Published in final edited form as:

Chembiochem. 2013 May 27; 14(8): 968–978. doi:10.1002/cbic.201300069.

Multiple site-selective insertions of non-canonical amino acids into sequence-repetitive polypeptides

I-Lin Wu^[a], Melissa A. Patterson^[a], Prof. Holly E. Carpenter Desai^{[a],[b]}, Prof. Ryan A. Mehl^[c], Prof. Gianluca Giorgi^[d], and Prof. Vincent P. Conticello^[a]

Vincent P. Conticello: vcontic@emory.edu

^[a]Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, GA 30322, USA, Fax: (+1)404-727-6586

^[b]Department of Chemistry, North Georgia College & State University, Dahlonega, GA 30597, USA

^[c]Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA

^[d]Department of Chemistry, University of Siena, via A. Moro, I-53100, Siena, Italy

Abstract

A simple and efficient method is described for introduction of non-canonical amino acids at multiple, structurally defined sites within recombinant polypeptide sequences. *E. coli* MRA30, a bacterial host strain with attenuated activity for release factor 1 (RF1), is assessed for its ability to support the incorporation of a diverse range of non-canonical amino acids in response to multiple encoded amber (TAG) codons within genetic templates derived from superfolder GFP and an elastin-mimetic protein polymer. Suppression efficiency and isolated protein yield were observed to depend on the identity of the orthogonal aminoacyl-tRNA synthetase/tRNA^{CUA} pair and the non-canonical amino acid substrate. This approach afforded elastin-mimetic protein polymers containing non-canonical amino acid derivatives at up to twenty-two positions within the repeat sequence with high levels of substitution. The identity and position of the variant residues was confirmed by mass spectrometric analysis of the full-length polypeptides and proteolytic cleavage fragments resulting from thermolysin digestion. The accumulated data suggest that this multi-site suppression approach permits the preparation of protein-based materials in which novel chemical functionality can be introduced at precisely defined positions within the polypeptide sequence.

Keywords

non-sense suppression; non-canonical amino acid; protein engineering; mutagenesis; protein material

Introduction

Protein polymers derived from sequence-repetitive polypeptides^[1] have been employed increasingly for the creation of novel functional materials in which the architectural uniformity of the biosynthetic macromolecules promotes self-assembly into structurally defined supramolecular materials that exhibit distinctive properties as a consequence of this

Correspondence to: Vincent P. Conticello, vcontic@emory.edu.

Supporting information for this article is available on the WWW under <http://www.chembiochem.org> or from the author.

sequence specificity (e.g., lamellar crystallites, dimensionally defined fibrils, surface-adhesive coatings, substrate-selective sorbents, electrically conducting nano-wires, liquid crystalline mesophases, environmentally-responsive nanoparticles and nanostructured hydrogels).^[2-9]

This approach contrasts with the synthesis of conventional organic polymers in which statistical mixtures are obtained as products and, consequently, the ability to create functional domains of defined architecture is limited to “living” polymerizations or more operationally complex methodologies.^[10] However, protein polymers have a significant disadvantage vis-à-vis conventional organic polymers in that their chemical diversity is generally limited to functional groups that comprise the side-chains of canonical amino acid residues.^[11] Several scenarios may be contemplated in which it would be advantageous to expand the functional group complexity of synthetic protein-based materials beyond that of the canonical amino acids. These situations include the creation of multi-component conjugates,^[12] the introduction of chemically unique and/or bio-orthogonal reactive groups,^[13] or substitution with amino acids that would enhance materials properties or ease of fabrication.

Recent advances in synthetic biology promise to push the boundary of chemical tolerance of biosynthetic systems to the point in which one might envision the preparation of sequence-repetitive polypeptides in which multiple non-canonical amino acids have been incorporated at defined positions in the polypeptide sequence in high yield and degree of substitution. Functionally enhanced protein polymers have been prepared previously through incorporation of non-canonical amino acids using biosynthesis.^[14-21] However, this approach has relied primarily on the native *E. coli* biosynthetic machinery to effect global substitution of genetically encoded canonical amino acids with structurally similar non-canonical amino acids.^[22] While significant changes in materials properties are often observed due to the chemical influence of the non-canonical amino acid, this sense-codon suppression strategy places restrictions on the chemical diversity of non-canonical amino acids that could be employed as substrates for incorporation into the target polypeptides. The endogenous aminoacyl-tRNA synthetases display significantly greater affinity for the native substrate than for the non-canonical amino acids, which limits the expression conditions to defined culture media that have been depleted of the canonical amino acid and usually require the use of auxotrophic *E. coli* strains. Moreover, the isolated yields of the target proteins are often low and the degree of substitution with the non-canonical amino acid is not necessarily quantitative.^[23]

In contrast, the most common procedure for site-selective substitution relies on the use of a biologically orthogonal aminoacyl-tRNA synthetase to charge a non-canonical amino acid to an orthogonal tRNA within an *E. coli* bacterial host system.²⁴ The non-canonical amino acid is inserted into a growing polypeptide chain during ribosomal biosynthesis in response to a codon that is complementary to the anti-codon of the charged tRNA species, usually a suppressor tRNA that recognizes the amber (TAG) termination codon. Schultz, et al.,^[24] among others,^[25] have employed protein evolution to engineer biologically orthogonal aminoacyl-tRNA synthetases that recognize and charge a chemically diverse range of non-canonical amino acids to suppressor tRNAs. However, while nonsense suppression can be effective for single insertions of non-canonical amino acids, the overall suppression efficiency decreases as the number of termination codons within the genetic template increases due to the multiplicative effect of competition with release factors at each suppression site. Thus, protein polymers that contain one or more suppression sites within the repeat sequence would not be expected to afford a high-level of substitution or a high yield of the full-length polypeptide.

