27a**−**h** (H6-32YEHKn-H6, $n = 1-8$), **28a**−**h** (C2-32YEHKn-H6, $n = 1-8$), **29a**−**d** (H6-32HKYEn-H6, $n = 1-4$), or **30a**−**h** (C2-32HKYEn-H6, $n = 1-6$).

*(d) Concatenation of **26a** (32YEHK) and Construction of Expression Vectors with Genes Coding High Molecular Weight Polypeptides.* 32YEHK fragments were further oligomerized with adaptive sequence **15**, were cloned to **19a** (pUC18/ *Bam*HI-*Eco*RI), and a plasmid harboring seven repeats of 32YEHK (**31**, 32YEHK7) was identified

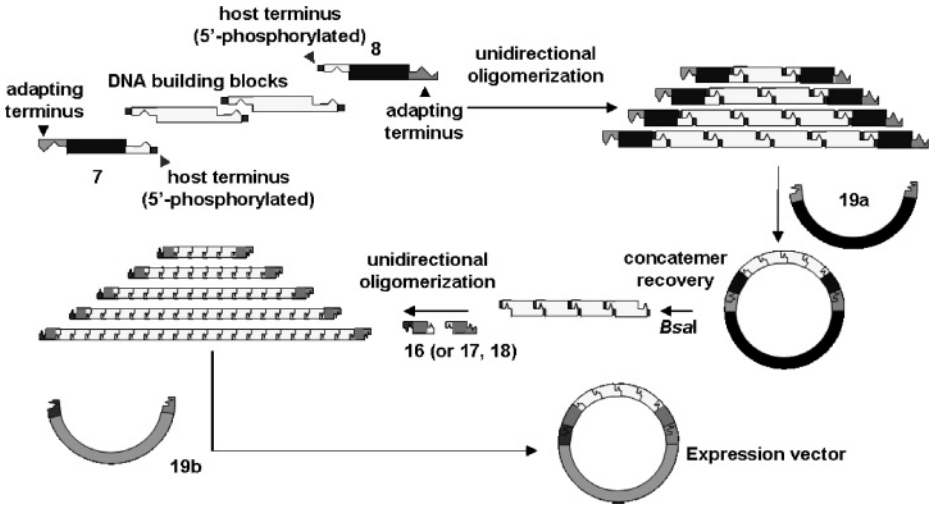

Figure 2. Construction of repetitive DNA sequences and insertion of these genes into the expression vector.

along with other various repeats units. **31** (32YEHK7) was digested by *Bsa*I, and the resulting 32YEHK7 fragment was purified, oligomerized with **16**, **17**, and **18**, and then was cloned to **19c** (pET-28/*Nco*I-*Bam*HI) to give a library of expression vectors harboring (**32a**−**c**, pET-28a/A2-32YEHKn, n = 14, 21, 28), (**27g**−**i**, pET-28/H6-32YEHKn-H6, n = 7, 14, 21), and (**28g**, **h**−**k**, pET-28/C2-32YEHKn-H6, n = 7, 14, 21, 28, 35). The representative synthetic scheme is depicted in Figure 2 and Figure 3.

(e) Construction of (YE)n, (HK)n, and Their Block-Copolymerization and Concatenation. 16YEn, 16HKn, (16YE8-16HK8)n, and (16HK8-16YE8)n repetitive coding sequences were similarly prepared from **26b** (32YEYE) and **26d** (32HKHK). First, a library of **33a** (16YE4), **33b** (16HK4), **33c** (16YE2-16HK2), and **33d** (16HK2-16YE2), a collection of **34a** (16YE8), **34b** (16HK8), **34c** (16YE4-16HK4), and **34d** (16HK4-16YE4) from **33a** (16YE4) and **33b** (16HK4), and then a library of **35a** (16YE16), **35b** (16HK16), **35c** (16YE8-16HK8), and **35d** (16HK8-16YE8) from **34a** (16YE8) and **34b** (16HK8) were constructed. Finally, **34a** (16YE8), **34b** (16HK8), and **35d** (16HK8-16YE8) were oligomerized with **17** and then were cloned into **19c** (pET-28/*Nco*I-*Bam*HI). **35c** (16YE8-16HK8) was oligomerized with **17′** and then was cloned to **19d** (pET-9a/*Nde*I-*Bam*HI). The resulting libraries of plasmids harboring repetitive coding sequences with H6-H6 adaptive sequence (**36a**−**d**, pET-28/H6-16YEn-H6, n = 8, 16, 24, 32; **37a**−**d**, pET-28/H6-16HKEn-H6, n = 8, 16, 24; **38a**−**b**, pET-9a/H6-(YE8HK8)n-H6, n = 1, 2; **39a**−**g** pET-28/H6-(HK8YE8)n-H6, n = 1−7) were obtained. KYEY (**40a**−**d**, pET-28/H6-32KYEYn-H6, n = 8, 16, 24, 32; **41a**−**g**, pET-28/C2-32KYEYn-H6, n = 8, 16, 24) and YKYE series (**42a**−**d**, pET-28/H6-32YKYEn-H6, n = 8, 16; **43a**−**c**, pET-28/C2-32YKYEn-H6, n = 8, 16, 24) were prepared from 16KY and 16YK obtained from **3**(8Y) and **6**(8K) and **24a** (16YE) and **24c** (16EY), respectively.

Expression and Purification of Repetitive Polypeptides. Expression vectors harboring coding sequences were used to transform expression hosts BLR(DE3)pLysS or BLR(DE3)pLysSRARE (42 °C for 90 s and then 4 volumes of SOC medium, 37 °C for 1 h).[10] The transformants were selected on LB agar plates containing chloramphenicol (34 μg/mL), 1% glucose, and 50 μg/mL of kanamycin sulfate without tetracycline. Selected colonies were inoculated into 2xYT broth (2−5 mL for preliminary expression on a 5−50 mL scale and in 200 mL for 4-L scale expression) containing kanamycin (50 μg/mL), chloramphenicol (34 μg/mL), and 1% glucose. The overnight precultures were used to inoculate the desired quantities of expression culture (2xYT broth containing kanamycin (10 μg/mL) and chloramphenicol (34 μg/mL). Expression of polypeptides was initiated between 1.0 and 2.0 optical density (OD) by the addition of isopropyl α-D-1-thiogalactopyranoside (IPTG) (final concentration 1 mM) and was induced for 3−4 h with aeration.[10] The expression and purification of the polypeptides were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The polyhistidine-tagged polypeptides were detected by western blot using a SuperSignal West HisProbe kit (Pierce, Rockford, IL).

Polypeptides expressed from plasmids harboring 8F and 16YF with A2 adaptive sequences were purified and cleaved according to the literature protocol[11] with an additional three washes with 1% Triton-X100. Small- and large-scale purification of other polyhistidine-tagged repetitive polypeptides, under denaturing conditions in 8 M urea, was possible using Ni-NTA Spin Column and Ni-NTA Superflow (Qiagen Inc., Valencia, CA), respectively. Solutions of 10, 20, 40, 100, 300, and 500 mM imidazole in 8 M urea and 1× PBS (phosphate-buffered saline, 0.1 M phosphate buffer (pH 7.4), 0.5 M NaCl at final concentration) were prepared for separation of the polyhistidine-tagged repetitive polypeptides. The required quantitites of 2 M imidazole (pH 7.4), urea and 8× stock solution of PBS were combined and then were filtered through a 0.45-μm filter.[12]

For large-scale purification, cells pelleted by centrifugation at 3500 G for 30 min at 4 °C were resuspended in H_2O (15 mL per 1-L culture) and were frozen. The frozen cells were lysed by freeze−thaw sonication method for 20 min in the presence of benzonase (2 μL per 1-L culture, Novagen, Inc., Madison, WI) and PMSF (phenylmethansulfonyl fluoride, 2 mM). β-Mercaptoethanol (5 μL/1-L culture, 0.2 μL/mL lysate) was added in the cases of C2-H6 polypeptides. The lysate was allowed to incubate for 30 min at room temperature (rt). For purification of carbamylated polypeptides, urea (11 g per 1-L culture, 8 M at final concentration) was added to the lysate and the solution was heated in a boiling water bath for 4 h with occasional mixing. For purification of unmodified peptides, the lysate was added to guanidine hydrochloride (16 g of guanidine hydrochloride per 15 mL H_2O of resuspended frozen stock, 6 M at final concentration) and then was incubated on boiling water bath for 1 h with occasional mixing. In both cases, the resultant solution was centrifuged at 15 000 G for 30 min at 20 °C, and then the supernatant was diluted twice with 10 mM imidazole in 8 M urea in PBS or H_2O with 10 mM imidazole and was recentrifuged at 25 000 G for 1 h at 20 °C. The supernatant was applied to Ni-NTA column, and the H6-H6 polypeptides adsorbed on the column. The column was eluted with 40 mM imidazole in 8 M urea in PBS and then with 300 mM imidazole in 8 M urea PBS while A2 and C2-H6 modified polypeptides were washed with 10−20 mM imidazole 8 M urea in PBS and were eluted with 100 mM imidazole 8 M urea in PBS. Carbamylation could also be performed after purification by incubation of the column eluent containing 8 M urea on boiling water bath for 1−2 h. In each case, the eluent was dialyzed against H_2O for at least 2 days (dialysis membrane MW cutoff = 3500 Da). The peptide products were usually isolated in concentrations greater than 10 mg/L.

Characterization of Repetitive Polypeptides. MALDI mass spectra were recorded on a TOF Spec 2E time-of-flight mass spectrometer (Micromass, Wythenshawe, United Kingdom) in linear positive ion mode. Ionization was achieved using pulsed N_2 laser at a wavelength

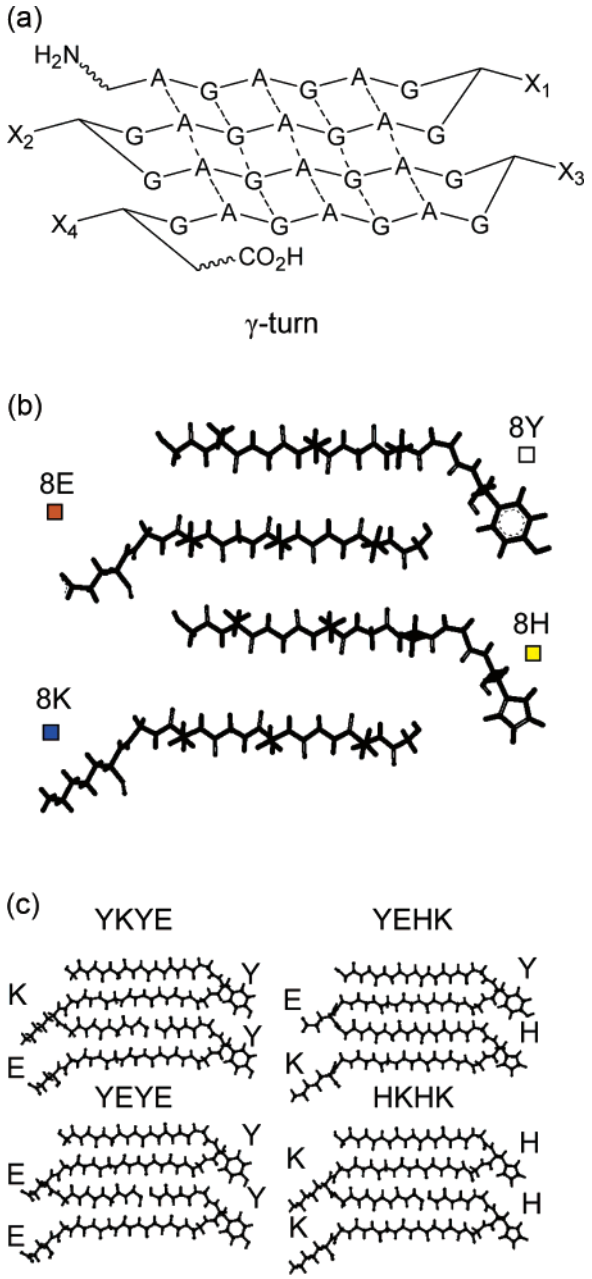

Figure 3. Design of repetitive unit architecture. (a) General design of polypeptide repetitive unit. (b) Peptide fragments which are coded by smallest DNA building blocks (strand + turn). (c) Representative repetitive units YKYE, YEHK, YEYE, HKHK.

of 337 nm. The laser power before passing the lenses was ~180 μJ. Typical instrument settings were pulse duration, 39 μs; pulse voltage, 3500 V; laser coarse, 50%; and laser fine, 80%. A voltage of 5−7 kV was applied to the high mass detector to improve the efficiency of the detection of the polypeptides. Data acquisition and data processing were performed using MassLynx 3.5 software. Typically, 200 shots were averaged in one scan. Sinapinic acid (Fluka, Switzerland) was used as a matrix (10 μg/mL). The mass scale was calibrated using bovine serum albumin (BSA) or insulin cluster ions as standards. The calibrants were spotted next to the analyte to allow acquisition during the same run. The accuracy of the mass measurements was better than 0.5% when a good signal-to-noise ratio (>10:1) was observed.

Most MALDI-TOF measurements utilized 10-fold diluted eluant since the peptides tend to aggregate on dialysis or during lyophilization. Samples were also prepared by dissolving the lyophilized polypeptides in 8 M urea, 6 M guanidine chloride and thiocyanate, or 70% formic acid, followed by 10-fold dilution.[13] The optimization of the matrix-to-analyte ratio was achieved by a series of dilutions of the sample−matrix solutions directly on a matrix assisted laser desorption ionization (MALDI) plate.[13a] For this purpose, 0.5 μL of the initial mixture was combined with 0.5 μL of the matrix. This solution (0.5 μL) was transferred to the next spot, and so on. Typically, three to five dilutions provided the optimum sample-to-matrix ratio.

Electrospray ionization mass spectra were obtained on an Agilent 1100 series LC/MSD-SL ion trap system (Germany). Samples were introduced into the ion source using the autosampler with a solvent (50:50 water-acetonitrile with 0.2% formic acid) flow rate of 200 μL/min. The data were acquired in a mass range from *m*/*z* 750 to *m*/*z* 1500. The ion trap parameters were optimized for detection of ions with *m*/*z* 1150. Instrument control, data acquisition, and data processing were performed using ChemStation 10.01 and IonTrap 5.2 software. Before injection, the samples were dialyzed or purified using ZipTips (Millipore, Bedford, MA).