Recently, several research groups have described an approach to multi-site non-sense suppression in genetically modified *E. coli* strains.^[26–27] These researchers were able to identify genetic modifications within the *E. coli* genome that enabled deletion of the *prfA* gene encoding RF1; the release factor that recognizes the amber (TAG) codon and induces hydrolysis of the peptidyl-tRNA bond. These *E. coli* strains support multi-site suppression with non-canonical amino acids for proteins in which up to ten amber codons had been introduced at distinct sites within the coding sequence. Target polypeptides could be produced in isolated yields that far surpassed those in the presence of wild-type RF1 levels and were sufficient for preparative scale isolation. While RF1 deletion represents a promising strategy for the preparation of proteins with multiple non-canonical amino acids encoded at defined sites within the polypeptide sequence, these mutant *E. coli* strains are not generally available and the corresponding wild-type strains require significant genetic modification to achieve the desired phenotype. However, several *E. coli* strains have been described in which simple mutations have been introduced into *prfA* gene that attenuate the translation termination activity of RF1 and support enhanced read-through of amber termination codons.^[28] We describe herein the use of one of these strains, *E. coli* MRA30, to facilitate incorporation of multiple non-canonical amino acids into recombinant polypeptides in high-yield and degree of substitution. This synthetic approach provides access to protein polymers that have been modified extensively with non-canonical amino acids at specific positions within the sequence, potentially enhancing the functional repertoire of this novel class of materials.

Results and Discussion

The success of this strategy requires a host strain that is permissive for suppression of amber codons and the presence of an orthogonal aminoacyl-tRNA synthetase that can charge a non-canonical amino acid to a suppressor tRNA. *E. coli* strain MRA30 (MG1655, *recA56 prfA1*)^[28] was employed as the bacterial host system for these experimental studies. This strain carries a conditionally lethal mutation in the *prfA* gene within its genome, which encodes a structural variant of RF1 with attenuated translation termination activity. Zhang, et al.,^[28] first reported that a single Arg137Pro mutation within the sequence of *E. coli* RF1 altered its interaction with the ribosome and supported higher levels of amber suppression within genetic templates when supplemented with genes encoding suppressor tRNA species. Strains with this mutation, *prfA1*, display a temperature-sensitive phenotype, in which growth is impaired at elevated temperature (42 °C). Even within the permissive temperatures range (30–37 °C), the activity of the mutant RF1 is weakened vis-à-vis the wild-type protein, although the bacterial cells maintained a growth phenotype that was only slightly reduced with respect to an isogenic *prfA*⁺ *E. coli* strain. For comparison, we employed *E. coli* DH10B, a common laboratory expression strain that supports single-site insertion of non-canonical amino acids, as a control host in these studies since it presumably displays wild-type levels of RF1 activity.

A number of orthogonal aminoacyl-tRNA synthetase/tRNA pairs have been developed that facilitate incorporation of single-site substitutions in response to amber codons. We chose to examine two different experimental systems to assess the generality of this synthetic approach. The first one was derived from the *M. barkeri* pyrrolysyl-tRNA synthetase/tRNA(Pyl)^[29] pair, which is naturally orthogonal in *E. coli* and need not be structurally modified to avoid tRNA cross-charging between the heterologous pair and the endogenous systems of the bacterial host. This pyrrolysyl-tRNA synthetase/tRNA(Pyl) pair comprises a native suppressor system in which the CUA anti-codon of tRNA(Pyl) selectively decodes the amber termination codon and co-translationally inserts the variant amino acid with high selectivity.^[30] Several mutants of pyrrolysyl-tRNA synthetase have been reported in which structural modifications have been introduced within the substrate recognition site on the

basis of analysis of the crystal structure of the homologous enzyme from *M. mazei*.^[31] In this study, the single (Tyr349Phe) and double (Tyr271Ala, Tyr349Phe) mutants of the wild-type *M. barkeri* PylRS were employed to direct incorporation of *N*-tBoc/Aloc-lysine and *N*-Cbz-lysine derivatives, respectively, (Figure 1) into polypeptide sequences in response to amber codons.

As a point of comparison, the second system was derived from a *M. jannaschii* tyrosyl-tRNA synthetase/tRNA(Tyr) pair, in which multiple rounds of mutagenesis and selection were employed to create an artificial state of orthogonality vis-à-vis the *E. coli* host system.^{[32],[33]} Schultz, et al., have developed methods for evolution of mutant tyrosyl-tRNA synthetases that are capable of activation of a chemically diverse range of substituted aromatic amino acids in *E. coli* host strains. We chose two of these mutant tyrosyl-tRNA synthetases for investigation; the *p*-benzoylphenylalanyl-tRNA synthetase (BpaRS)^[34–35] and *p*-azidophenylalanyl-tRNA synthetase (AzFRS),^[36–37] in combination with an orthogonal tRNA, Nap1, that was derived from MjtRNA^{CUA}(Tyr).^[33b] Taken together, mutant aminoacyl-tRNA synthetases from these two classes of organisms should be able to support the incorporation of a variety of non-canonical amino acids into recombinant polypeptides (Figure 1).

Codon-optimized synthetic genes encoding the respective aaRS mutants were cloned into the plasmid-based vector into which we had previously cloned the gene encoding the appropriate suppressor tRNA under the control of the constitutive *lpp* promoter.^[38] The aaRS genes were placed under the transcriptional control of the *P_{Ltet}* promoter^[39] and were expressed constitutively in a vector background based on the low copy p15A replicon (Supporting Information).^[19a]

We constructed two test substrates to assess the suppression efficiency of the RF1-attenuated *E. coli* host system and evaluate its suitability for incorporation of non-canonical amino acids at multiple positions within a recombinant polypeptide sequence. The first genetic construct was derived from a codon-optimized variant of the green fluorescent protein superfolder (sfGFP),^[40] in which three amber (TAG) codons were inserted at permissible loop positions (Gln157/Lys158, Glu172/Asp173, and Leu194/Leu195) within the coding sequence.^[41] This fluorescent reporter system permitted the use of flow cytometric analysis to monitor the incorporation of the non-canonical amino acids into the test substrate. The second genetic construct was derived from an elastin-mimetic protein polymer^[42–43] based on a repeat sequence in which the introduction of an amber codon was employed to specify the position and frequency of occurrence of a non-canonical amino acid (Xaa) (*vide infra*). This genetic template permitted evaluation of the potential of our approach for the preparation of sequence-repetitive polypeptides that had been substituted with multiple non-canonical amino acids. In addition, we could interrogate the efficiency of decoding as a function of the number of termination sites within the polypeptide coding sequence. The genes encoding the respective test substrates were which supports inducible expression in response to the presence of the gratuitous inducer IPTG.