Molecular weight of repetitive polypeptides was also estimated by SDS-PAGE (12%). The mobilities of the molecular weight marker proteins were plotted versus the logarithm of the corresponding molecular weight. Molecular weight fit by logarithmic approximation ($R^2 \geq 0.99$) was employed for molecular weight estimation.

Results and Discussion

1. Polypeptide Design and Construction. *1.1. Libraries of Antiparallel β-Sheet Forming Repetitive Polypeptides.* Repetitive polypeptides have previously been prepared in the course of investigations of elastin-based polymers. Several groups have also prepared high molecular weight polypeptides which can adopt a β-sheet conformation.[2e,11,14,15] Two principal approaches were developed for creation of β-sheet polypeptide libraries. In the first approach, a polypeptide homopolymer with an appropriate repeating unit was constructed by a single concatenation of the repetitive unit. An alternate approach based on the formation of polypeptide libraries where the peptide would assume a β-sheet structure relied upon a binary patterning of β-strand elements, that is, alternation of hydrophobic or hydrophilic amino acids. This assembly did, however, consistently employ a specified amino acid sequence for turn introduction at the desired site.[14b−h] In our work, each β-sheet forming polypeptide block is derived from the assembly of DNA building blocks that encoded the elementary constructs for a single strand and turn (Figure 3a and 3b). This strategy affords precise control of the polypeptide sequence while permitting flexible variation of the polymer architecture. The constituents of any particular turn or the position of a strand within the repetitive polypeptide can be easily designated.

1.2. Polypeptide Repeats Unit Design. The repetitive polypeptide design was based on early work by Krejechi and coworkers[11,15] where GA repeats were utilized for the β-strand forming motif with pendent amino acids X at the turns (Figure 3a). It has been reported that $(GA)_3GX$ repetitive polypeptides form antiparallel β-sheets by regular adjacent-reentry chain-folding through γ-turns that result in 0.5 nm × 0.7 nm rectangular amphiphilic structures as shown in Figure 3a.

First-Generation Polypeptide Libraries. Our initial experiments with the adaptive DNA methods reproduced literature precedent[15d,e] in construction and expression of both **22** (A2-8F13, Figure 3a, $X^1=X^2=X^3=X^4=F$) and the slightly and asymmetrically modified **23** (A2-16YF12, Figure 3a, $X^1=X^3=Y$, $X^2=X^4=F$). It was determined that the hydrophobicity of these peptides imparted a profound tendency to aggregate.

Second-Generation Polypeptide Libraries. To improve the solubility of the polypeptides, various functional groups were incorporated at the turn sites with a resultant introduction of amphiphilicity. The amphiphilic polypeptide design included

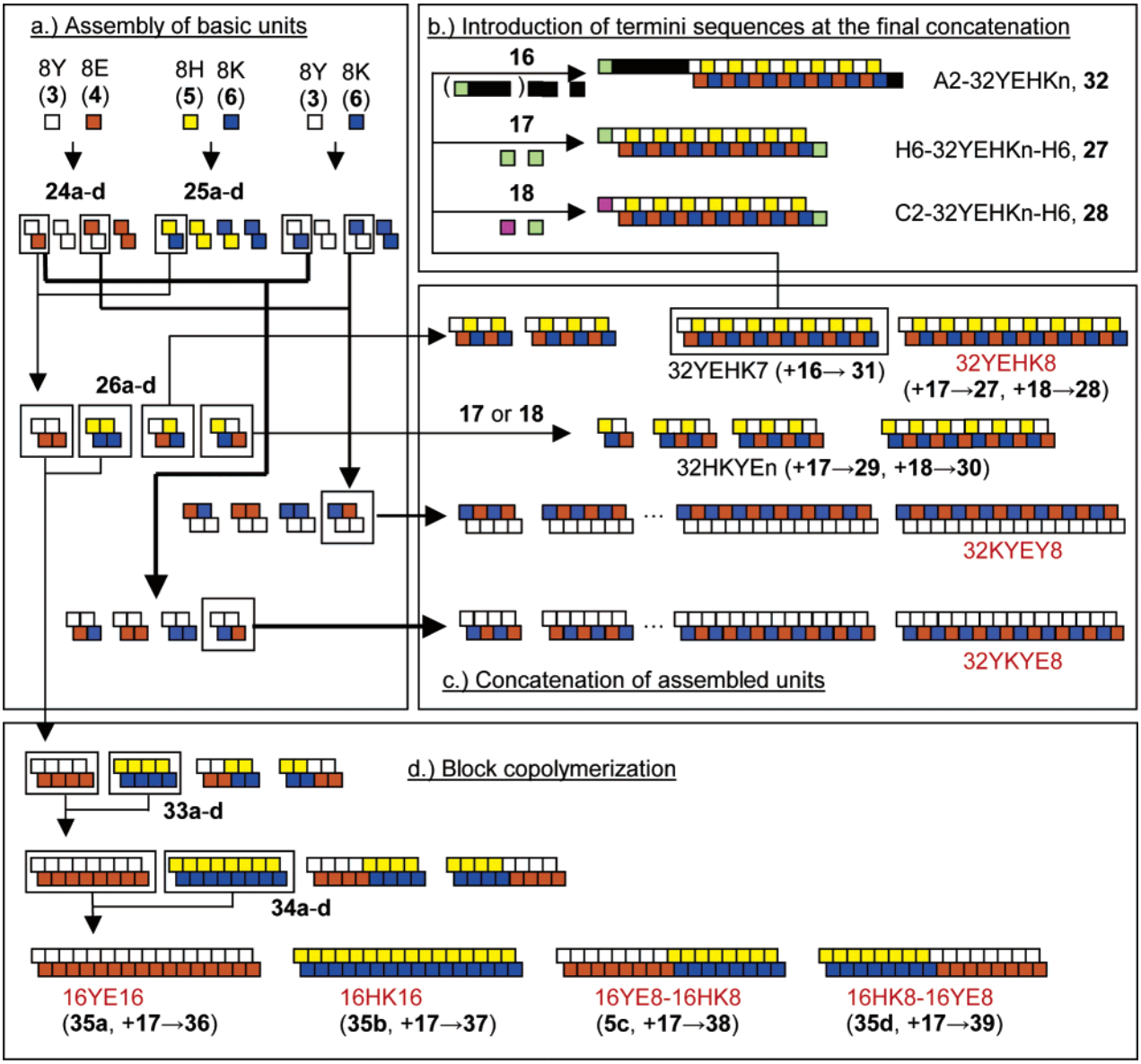

Figure 4. Schematic representation of the construction of the polypeptide libraries. Alternate boxes depict peptide β-strand subunits to illustrate relative relationship of β-turn residues along a β-sheet construct.

the introduction of hydrophilic and hydrophobic edges to the β-sheet to enhance water solubility. Starting with the four building blocks (8Y, 8E, 8H, and 8K where each coding block consists of a β-strand forming AGAGAG sequence and a turn sequence GX, Figure 3b), the heterogeneous unit YEHK was assembled. Negatively charged glutamate and protonated lysine, introduced to generate a hydrophilic edge, could form stabilizing salt bridges. The aromatic tyrosine and histidine residues were postulated to form an intramolecular π-stacked hydrophobic array. This system of β-strand subunits enabled examination of the impact of the systematic displacement of the turn residues, for example, creation of YKYE where histidinyl turn residues of the YEHK repeat are replaced with tyrosines. Deletion of the lysine residues from YKYE forms the simple distrand repeat 16YE which results in an unbalanced negative charge in the resultant polypeptides. By analogy, the HK peptide presents histidine residues along one edge and array of protonated lysines on the other. (Figure 3c)

1.3. Library Construction. Overall steps of library construction are described in Figure 4.

Building Block Assembly and Oligomerization. Oligonucleotide units were assembled on the basis of unidirectional, head-to-tail oligomerization[8f,g,h,j,m] with adaptive DNA sequences for incorporation at the appropriate cloning sites.[4a,8j] Each of the minimal DNA coding units was joined at *Ban*I sites allowing unidirectional concatenation and reduction of concatenated units to minimal coding units. Inclusion of the adaptive DNA sequences allows cloning at virtually any restriction site with common, conventional, commercially available cloning vectors. Two different types of adaptive DNA sequences were employed.

The Adaptive DNA for Cloning. A1 adaptive DNA (**15**) contains the recognition sites for the type IIs restriction endonuclease, *Bsa*I, where cloning at both the 3′ and 5′ termini leads to facile recovery of assembled DNA units, facilitating further recursive, seamless oligomerization and block copolymerization.[4a,8a,8i,8k,8l,14a] Self-ligation of the adaptive sequences during the concatenation was suppressed by selective phosphorylation of 5′-hydroxyl groups of synthetic adaptive DNA strands that constitute the termini of the oligomerized repetitive DNA units (**7b**, **8a**, **9b**, **10a**, **12b**, **13a**, and **14b**). Oligomerized DNA containing the adaptive sequences was purified by agarose gel electrophoresis. The bands containing the desired adapted

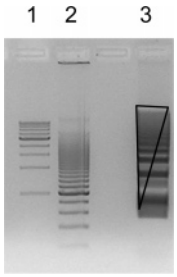

Figure 5. Oligomerization with adaptive sequences, agarose gel electrophoresis, and enrichment of longer adapted oligomers. Lane 1: 1 kb molecular marker ladder, lane 2: high 200 bp molecular marker ladder, lane 3: C2-32YEHKn-H6 oligomers. Enclosed triangular area was purified and used for cloning.

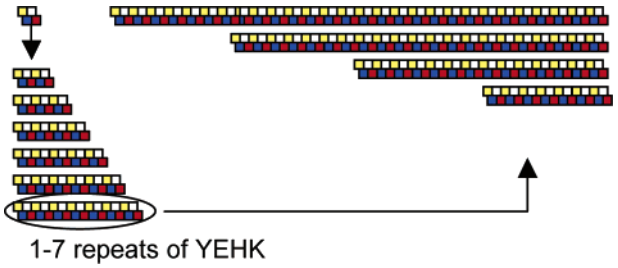

Figure 6. Schematic representation of the extent of oligomerization at different steps of concatenation.

oligomers were enriched as shown in Figure 5 and then were cloned into recipient vectors.

Adaptive DNA for Expression. Adaptive DNA for expression was utilized in the last concatenation step to terminate the oligomerization and to enable insertion of the construct into the desired expression vector. Simultaneously, the appropriate fusion amino acid sequences at both the head and tail of the repetitive and block-copolymerized polypeptides were introduced thereby enabling rapid shotgun library construction with various terminal amino acids (Figure 4b). Adaptive DNA sequence **16** was designed to facilitate the preparation of previously published polypeptides[15] by cloning into the pET-3a or pET-9a vectors. Resulting polypeptides have the repetitive sequence fused with a 4.9 kDa domain that could be easily removed enabling recovery of the purified repetitive polypeptide sequence. Adaptive sequence **17** was designed to introduce a hexahistidinyl track not only to facilitate polypeptide purification but also to terminate the β-sheet construct with a specific unique β-strand. Sequence **18** allowed differentiation of the *C*- and *N*-terminus with specific sequences, for example, a hexahistidinyl track at the *C*-terminus and a Cys-Cys repeat at *N*-terminus.

Extent of Oligomerization. In contrast with previously described methods,[9,1e] the sequential oligomerization of the DNA product recovered from the first concatenation step was affected in the presence of the appropriate adaptive DNA sequence prior to insertion into an expression vector (Figures 4c and Figure 6). This approach permits a second level of concatenation. In a single step, a library of repetitive polypeptides can be formed bearing the appropriate *C*- and *N*-terminal fusion sequences where there is a gradual increase in molecular weight, for example, 1X, 2X, 3X, 4X, and 5X (where X is the selected concatemer from previous oligomerization). This approach affords a diversity of oligomers not accessible by arithmetic synthesis on the basis of ligation of two blocks, for example, 1X, 2X, 4X, 8X, and 16X series.[9]

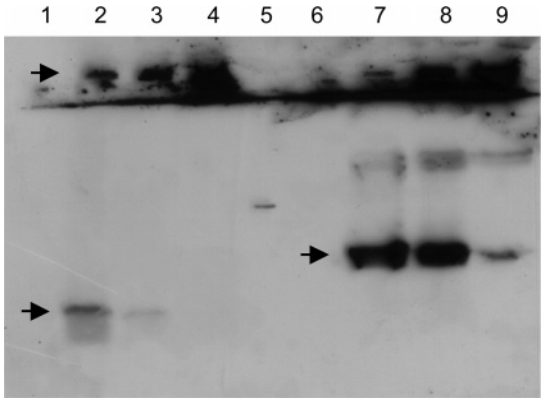

Figure 7. Western-blot analysis of expression of A2-8F13 (**22**) and A2-16YF4 (**23**). Lanes 1−4: **22**, lanes 6−9: **23**, lanes 1, 6: before induction, lanes 2, 7: 30 min, lanes 3, 8: 1 h, lanes 4, 9: 2 h after induction, lane 5: molecular weight marker (29 kD protein showing the signal).

Block-Copolymerization. Block-copolymers prepared via this stepwise concatenation strategy have distinct domains comprised of the simple motifs (Figure 4d).

2. Expression, Purification, and Characterization. RecA deficient hosts were used for both cloning and expression since DNA repetitive sequences frequently result in rapid recombination. Rare codon supplemented expression hosts were utilized to maximize the yields of target peptides. In most cases, since the product repetitive polypeptides are harmful to the bacterial hosts, 1% of glucose was added as a supplement to the agarose plates and preculture media to suppress expression leakage. Expression was induced at relatively high OD (>1) since in many cases, hosts stop growing 1 h after induction by 1 mM IPTG.

2.1. First Generation of Polypeptides: Aggregation during the Expression. Earlier, the tendency of the product peptide **22** (A2-8F13) and **23** (A2-16YF12) to aggregate was discussed. Figure 7 illustrates the time course of polypeptide expression. Within 2 h, the majority of the expression products remained in the stacking gel wells (at the top of the figure) and was transferred directly to the membrane. Once aggregation occurred, with **22** (A2-8F13) and **23** (A2-16YF12), the aggregates could not be dissociated using denaturing reagents suitable for use with Ni-NTA column chromatography such as urea or guanidine hydrochloride.

2.2. Second-Generation Repetitive Polypeptides: Introduction of Amphiphilic Character. Repetitive and block-copolymerized polypeptides **27**, **28**, **32** (YEHKn), **29**, **30** (HKYEn), **36** (16YEn), **37** (16HKn), **38** ((YE8HK8)n), **39** ((HK8YE8)n), **40**, **41** (KYEYn), and **42**, **43** (YKYEn) were constructed with various *N*- and *C*-terminal sequences by combination of 8Y, 8E, 8H, and 8K coding DNA units. In contrast to **22** (A2-8F13) and **23** (A2-16YF12) polypeptides, the second-generation polypeptides were apparently monomeric in 8 M urea or 3−6 M guanidine hydrochloride during the purification and remained soluble after dialysis against water.