Multiple-site suppression of sfGFP derivatives

Flow cytometric analysis of expression cultures of the sfGFP reporter was employed to monitor incorporation of the non-canonical amino acids in *E. coli* strains DH10B and MRA30 (Figure 2 and Supporting Information). The data clearly indicate an increase in the fluorescence activity of the sfGFP-(TAG)₃ derivative in *E. coli* strain MRA30 in the presence of the non-canonical amino acids, with respect to the negative control, i.e., absence of analogue, but only under conditions of co-expression of the mutant aaRS/tRNA pair. In comparison, the fluorescence activity under identical conditions for sfGFP derivatives in *E. coli* strain DH10B was significantly reduced with respect to that in *E. coli* MRA30 and often

barely distinguishable from the negative control, even in the presence of the aaRS/tRNA pair. Taken together, the flow cytometry data demonstrate that *E. coli* strain MRA30 supports a higher level of sfGFP expression in the presence of the non-canonical amino acids, which suggests that the suppressor tRNA competes more effectively for amino acid insertion due to the attenuated activity of the mutant RF1. However, we note that the fluorescent signal of the sfGFP derivative falls below that of the positive control to varying degrees for the different non-canonical amino acids. The observed decrease in fluorescence might result in part from the loop insertions within the sfGFP-(TAG)₃ and/or the presence of the non-canonical amino acids, either of which may inhibit protein folding and fluorophore formation. However, the relative efficiency of co-translational insertion of the amino acid analogue also appears to play a significant role. On the basis of the flow cytometry data, the effective level of incorporation of the pyrrolysine analogues decreases in the order; tBoc $\approx$ Aloc > Cbz, which agrees with previous experimental evaluations of incorporation efficiency in single-site suppression studies (Figure 2 and Supporting Information).^[31b, 31c] In addition, the incorporation of AzF into the sfGFP template appears significantly more facile than that of Bpa; an observation that is supported from preparative-scale expression studies (*vide infra*). The observed differences between non-canonical amino acids may reflect several criteria, including the efficiency of substrate recognition and tRNA charging, the binding efficiency of the charged tRNA to EF-Tu and the ribosome, and residual competition between the aminoacylated suppressor tRNA and the mutant release factor.

In order to quantitate the effectiveness of suppression for the various analogues, preparative-scale expression of the sfGFP derivatives was performed in *E. coli* MRA30 in the presence of the respective non-canonical amino acids. The sfGFP derivatives were isolated from LB or TB expression cultures supplemented with the amino acid analogues using immobilized metal affinity chromatography and the yield was determined using the Bradford assay (Table 1). Western blot analyses of whole-cell lysates indicated a significantly higher level of expression of the target protein in *E. coli* strain MRA30 in comparison to expression under identical conditions in *E. coli* strain DH10B. The isolated yields differ between the non-canonical amino acid derivatives in a manner that reflects the trend observed in the flow cytometric analysis of the sfGFP derivatives and Western blot analyses. The deconvoluted ESI-mass spectrometric analysis of the purified sfGFP variants indicated that the mass of the main peak was greater than that of the wild-type sfGFP sequence (Figure 3 and Table 1). The mass increment in each case was consistent with the presence of three equivalents of the non-canonical amino acid that had been added to supplement the respective expression culture. No evidence could be detected for infiltration of canonical amino acids into the target protein within the limits of experimental detection for any of the cases under consideration. For the most effective analogues, the expression level of the modified sfGFP proteins compared well to that of recombinant polypeptide derivatives containing the same number of suppression sites that had been expressed in *E. coli* strains in which *prfA* had been deleted.^[26–27]

Multi-site suppression of sequence-repetitive polypeptides

The sfGFP expression studies demonstrated that multiple termination codons could be decoded with a structurally diverse variety of non-canonical amino acids to afford chemically modified proteins in good yield. However, questions lingered regarding limitations in the scope of the process, in particular, the number of termination codons that could be efficiently decoded as non-canonical amino acids. This issue is of critical importance for the synthesis of protein polymers modified with non-canonical amino acids, as each sequence repeat may contain one or more suppression sites. Therefore, sequence-repetitive polypeptides of modest size may contain a significant number of amber termination sites within the coding sequence, any of which may result in truncation of the

full-length construct upon recognition with endogenous RF1. To address this issue, we assessed the suppression efficiency of the RF1-attenuated *E. coli* system using a test substrate based on a sequence-repetitive polypeptide derived from an elastin-mimetic repeat sequence.

Elastin is a native protein-based biomaterial that is the primary structural component underlying the elastomeric mechanical response of compliant tissues in vertebrates.^[43] Polypeptides that mimic the physiologically relevant properties of native elastin have been the focus of significant research interest for their potential as responsive biomaterials in tissue-engineering and biotechnological applications.^[44] The macromolecular properties of the elastomeric domains can be emulated by synthetic polypeptides that are composed of a concatenated sequence of native oligopeptide repeat motifs; the most common of which is the pentapeptide (Val-Pro-Gly-Val-Gly). In addition, elastin-mimetic polypeptides display a well-defined correlation between repeat sequence and macromolecular properties (*vide infra*), which enables the creation of a wide variety of synthetic elastin analogues with tailorable physical properties.^[42] As such, elastin-mimetic polypeptides represent the best-characterized biomimetic protein-based materials that have been prepared and analyzed to date.