3. Post-Translational Chemical Modification. The lysine residues were carbamylated in solutions containing 8 M urea (in most cases, eluent from Ni-NTA column) on boiling water bath for 1−2 h.[16] Figure 8 illustrates the discrete carbamylation of short repetitive polypeptides **27d** (H6-YEHK4-H6) and **28d** (H6-HKYE4-H6). Although **27d** and **28d** have identical molecular weights and each of the native polypeptides demonstrated nearly the same gel mobility (Figure 8, lanes 1 and 9, Figure 9, empty (white) series), the carbamylated peptides appear at

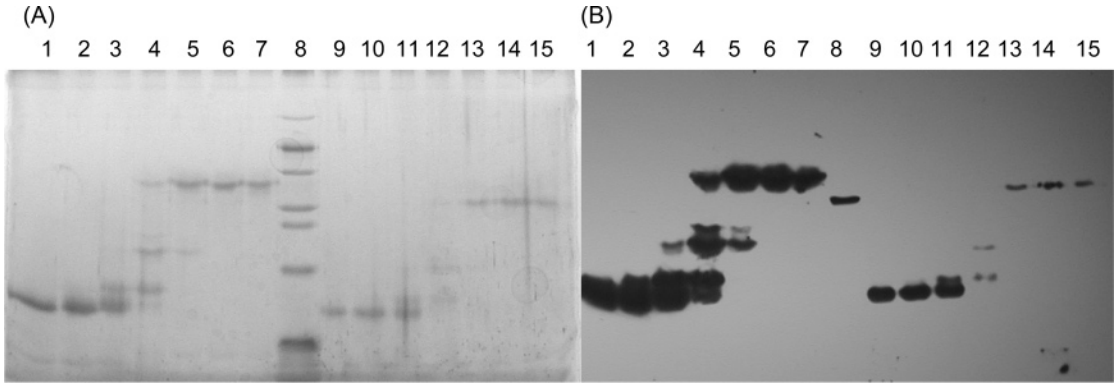

Figure 8. Time course of carbamylation of H6-YEHK4-H6 **27d** and H6-HKYE4-H6 **29d**. (A) 15% SDS-PAGE, (B) western blot, lanes 1−7, 16: **27d**, lane 8: molecular weight marker (from the top: 66, 45, 36, 29, 24, 20, 14.2 kD; 29 kD marker also shows the western blot signal), lanes 9−15: **29d**, lanes 1 and 9: as eluted (300 mM imidazole, 8 M urea PBS), lanes 2 and 10: boiling in 8 M urea for 2 min, lanes 3 and 11: 5 min, lanes 4 and 12: 10 min, lanes 5 and 13: 20 min, lanes 6 and 14: 60 min, lanes 7 and 15: 120 min.

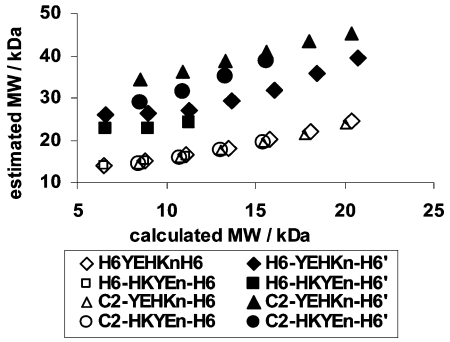

Figure 9. A comparison of estimated molecular weight from the mobility of short polypeptides on 12% SDS-PAGE with expected molecular weight. Empty: intact, filled: carbamylated. H6-YEHKn-H6 **27** and **27′**, H6-HKYEn-H6 **29** and **29′**, *ρ*: C2-YEHKn-H6 **28** and **28′**, C2-HKEYn-H6 **30** and **30′**.

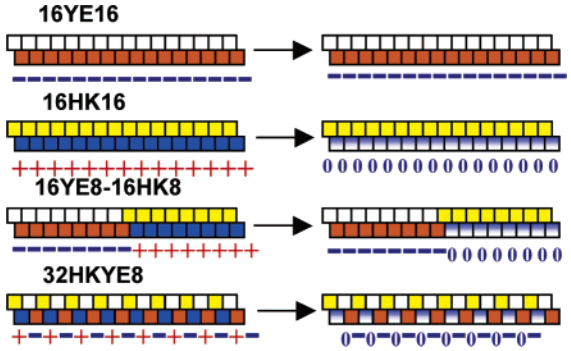

Figure 10. The schematic representation of the polypeptide charge distribution before and after post-translational chemical modification.

different positions (Figure 8, lanes 7 and 15, Figure 9, filled (black) series). The resultant carbamylated polypeptides were less mobile on SDS-PAGE and were difficult to visualize by Coomassie or silver staining or western blotting.

A comparison of the ESI mass spectra of **27d** (H6-YEHK4-H6) with **27d′** (carbamylated H6-YEHK4-H6) and **28d (**H6-HKYE4-H6) with **28d′** (carbamylated H6-HKYE4-H6) after dialysis confirmed tetracarbamylation of the deformylmethionylated polypeptides (see Supporting Information for more information).

4. Mobility of Polypeptides on Gel Electrophoresis. The polypeptides constructed by the described method are comprised of the same small peptide building blocks yet differ in the architecture of repetitive units. The degree of polymerization or block-copolymerization of the polypeptides varies as do the fusion sequences employed at the *C*- and *N*-termini. Post-translational chemical transformations allow quantitative modifications of all selected residues in the polymers with a consequent variation of the polymer properties. Together, these systematic changes enable investigation of the influence of the polypeptide structure on the physical properties of the polypeptide. As an example, gel electrophoresis under the denaturing conditions was employed to demonstrate structure−physical property correlations. Under denaturing conditions, these polypeptides can be viewed as unstructured biopolymers with incorporation of charged amino acids at precise position in the molecule.

Polypeptides with Low Molecular Weight. Figure 9 illustrates the mobility of polypeptides relative to the molecular weight ladders from four libraries with the repetitive sequences YEHK and HKYE flanked with two different fusion constructs at the termini (**27b−h)** H6-YEHKn-H6, (**28c−h)** C2-YEHKn-H6, (**29b−d)** H6-HKYEn-H6, and (**30c−f**) C2-HKYEn-H6, both before and after carbamylation. Intact polypeptides (**27−30**) where positive and negative charge was balanced (Figure 9 (empty series) and Figure 10) demonstrated little if any difference in gel mobility among the polypeptides with equivalent molecular weight. However, the mobility of **27−30** was retarded (Figure 9, empty series). The corresponding carbamylated polypeptides (**27′−30′**) showed a substantial decrease in mobility (Figure 9, filled series). Carbamylation of lysine residues and elimination of positively charged centers result in an overall uncompensated negative charge proportional to the number of the glutamic residues (Figure 10).

Though SDS binds the polypeptides in a molecular weight specific ratio and therefore enables polypeptide separation by the molecular weight, side chain modifications, such as glycosylation or phosphorylation, significantly affect gel mobility.[10] In contrast to native polypeptides, the mobility of carbamylated compounds was sensitive to the repetitive unit and terminal sequences. The HKYE repetitive polypeptides had greater gel mobility than the corresponding YEHK derived polypeptide. Interestingly, polypeptide libraries with H6 sequences at both termini had a slightly higher mobility than those with a combination of C2 and H6 termini.