The elastin-mimetic repeat sequence, [(Val-Pro-Gly-Val-Gly)₂(Val-Pro-Gly-Xaa-Gly)(Val-Pro-Gly-Val-Gly)₂] elastin-(TAG), was used as the basis for design of a synthetic DNA cassette (i.e., DNA monomer) in which the introduction of an amber (TAG) codon was employed to specify the position and frequency of occurrence of a non-canonical amino acid (Xaa) (Supporting Information). Previous studies have demonstrated that the elastin-mimetic repeat sequence is tolerant of non-conservative substitutions at the fourth position of the pentapeptide unit,^[42] which makes this position an ideal site for introduction of non-canonical amino acids. A seamless cloning procedure^[45] was employed for head-to-tail self-ligation of the DNA cassette to afford a pool of DNA concatemers,^[46] from which individual clones corresponding to twelve, elastin-(TAG)₁₂, and twenty-two, elastin-(TAG)₂₂, repeats of the DNA monomer were isolated (Figure 4). An important consequence of this procedure is that the concatemeric genes encode direct sequence repeats of the DNA monomer. Since release factor recognition of termination codons depends on the local sequence context,^{[28],[47]} the repetitive sequence of the elastin-mimetic polypeptide suggests that suppression efficiency should be uniform for each amber site within the sequence.

SDS-PAGE electrophoretic analysis of the whole-cell lysates from the expression cultures was employed to evaluate the production of the two elastin-mimetic polypeptides under IPTG induction in the presence of the non-canonical amino acids (Figure 4). In each case, a significant level of protein accumulation was observed for *E. coli* MRA30, but only in the presence of the respective aaRS/tRNA pair that was specific for the corresponding non-canonical amino acid. In contrast, *E. coli* DH10B strain did not display significant levels of protein production under identical conditions, nor was protein accumulation observed in the *E. coli* MRA30 strain upon induction in the absence of the non-canonical amino acid. The substituted elastin-mimetic polypeptides could be purified from the cell lysate using either immobilized metal affinity chromatography or inverse temperature transition cycling (4 °C/ 37 °C).^[48] SDS-PAGE electrophoretic analysis of the purified polypeptides indicated that the bands corresponding to the purified elastin derivatives migrated at the correct mass relative to the protein standards (Supporting Information).

For the pyrrolysyl-tRNA synthetase/tRNA(Pyl) system, the isolated yields of the elastin derivatives reflected a similar trend to those of the corresponding sfGFP derivatives in that the *N*-³-tBoc-lysine derivative afforded the highest yield, while the *N*-³-Cbz-lysine derivative provided the lowest yield (Table 2). The elastin-(Cbz)₂₂ polypeptide could not be

isolated using either purification protocol, as substantial degradation of the polypeptide occurred during the isolation. However, Western blot analysis provided strong evidence for initial production of the corresponding full-length polypeptide. As observed earlier for the sfGFP derivatives, the yields of the AzF-substituted elastin-mimetic polypeptides were higher than for corresponding Bpa-substituted analogues. The discrepancies in protein yield observed between non-canonical amino acids for a given aaRS/tRNA system presumably reflect the considerations described above for the corresponding sfGFP derivatives. These considerations include differences in the enzymatic activity of the respective aaRS mutants for the non-canonical amino acids, and in suppression efficiency of the charged suppressor tRNAs during ribosomal chain elongation.

MALDI-TOF mass spectrometry was employed to assess the efficiency of production of the full-length elastin-mimetic polypeptides that resulted from multi-site suppression at the encoded amber positions. While mass spectra could be obtained for all of the elastin-(Xaa)₁₂ derivatives, only elastin-(Aloc)₂₂ gave reliable data among the longer elastin derivatives. Presumably, this situation reflected the difficulties in desorption/ionization of relatively large, hydrophobic proteins and/or concomitant gas-phase decomposition of the polypeptides in the mass spectrometer. The MALDI-TOF mass spectra in each case were consistent with a high level of substitution of the respective non-canonical amino acid into the corresponding elastin polypeptide sequence (Table 2). Notably, in the case of elastin-(Aloc)₂₂, the major species in the MALDI-TOF mass spectrum was observed at the correct *m/z* value for intact full-length polypeptide in which the appropriate number of non-canonical amino acids had been inserted (Table 2). In addition to the major peak, the MALDI-TOF mass spectra of proteins isolated using inverse transition temperature cycling indicated the presence of a series of species at lower *m/z* values (Figure 5). For the elastin-(Aloc)_{*n*} derivatives (*n* = 12, 22), the difference between successive peaks corresponded to an average mass of 2143 units (u), which approximates the calculated mass of 2160 u for the [(Val-Pro-Gly-Val-Gly)₂(Val-Pro-Gly-Aloc-Gly)(Val-Pro-Gly-Val-Gly)₂] repeat unit. Assuming the peaks correspond to singly charged species, the observed masses were consistent with truncation products that result from termination at amber suppression sites upstream of the TAA codon that specifies termination of the entire open reading frame of the elastin expression cassette. These species were presumably carried through with the major product, i.e., full-length polypeptide, during the inverse transition temperature purification process. The truncation peaks are much more noticeable in the mass spectrum of elastin-(Aloc)₂₂, especially at lower *m/z* values. This situation may reflect the greater difficulty in achieving gas-phase ionization of the larger elastin derivative, rather than a skewing of the distribution of truncation products towards lower molar mass species. These observations suggest that termination at amber sites can compete with suppression, however the SDS-PAGE analysis of the purified proteins indicates that the efficiency of suppression in the RF1-attenuated strains is such that full-length polypeptides are obtained as the major product even at the relatively high content of amber codons present within these sequence-repetitive polypeptides. Notably, the truncation products were not observed in the MALDI-TOF mass spectra of protein samples that had been purified using immobilized metal affinity chromatography, which requires the presence of an intact C-terminal deca-histidine tag of the full-length polypeptide sequence in order to effect the separation. This observation provides further support for the hypothesis that the truncation products resulted from termination at internal amber sites within the coding sequence of the elastin-mimetic polypeptides.

In order to ascertain the presence and position of the non-canonical amino acid, high-resolution tandem mass spectrometry was performed on cleavage fragments from proteolytic digestion of the elastin-mimetic polypeptides. However, the sequence repeats of the protein polymer lacked convenient sites for residue-specific endoproteinases. The metalloproteinase

thermolysin was chosen for these analyses since it catalyzes the selective hydrolysis of peptide bonds at positions that are located directly upstream of hydrophobic amino acid residues within the polypeptide sequence. The high content of valine residues within the repeat units of the elastin-mimetic polypeptides suggested that the largest cleavage fragments would correspond to peptide sequences that incorporated the non-canonical amino acid residue. Tandem MALDI-TOF/TOF mass spectrometry of the products from thermolysin cleavage of the elastin-mimetic polypeptides were consistent with incorporation of the respective non-canonical amino acids at the amber sites within the coding sequence of the elastin repeat unit (Supporting Information). Mass spectrometric analyses of thermolysin cleavage products derived from protein polymers corresponding to either 12 or 22 repeats (where available) of the elastin-mimetic sequences were similar in terms of the masses and relative population of cleavage products. In addition, the peptide fragmentation pattern derived from collision-induced dissociation of selected gas-phase ions corresponding to the proteolytic cleavage products indicated the presence of the non-canonical amino acid at the appropriate position within the sequence. Mass differences were observed between fragment ions that could be attributed to the presence of *N*-allyloxycarbonyl lysine (212 u), *N*-benzyloxycarbonyl lysine (262 u), or *p*-benzoylphenylalanine (251 u) residues. In the case of the tBoc-lysine derivative, the tBoc group was lost in the gas phase and the observed mass difference (128 u) corresponded to that of the deprotection product lysine rather than the tBoc-lysine derivative (228 u). MALDI-TOF/TOF mass spectra of the *p*-azidophenylalanine-containing peptide fragments indicated reaction of the azido group under laser irradiation.

The identities of the peptides corresponding to the thermolysin cleavage fragments were confirmed independently by accurate mass measurements and by MSⁿ experiments carried out by using electrospray ionization coupled to a linear ion trap and an Orbitrap analyzer (Table 3 and Figure 6). The experimental errors on the mass measurements were within 3 parts per million with respect to the theoretical values calculated from the elemental composition for the suppositious gas-phase species. In addition, good agreement was observed between measured and theoretical isotopic clusters. Tandem MS data indicated the presence of the non-canonical amino acids in terms of differences in mass between secondary ion cleavage fragments and in terms of loss of reactive functionality associated with the unique amino acid (Fig 5 and Supporting Information). These data provided strong evidence for the presence of the non-canonical amino acids, Aloc, Cbz, Bpa and AzF, in the appropriate sequence context within the elastin-mimetic repeat unit. As observed previously in the tandem MALDI-TOF/TOF mass spectrometric analyses, the presence of the tBoc derivative was inferred from the appearance of lysine within the fragment ions, which resulted from gas-phase loss of the protecting group. The mass spectrometric analysis did not provide evidence for the presence of canonical amino acids at amber codon sites within the elastin repeat sequence within the experimental limits of detection. These data, in combination, suggest that site-specific substitution of the non-canonical amino acids occurs in response to amber codons within the elastin template and that the sole detectable alternative to suppression is low-level termination due to the attenuated activity of RF1 within *E. coli* MRA30.

Conclusion

This biosynthetic approach enables the preparation of sequence-repetitive polypeptides in which multiple non-canonical amino acids have been incorporated at defined positions in the polypeptide sequence. The accumulated data suggest that attenuation of RF1 activity can significantly enhance the incorporation of non-canonical amino acids through decreased effectiveness in competition with aminoacylated suppressor tRNAs for recognition of the amber codon sites encoded within the translational reading frame of the recombinant

polypeptide. The efficacy of the process depends strongly on the identity of the orthogonal aminoacyl-tRNA synthetase/suppressor tRNA pair; nonetheless, expression yields approached those estimated from similar studies conducted within RF1 knockout strains of *E. coli*. Significant levels of recombinant protein expression were detected with sequence-repetitive proteins encoding up to 22 amber suppression sites. While the process remains to be optimized with respect to isolated protein yield, these results suggest that sequence-repetitive polypeptides, i.e., protein polymers, can be prepared with high levels of incorporation of non-canonical amino acids with selectivities that eclipses those of the sense-suppression approach. These substitutions occur within a uniform and controllable sequence context, which should enhance the potential for coupling of the chemical properties of the variant side-chains to the materials properties of the protein polymer. The Bpa^[49] and AzF^[50] residues are of particular interest in this regard as they provide photo-crosslinkable sites that may facilitate processing of the corresponding elastin-mimetic protein polymers into elastic networks under environmentally benign conditions.

Schultz, et al., have reported conditions for the co-translational incorporation of a vast repertoire of chemically diverse phenylalanine derivatives into single amber sites encoded within recombinant polypeptides.^[24] In addition, several research laboratories have described mutant versions of pyrrolysyl-tRNA synthetases that enable insertion of functionally modified derivatives using a similar approach.^{[31], [51]} The multi-site suppression method described herein can leverage the utility of these systems for the preparation of substituted protein polymers in which the materials properties can be enhanced through incorporation of novel chemical functionality at precisely defined positions within the sequence. In addition, many native protein-based materials incorporate repetitive sequences that include multiple sites of post-translational modification that are critical for appropriate biological function, e.g., *trans*-hydroxyproline in collagens,^[52–53] or DOPA residues in mussel adhesive proteins.^[54] The multi-site suppression techniques represents an attractive approach to the synthesis of these functional materials, which are challenging to obtain from *in vitro* chemical or *in vivo* enzymatic modification of the native protein sequences.

Experimental Section

Amino acid analogues

The amino acid derivatives *N*-(*tert*-butoxycarbonyl)-*L*-lysine (tBOC), *N*-benzyloxycarbonyl-*L*-lysine (Cbz), *para*-benzoyl-*L*-phenylalanine (Bpa) and *para*-azido-*L*-phenylalanine were purchased Bachem Bioscience, Inc. (King of Prussia, PA), and *N*-allyloxycarbonyl-*L*-lysine (Aloc) was obtained from Sigma Chemical Co.