The polypeptides in their modified and unmodified forms represent different types of charged polymers (Figure 10): negatively charged polypeptides: type 1a possesses a uniform distribution of charge, type 1b with cluster of negative charge; positively charged polypeptides: type 2; neutral polypeptides: type 3a where there is a uniform distribution of positive and negative charge, type 3b with clusters of negative and positive charge, type 3c composed of neutral amino acids.

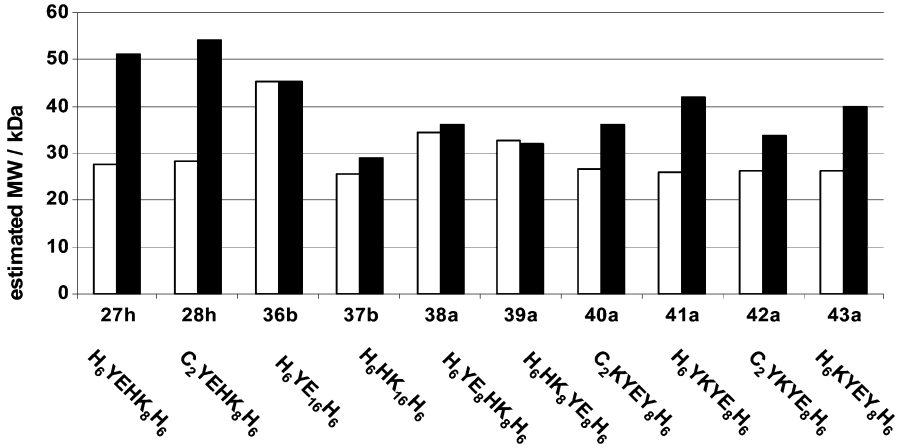

Figure 11. Comparison of mobility of native and carbamylated polypeptides comprised of 271 amino acids on 12% SDS-PAGE. White: native, black: carbamylated. For reference, these constructs with different repetitive units were shown in Figure 4 and were labeled in red color.

A comparison of different polypeptides with the same number of residues (271 amino acids) in modified and unmodified state is shown in Figure 11. From Figure 11, it is clear that polypeptide mobility varies significantly; however, several tendencies can be observed: (1) negatively charged polypeptides (type 1) show very low mobility (**36b** (H6-16YE16-H6), **27h′** (H6-YEHK8-H6), **28h′** (H6-HKYE8-H6)). (2) The YEHK polypeptide family has the greatest shift in mobility on carbamylation (type 3a to type 1a transition). (3) Carbamylation has very little effect on the mobility of the HK polypeptide **37a** (H6-16HK16-H6) (type 2 to type 3c transition). (4) Block-copolymerized polypeptides **38a** (H6-YE8HK8-H6) and **39a** (H6-HK8YE8-H6) demonstrate the least change in mobility on carbamylation (type 3b to 1b transition).

The polypeptides **27h** (H6-YEHK8-H6), **38a (**H6-YE8HK8-H6), and **39a (**H6-HK8YE8-H6) have exactly the same amino acid composition and yet migrate differently on the gel both in their modified and unmodified states. The difference in mobility can be understood by consideration of the distribution of charge within the polypeptides. In the case of unmodified polypeptides (**27h** (H6-YEHK8-H6) type 3a and **38a (**H6-YE8HK8-H6) or **39a (**H6-HK8YE8-H6) type 3b), negative charge is localized in one domain of the overall neutral polypeptide and therefore apparently influences the polypeptide mobility. On carbamylation, polypeptide (**27h′** (H6-YEHK8-H6) represents a type 1a peptide and has a uniform distribution of negative charge whereas modified **38a′** (H6-YE8HK8-H6) or **39a′** (H6-HK8YE8-H6) polypeptides have a localized negative charge domain.

Polypeptides with High Molecular Weights. To investigate the effect of terminal fusion sequences on the mobility of the polymer, the long but easily cleavable A2 fusion adaptive sequence (Figure 1), 45 amino acids, 5 kDA at the *N*-terminal or alternative short fusion sequences at both polypeptide termini were employed (Figure 4b and Figure 12). Figure 12 demonstrates that the addition of small fusion sequences does not have a pronounced effect on the mobility of larger repetitive polypeptides. Both untreated and carbamylated YEHK21 polypeptides **27j** (H6-YEHK21-H6) and **28j** (C2-YEHK21-H6) have nearly the same mobility as the YEHK21 polypeptide, **32d** (YEHK21) absent the terminal strands. However, the addition of the quite large A2 fusion adaptive sequence slightly increase the mobility of polypeptide.

The tendencies observed for low molecular weight polypeptide such as the influence of the repetitive building block sequence, post-translational chemical modification, or block-copolymerization could be extended over the full range of polypeptide molecular weights (Figures 13 and 14).

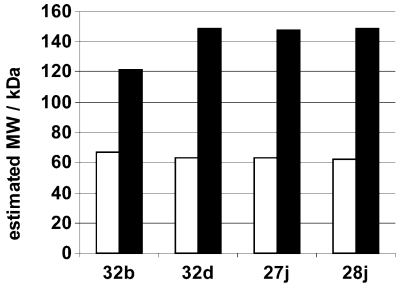

Figure 12. The effect of *N*-terminal sequences on the mobility of the YEHK21. A2 adaptive sequence (**32b**), the CNBr cleaved A2 (**32d**), H6-H6 (**27j**), and C2-H6 (**28j**) on 8% SDS PAGE. White: intact, black: carbamylated.

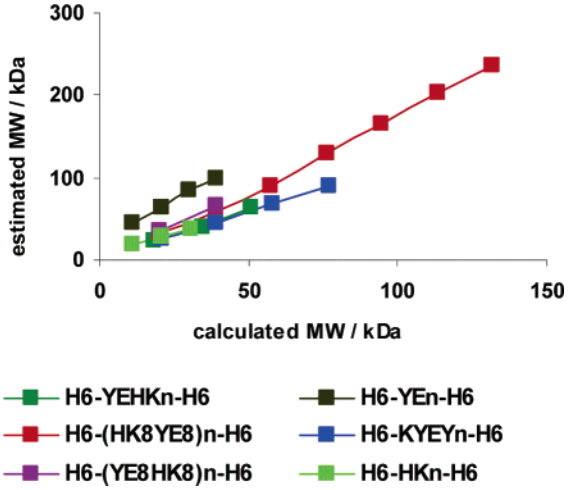

Figure 13. Calculated and estimated molecular weight from the mobility of the polypeptide from libraries with different architectures of repetitive unit but the same *N*- and *C*-terminal groups, green: H6-YEHKn-H6 (**27**), black: H6-YEn-H6 (**36**), light green: H6-HKn-H6 (**37**), red: H6-(HK8YE8)n-H6 (**39**), blue: H6-KYEYn-H6 (**40**), indigo: H6-(YE8HK8)n-H6 (**38**).