Plasmid construction

A synthetic gene encoding a superfolder GFP variant, sfGFP(TAG)₃, with three amber codons inserted at sites between amino acid residues 157–158, 172–173, and 194–195 was isolated via restriction digestion with *Bam*H I and *Eco*R I and cloned into compatible sites in plasmid pIL5, an expression vector derived from pQE-80L with a modified polylinker (supporting information), in which the GFP variant was cloned upstream of an in-frame deca-histidine tags at the *C*-terminus of the full-length reading frame. The sequence of the expression cassette was verified after isolation of the plasmid clone.

A synthetic DNA monomer encoding the elastin repeat sequence [(Val-Pro-Gly-Val-Gly)₂-(Val-Pro-Gly-(amber)-Gly)-(Val-Pro-Gly-Val-Gly)₂] was synthesized and cloned into the *Bam*H I/*Hin*D III-digested pZER0-1 to generate pIL41. A seamless DNA concatemerization procedure^[46] was employed to afford a pool of DNA concatemers. The pool of multimers was isolated from the gel and cloned into the *Bbs* I/*Bsm*B I sites of plasmid pIL41. The

ligation product was transformed into competent cells of *E. coli* strain Top10F. Plasmids containing 12 and 22 repeats of the Elastin-UAG monomer were isolated. The concatemeric DNA cassettes were liberated from the cloning plasmid by sequential restriction digestion with endonucleases *Bbs*I and *Bsm*B I and isolated using agarose gel electrophoresis. The purified DNA cassettes were cloned into the compatible *Bsa*I sites of pIL5 to afford plasmids pIL43.5 and pIL43.11, which contained 12 and 22 repeats of Elastin-UAG, respectively.

A duplex DNA cassette encoding tRNA sequence was synthesized by annealing the synthetic oligonucleotide cassettes. The synthetic gene was cloned into compatible *Eco*R I/ *Pst*I sites within plasmid pSU81.^[38] Plasmid pSU81 was digested with *Nhe*I and *Pvu*I to afford the tRNA gene as a duplex DNA cassette under the transcriptional control of the *lpp* promoter and *rrn*C terminator sequences. The synthetic DNA encoding the RNA expression cassette was cloned into compatible sites within plasmid pHEC2, a modified version of plasmid pME1 into which *Nhe*I and *Pvu*I restriction sites had been inserted using inverse PCR from synthetic oligonucleotide primers. The synthetic gene encoding wild-type aminoacyl-synthetase was cloned into compatible *Kpn*I/*Xba*I restriction sites within pHEC2 in which the aaRS gene was under control of the P_{Ltet} promoter. Single Y271A and double Y371A/Y349F mutations were introduced into the wild-type MbPylRS gene template using Quik-Change mutagenesis from gene-specific primers. The optimized BpaRS,^[35] with the following mutation sites: A31V, Y32G, E107P, D158S, I159S, was introduced into the MjTyrRS gene template using Quik-Change mutagenesis from gene-specific primers. The primers and details for plasmid construction are described in the Supporting Information.

Incorporation of amino acid analogues into recombination protein

The expression vectors encoding sfGFP(TAG)₃ or elastic-mimetic proteins were co-transformed with aminoacyl-synthetase/tRNA pairs into *E. coli* strains DH10B or MRA30. Single colonies of the expression strains were inoculated into sterile LB or TB broth supplemented with the appropriate antibiotics (100 µg/mL ampicillin and 34 µg/mL chloramphenicol) as required for plasmid maintenance. The overnight culture was diluted 50-fold the next morning into fresh media, and cells were allowed to grow at 30°C until the OD₆₀₀ reached approximately 1.0. A final concentration of 1mM IPTG and amino acid analogues were added to induce recombinant protein expression, and cells were harvested 12h post-induction.

Purification of recombinant protein

Harvested cells were re-suspended in lysis buffer (50 mL, 50 mM sodium phosphate, 300 mM NaCl, pH 7.0) and stored at -80°C. The frozen cells were lysed by three freeze/thaw cycles. Lysosome (1 mg/mL), EDTA-free protease inhibitor cocktail, benzonase (25 units/mL), and MgCl₂ (1mM) were added to the lysate and the mixture was incubated shaking at 4 °C for 12 h. The cell lysate was centrifuged at 14,000g for 30 min at 4 °C. The supernatant from the sfGFP expression was loaded directly onto a column containing cobalt-charged TALON resin and followed by a series of washes with lysis buffer (20mL) containing 20 mM imidazole. The target protein was eluted with elution buffer (20 mL, 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0) and dialyzed (MWCO = 10 kDa) against distilled de-ionized water (5 × 4 L). The dialysate was concentrated by ultrafiltration using Amicon filters (NMWL 10 kDa) to a final volume of 1 mL. Protein concentration was measured by the Bradford assay. Elastins were purified through repetitively cycling through the inverse temperature transition between 30 °C and 4 °C. A final concentration of 2 M NaCl was added into the initial supernatant from the elastin protein expression cultures to induce phase separation above the lower critical solution temperature (LCST). The mixture

was incubated at 30 °C for 30 min to precipitate the target protein. At this temperature, the elastins derivatives became insoluble and formed large micron-sized aggregates. The suspension was centrifuged at 9500g and 25 °C for 20 min. The pellet containing the precipitated elastin derivative was re-suspended in cold (4 °C) lysis buffer, incubated on ice for 30 min and centrifuged at 20000g and 4 °C for 20 min. This thermal cycling process was repeated until elastin derivatives were purified. The pure elastins were dialyzed against distilled deionized water (5 × 4 L). The dialysate was lyophilized to produce colorless solids and the protein yields were measured in terms of dry weight per unit volume of culture.

Flow cytometric analysis

Aliquots (1 mL) of *E. coli* cells from expression cultures were grown until the OD₆₀₀ reached approximately 1.0 absorbance units. The cultures were centrifuged at 4000g and 4 °C for 10 min and re-suspended in 1 mL of phosphate-buffered saline pH 7.4. Flow cytometry was performed using a LSRII flow cytometer (Beckton Dickinson) equipped with a 100 mW solid-state laser emitting at 488 nm for the excitation of sfGFP, a 505 nm LP dichroic mirror and a 530/30 bandpass filter. Forward scatter (FSC), sideward scatter (SSC) and green fluorescence were acquired by FACSDiva software. The specific instrumental gain settings for these measurements were as follows: FSC = 250, SSC = 300, F1 = 302. The maximum of each fluorescence histogram (number of events as a function of fluorescence) was scaled to 10000 to facilitate comparison of the histograms of sfGFP with different non-canonical amino acids. Data was analyzed using FlowJo software (TreeStar.com).

Thermolysin cleavage reactions

Purified elastin derivatives were dissolved in sterile water at a concentration of approximately 1 mg/mL. Dithiothreitol (DTT) was added as final concentration 10 mM and the mixture was incubated at 100 °C for 30 min to denature the protein. After the reaction mixture was cooled to 37 °C, thermolysin was added to the denatured protein solution to a concentration of 1 part in 50 (w/w) with respect to protein and the reaction mixture was incubated at 37 °C for 12 h. The products from the proteolysis reaction were passed through a C18 spin column (Thermo Scientific, Inc.) to remove salts and thermolysin. The peptide fragments were eluted using acetonitrile:water (70/30, v/v) mixture and the solution was employed directly for mass spectrometric analysis.

Mass spectrometry

ESI mass spectra of sfGFP derivatives were acquired by LTQ-FT mass spectrometer (ThermoElectron, San Jose, CA) in the positive ion mode (200–2000 *m/z*) using a spray voltage of 4 kV, a capillary voltage of 41 V, a sheath gas flow rate of 20 arbitrary units. The tube lens voltage was 215 V, and the AGC setting was 5e+05. Molar masses of elastins with pyrrolysine analogues were determined by MALDI-TOF MS on an Applied Biosystems Voyager System STR mass spectrometer in the positive linear mode. Ferulic acid was used as matrix at a concentration of 10 mg/mL in a mixture of 75 % acetonitrile and 0.2% formic acid in deionized water. The protein solution (1 mg/mL in distilled water) was mixed with the matrix solution in a ratio of 1:10 and dried under air. Bovine serum albumin (BSA) was used as a standard for external calibration.

Tandem MS of selected thermolysin-digested peptides were carried out using ESI and MALDI mass spectrometry. Electrospray measurements have been carried out on a LTQ-XP-Orbitrap instrument (Thermo, Bremen, Germany). Operating conditions of the ESI source were as follows: spray voltage 4.5 kV; capillary temperature 200°C; sheath gas (nitrogen) flow rate ca. 0.75 L/min. Solutions of each peptide (1×10⁻⁵ M, methanol) were infused with a rate of 5 µL/min. CID collision energy: 0.5–1.0 eV (laboratory frame). MALDI spectra were recorded on an Ultraflex TOF/TOF instrument (Bruker Daltonics

GmbH, Bremen, Germany) equipped with LIFT capability. A pulsed Nd:YAG laser at a wavelength of 355 nm was operated at a frequency of 100 Hz. The source was operated in the positive mode with an acceleration voltage of 25.0 kV, and a delayed extraction time of 20 ns was applied. A solution of the alpha-cyano-4-hydroxycinnamic acid (4-HCCA, Sigma-Aldrich, Milan, Italy) matrix in water/acetonitrile (1/1, v/v) at a concentration of 10 mg/ml was mixed with the peptide sample in equal amount and 0.5 µl of this solution was deposited onto the MALDI target. Mass spectra were acquired from 100 laser shots in the range m/z 500–3000 in the reflectron mode. MS/MS experiments were performed under LID conditions with the LIFT cell voltage parameters set at 19.0 kV for a final acceleration of 29.5 kV (reflector voltage) and a pressure in the LIFT cell around 4×10^{-7} mbar. Data analysis of MS/MS data has been carried out using the package Bio Tools™ (Bruker Daltonics, Bremen, Germany).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

V.P.C. acknowledges the NIH (7R01HL083867) and the Emory University Research Committee for financial support. We are grateful to Dr. Monica Ryden-Aulin and Jaroslav Belotserkovsky for providing *E. coli* strain MRA30 and to Professor William McClain for providing plasmid pSU81.