The uniformly negatively charged polypeptides H6-YEn-H6 (**36**) and carbamylated C2-YEHKn-H6 (**28′**) (type 1a) demonstrate the lowest gel mobility, while positively charged H6-HKn-H6 (**37**) (type 2) polypeptides and neutral polypeptides with uniformly distributed charge H6-YEHKn-H6 (**27**), H6-HKYEn-H6 (**29**), H6-KYEYn-H6 (**40**), and H6-YKYE-H6 (**42**) (type 3a) are more mobile. The neutral block-copolymers such as H6-(YE8HK8)n-H6 (**38**) and H6-(HK8YE8)n-H6 (**39**) (type 3b), where negative and positive charge lies in separate domains,

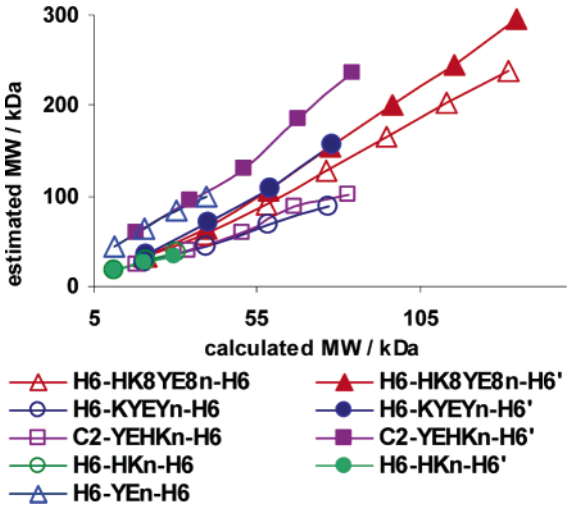

Figure 14. Comparison of calculated and estimated molecular weight of selected repetitive polypeptides in their modified and unmodified state Empty: intact polypeptides, filled: carbamylated polypeptides, indigo: C2-YEHKn-H6 (**28**, **28′**), blue: H6-YEn-H6 (**36**), green: H6-HKn-H6 (**37**, **37′**), red: H6-(HK8YE8)n-H6 (**39**, **39′**), dark blue: H6-KYEYn-H6 (**40**, **40′**).

have an intermediate gel mobility that only decreases slightly on carbamylation.

Conclusion

Libraries of polypeptides were constructed from a small number of specific building blocks; building blocks that were assembled initially into monomeric repetitive sequences were then oligomerized or propagated to form block-copolymers. These libraries enable an investigation of the correlation of the polypeptide structure with physical and biophysical properties such as folding, aggregation, amyloid formation, and metal or nanoparticle recognition. In this paper, the repetitive unit design, the post-translational chemical modification, the nature of *C*- and *N*-fusion constructs, and the degree of polymerization or block-copolymerization were shown to define the physical properties of biopolymers as revealed by SDS-PAGE.

Acknowledgment. We gratefully acknowledge Profs. Richard P. Cunningham, Charles P. Scholes, and Caro-Beth Stewart and Drs. Evelina Loghin and Sara Seepo at The University at Albany-SUNY for their kind help and discussions. This research was supported by the Microelectronics Advanced Research Corporation (MARCO), the Defense Advanced Research Programs Agency (DARPA), and the New York State Office of Science, Technology and Academic Research (NYSTAR) through the Interconnect Focus Center (IFC) and the Materials, Structures, and Devices (MSD) Center under the auspices of Prof. Alain E. Kaloyeros. MALDI and ESI mass spectra were measured at the Mass Spectrometry Facility of the Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute. The National Science Foundation is gratefully acknowledged for the purchase of the MALDI-TOF instrument (NSF Grant CHE-0078056) and the LC/MS system (NSF Grant CHE-0091892).

Supporting Information Available. Mass spectral data and molecular weight estimations derived from gel electrophoretic measurements. Comparisons of electrophoretic gels from post-translationally modified polypeptides and electrophoretic gels from peptides with different *N*-, and *C*-terminal sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) (a) *Protein-Based Materials*; McGrath, K. P.,; Kaplan, D., Eds.; Birkhauser: Boston, MA, 1999. (b) Zhang, S. *Nat. Biotechnol.* **2003**, *21*, 1171−1178. (c) Langer, R.; Tirell, D. A. *Nature* **2004**, *428*, 487−492. (d) Ratner, B. D.; Bryant, S. J. *Annu. Rev. Biomed. Eng.* **2004**, *6*, 41−75. (e) Wright, E. R.; Conticello, V. P. *Adv. Drug Delivery Rev.* **2002**, *54*, 1057−1073. (f) Urry, D. W. *Biorefineries: Ind. Processes Prod.* **2006**, *2*, 217−251.

(2) (a) Lee, T. A. T.; Cooper, A.; Apkarian, R. P.; Conticello, V. P. *Adv. Mater.* **2000**, *12*, 1105−1110. (b) Wright, E. R.; McMillan, R. A.; Cooper, A.; Apkarian, R. P.; Conticello, V. P. *Adv. Funct. Mater.* **2002**, *2*, 149−154. (c) Panitch, A.; Yamaoka, T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1999**, *32*, 1701−1703. (d) Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A. *Science* **1998**, *281*, 389−392. (e) Winkler, S.; Wilson, D.; Kaplan, D. L. *Biochemistry* **2000**, *39*, 12739−12746. (f) Winkler, S.; Szela, S.; Avtges, P.; Valluzzi, R.; Kirschner, D. A.; Kaplan, D. L. *Int. J. Biol. Macromol.* **1999**, *24*, 265−270.

(3) (a) Haider, M.; Megeed, Z.; Ghandehari, H. *J. Controlled Release* **2004**, *95*, 1−26. (b) Haider, M.; Leung, V.; Ferrari, F.; Crissman, J.; Cappello, J.; Ghandehari, H. *Mol. Pharmaceutics* **2005**, *2*, 139−150. (c) Cappello, J.; Crissman, J. W.; Crissman, M.; Ferrari, F. A.; Textor, G.; Wallis, O.; Whitledge, J. R.; Zhou, X.; Burman, D.; Aukerman, L.; Stedronsky, E. R. *J. Controlled Release* **1998**, *53*, 105−117. (d) Urry, D. W.; Woods, T. C.; Hayes, L. C.; Xu, J.; McPherson, D. T.; Parker, T. M.; Iwama, M.; Furutan, M.; Hayashi, T.; Murata, M. *Tissue Eng. Novel Delivery Syst.* **2004**, 31−54.

(4) (a) McPerson, D. T.; Xu, J.; Urry, D. W. *Protein Expression Purif.* **1996**, *7*, 51−57. (b) Meyer, D. E.; Trabbic-Carlson, K.; Chilkoti, A. *Biotechnol. Prog.* **2001**, *17*, 720−728. (c) Meyer, D. E.; Chilkoti, A. *Nat. Biotechnol.* **1999**, *17*, 1112−1115.

(5) (a) Zurdo, J. *Protein Pept. Lett.* **2005**, *12*, 171−187. (b) Lednev, I. K.; Ermolenkov, V. V.; Higashiya, S.; Popova, L. A.; Topilina, N. I.; Welch, J. T. *Biophys. J.* **2006**, *91*, 3805−3818.

(6) (a) Scherzinger, E.; Sittler, A.; Schweiger, K.; Heiser, V.; Lurz, R.; Hasenbank, R.; Bates, G. P.; Lehrach, H.; Wanker, E. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999** *96*, 4604−4609. (b) Lynn, D. G.; Meredith, S. C. *J. Struct. Biol.* **2000**, *130*, 153−173. (c) Fraser, P. E.; Nguyen, J. T.; Surewicz, W. K.; Kirschner, D. A. *Biophys. J.* **1991**, *60*, 1190−1201. (d) Sunde, M.; Serpell, L. C.; Bartlam, M.; Fraser, P. E.; Pepys, M. B.; Blake, C. C. *J. Mol. Biol.* **1997**, *273*, 729−739. (e) Jaikaran, E. T.; Higham, C. E.; Serpell, L. C.; Zurdo, J.; Gross, M.; Clark, A.; Fraser, P. E. *J. Mol. Biol.* **2001**, *308*, 515−525. (f) Topilina, N. I.; Higashiya, S.; Rana, N.; Ermolenkov, V. V.; Kossow, C.; Carlsen, A.; Ngo, S. C.; Wells, C. C.; Eisenbraun, E. T.; Dunn, K. A.; Lednev, I. K.; Geer, R. E.; Kaloyeros, A. E.; Welch, J. T. *Biomacromolecules* **2006**, *7*, 1104−1111.

(7) (a) Mershin, A.; Cook, B.; Kaiser, L.; Zhang, S. *Nat. Biotechnol.* **2005**, *23*, 1379−1380. (b) Zhao, X.; Zhang, S. *Trends Biotechnol.* **2004**, *22*, 470−476. (c) Lamm, M. S.; Rajagopal, K.; Schneider, J. P.; Pochan, D. J. *J. Am. Chem. Soc.* **2005**, *127*, 16692−16700. (d) Aggeli, A.; Bell, M.; Boden, N.; Carrick, L. M.; Strong, A. E. *Angew. Chem., Int. Ed.* **2003**, *42*, 5603−5606. (e) Maskarinec, S. A.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2005**, *16*, 422−426. (f) Clark, J.; Singer, E. M.; Korns, D. R.; Smith, S. *Biotechniques* **2004**, *36*, 992−996, 998−1001.

(8) (a) Ferrari, F. A.; Cappello, J. In *Protein-Based Materials*; McGrath, K. P., Kaplan, D., Eds.; Birkhauser: Boston, MA, 1997; pp 37−60. (b) Tirrell, J. G.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. In *Protein-Based Materials*; McGrath, K. P., Kaplan, D., Eds.; Birkhauser: Boston, MA, 1997; pp 61−99. (c) McPherson, D. T.; Morrow, C.; Minehan, D. S.; Wu, J.; Hunter, E.; Urry, D. W. *Biotechnol. Prog.* **1992**, *8*, 347−352. (d) Feeney, K. A.; Tatham, A. S.; Gilbert, S. M.; Fido, R. J.; Halford, N. G.; Shewry, P. R. *Biochim. Biophys. Acta* **2001**, *1546*, 346−355. (e) Kostal, J.; Mulchandani, A.; Chen, W. *Macromolecules* **2001**, *34*, 2257−2261. (f) Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, Marquet, M.; Ferrari, F. *Biotechnol. Prog.* **1990**, *6*, 198−202. (g) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213−1214. (h) Fukushima, Y. *Biopolymers* **1998**, *45*, 269−279. (i) McMillan, R. A.; Lee, T. A. T.; Conticello, V. P. *Macromolecules* **1999**, *32*, 3643−3648. (j) McGrath, K. P.; Tirrell, D. A.; Kawai, M.; Mason, T. L.; Fournier, M. J. *Biotechnol. Prog.* **1990**, *6*, 188−192. (k) Padgett, K. A.; Sorge, J. A. *Gene* **1996**, *16*, 31−35. (l) Zhou,

Y.; Wu, S.; Conticello, V. P. *Biomacromolecules* **2001**, *2*, 111−125. (m) Ferrari, F. A.; Richardson, C.; Chambers, J.; Causey, S. C.; Pollock, T. J.; Capello, J.; Crissman, J. W. U.S. patent 5,243,-038, 1993. (n) Lixin, M. *Biomacromolecules* **2006**, *7*, 2099−2107.

(9) (a) Won, J.; Barron, A. E., *Macromolecules* **2002**, *35*, 8281−8287. (b) Meyer, Dan E.; Chilkoti, A. *Biomacromolecules* **2002**, *3*, 357−367.

(10) Sambrook, J.; Russel, D. W. *Molecular Cloning*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

(11) (a) Krejechi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427−1432. (b) Wang, J.; Parkhe, A. D.; Tirrell, D. A.; Thompson, L. K. *Macromolecules* **1996**, *29*, 1548−1553. (c) Cantor, E. J.; Atkins, E. D. T.; Cooper, S. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Biochem.* **1997**, *122*, 217−225. (d) Panitch, A.; Matsuki, K.; Cantor, E. J.; Cooper, S. J.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1997**, *30*, 42−49.

(12) Amersham Pharmacia Biotech. HisTrap Kit (Product Code 18-1212-00 AB), Wikströms, Sweden, 2000.

(13) (a) Grimm, R.; Huber, R.; Neumeier, T.; Seidl, A.; Haslbeck, M.; Seiberta, F. S. *Anal. Biochem.* **2004**, 330, 140−144. (b) Whitelegge, J. P.; Le Coutre, J.; Lee, J. C.; Engel, C. K.; Privé, G. G.; Faull, K. F.; Kaback, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10695−10698.

(14) (a) Goeden-Wood, N. L.; Conticello, V. P.; Muller, S. J.; Keasling, J. D. *Biomacromolecules* **2002**, *3*, 874−879. (b) Hecht, M. H.; Das, A.; Go, A.; Bradley, L. H.; Wei, Y. *Protein Sci.* **2004**, *13*, 1711−1723. (c) Brown, C. L.; Aksay, I. A.; Saville, D. A.; Hecht, M. H. *J. Am. Chem. Soc.* **2002**, *124*, 6846−6848. (d) Wang, W.; Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2760−2765. (e) Xu, G.; Wang, W.; Groves, J. T.; Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3652−3657. (f) West, M. W.; Wang, W.; Patterson, J.; Mancias, J. D.; Beasley, J. R.; Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11211−11216. (g) West, M. W.; Beasley, J. R.; Hecht, M. H. *Protein Eng.* **1997**, *10* (Suppl.), 38. (h) Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8729−30. (i) Bradley, L. H.; Thumfort, P. P.; Hecht, M. H. In *Protein Design*; Guerois, R., de la Paz, M. L., Eds.; Humana: Totowa, NJ, 2006; pp 53−69.

(15) (a) Krejchi, M. T.; Cooper, S. J.; Deguchi, Y.; Atkins, E. D. T.; Fournier, M. J. Mason, T. L.; Tirrell, D. A. *Macromolecules* **1997**, *30*, 5012−5024. (b) Parkhe, A. D.; Cooper, S. J.; Atkins, E. D. T.; Fournier, M. J. Mason, T. L.; Tirrell, D. A. *Int. J. Biol. Macromol.* **1998**, *23*, 251−258. (c) Smeenk, J. M.; Otten, M. B. J.; Thies, J.; Tirrell, D. A.; Stunnenberg, H. G.; van Hest, J. C. M. *Angew. Chem., Int. Ed.* **2005**, *44*, 1968−1971. (d) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471−5475. (e) Dougherty, M. J. Ph.D. Dissertation, University of Massachusetts at Amherst, 1993.

(16) Zhen, L.; Janusz, P. *J. Proteome Res.* **2004**, *3*, 567−571.

BM061098Y