References

1. Payne, SC.; Patterson, M.; Conticello, VP. Protein Engineer Handbook. Lutz, S.; Bornscheuer, UT., editors. Vol. 2. Wiley-VCH; 2009. p. 915-938.
2. a) Krejchi MT, Atkins ED, Waddon AJ, Fournier MJ, Mason TL, Tirrell DA. *Science*. 1994; 265:1427–1432. [PubMed: 8073284] b) Cantor EJ, Atkins ED, Cooper SJ, Fournier MJ, Mason TL, Tirrell DA. *J Biochem*. 1997; 122:217–225. [PubMed: 9276692] c) Parkhe AD, Cooper SJ, Atkins ED, Fournier MJ, Mason TL, Tirrell DA. *Int J Biol Macromol*. 1998; 23:251–258. [PubMed: 9849622]
3. a) West MW, Wang W, Patterson J, Mancias JD, Beasley JR, Hecht MH. *Proc Natl Acad Sci U S A*. 1999; 96:11211–11216. [PubMed: 10500156] b) Topilina NI, Higashiya S, Rana N, Ermolenkov VV, Kossow C, Carlsen A, Ngo SC, Wells CC, Eisenbraun ET, Dunn KA, Lednev IK, Geer RE, Kaloyeros AE, Welch JT. *Biomacromolecules*. 2006; 7:1104–1111. [PubMed: 16602727] c) Higashiya S, Topilina NI, Ngo SC, Zagorevskii D, Welch JT. *Biomacromolecules*. 2007; 8:1487–1497. [PubMed: 17388563] d) Topilina NI, Sikirzhytsky V, Higashiya S, Ermolenkov VV, Lednev IK, Welch JT. *Biomacromolecules*. 2010; 11:1721–1726. [PubMed: 20553038]
4. a) Tulpur A, Henderson DB, Mao M, Caba B, Davis RM, Van Cott KE, Ducker WA. *Langmuir*. 2005; 21:1497–1506. [PubMed: 15697300] b) Henderson DB, Davis RM, Ducker WA, Van Cott KE. *Biomacromolecules*. 2005; 6:1912–1920. [PubMed: 16004428]
5. a) Kostal J, Mulchandani A, Gropp KE, Chen W. *Environ Sci Technol*. 2003; 37:4457–4462. [PubMed: 14572100] b) Lao UL, Chen A, Matsumoto MR, Mulchandani A, Chen W. *Biotechnol Bioeng*. 2007; 98:349–355. [PubMed: 17486654] c) Wang Y, Kiick KL. *J Am Chem Soc*. 2005; 127:16392–16393. [PubMed: 16305215] d) Top A, Kiick KL. *Adv Drug Deliv Rev*. 2010; 62:1530–1540. [PubMed: 20562016]
6. a) Scheibel T, Parthasarathy R, Sawicki G, Lin XM, Jaeger H, Lindquist SL. *Proc Natl Acad Sci U S A*. 2003; 100:4527–4532. [PubMed: 12672964] b) Sharma N, Top A, Kiick KL, Pochan DJ. *Angew Chem Int Ed Engl*. 2009; 48:7078–7082. [PubMed: 19691075]
7. a) Yu SM, Conticello VP, Zhang G, Kayser C, Fournier MJ, Mason TL, Tirrell DA. *Nature*. 1997; 389:167–170. [PubMed: 9296493] b) Yu SM, Tirrell DA. *Biomacromolecules*. 2000; 1:310–312. [PubMed: 11710117] c) Zhuravel MA, Davis NE, Nguyen ST, Koltover I. *J Am Chem Soc*. 2004; 126:9882–9883. [PubMed: 15303837]

8. a) Lee TAT, Cooper A, Apkarian RP, Conticello VP. *Adv Mater* (Weinheim, Ger). 2000; 12:1105–1110. b) Dreher MR, Simnick AJ, Fischer K, Smith RJ, Patel A, Schmidt M, Chilkoti A. *J Am Chem Soc*. 2008; 130:687–694. [PubMed: 18085778] c) Kim W, Thevenot J, Ibarboure E, Lecommandoux S, Chaikof EL. *Angew Chem Int Ed Engl*. 2010; 49:4257–4260. [PubMed: 20446331]
9. a) Wright ER, McMillan RA, Cooper A, Apkarian RP, Conticello VP. *Adv Funct Mater*. 2002; 12:149–154. b) Wright ER, Conticello VP. *Adv Drug Deliv Rev*. 2002; 54:1057–1073. [PubMed: 12384307] c) Nagapudi K, Brinkman WT, Leisen J, Thomas BS, Wright ER, Haller C, Wu X, Apkarian RP, Conticello VP, Chaikof EL. *Macromolecules*. 2005; 38:345–354. d) Petka WA, Harden JL, McGrath KP, Wirtz D, Tirrell DA. *Science*. 1998; 281:389–392. [PubMed: 9665877] e) Mi L, Fischer S, Chung B, Sundelacruz S, Harden JL. *Biomacromolecules*. 2006; 7:38–47. [PubMed: 16398496] f) Fischer SE, Mi L, Mao HQ, Harden JL. *Biomacromolecules*. 2009; 10:2408–2417. [PubMed: 19655714] g) Haghighpanah JS, Yuvienko C, Civay DE, Barra H, Baker PJ, Khapli S, Voloshchuk N, Gunasekar SK, Muthukumar M, Montclare JK. *ChemBioChem*. 2009; 10:2733–2735. [PubMed: 19806626] h) Haghighpanah JS, Yuvienko C, Roth EW, Liang A, Tu RS, Montclare JK. *Mol Biosyst*. 2010; 6:1662–1667. [PubMed: 20480093] i) Qu Y, Payne SC, Apkarian RP, Conticello VP. *J Am Chem Soc*. 2000; 122:5014–5015. j) Xu C, Breedveld V, Kopecek J. *Biomacromolecules*. 2005; 6:1739–1749. [PubMed: 15877401] k) Xu C, Kopecek J. *Pharm Res*. 2008; 25:674–682. [PubMed: 17713844]
10. Badi N, Lutz JF. *Chem Soc Rev*. 2009; 38:3383–3390. [PubMed: 20449057]
11. van Hest JC, Tirrell DA. *Chem Commun*. 2001:1897–1904.
12. de Graaf AJ, Kooijman M, Hennink WE, Mastrobattista E. *Bioconj Chem*. 2009; 20:1281–1295. [PubMed: 19191567]
13. Lim RK, Lin Q. *Chem Commun*. 2010; 46:1589–1600.
14. Yoshikawa E, Fournier MJ, Mason TL, Tirrell DA. *Macromolecules*. 1994; 27:5471–5475.
15. Kothakota S, Mason TL, Tirrell DA, Fournier MJ. *J Am Chem Soc*. 1995; 117:536–537.
16. Deming T, Fournier M, Mason T, Tirrell D. *J Macromol Sci Pure Appl Chem*. 1997; 34:2143–2150.
17. van Hest JCM, Kiick KL, Tirrell DA. *J Am Chem Soc*. 2000; 122:1282–1288.
18. a) Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF, Tirrell DA. *Angew Chem Int Ed Engl*. 2001; 40:1494–1496. [PubMed: 11317312] b) Tang Y, Tirrell DA. *J Am Chem Soc*. 2001; 123:11089–11090. [PubMed: 11686725]
19. a) Kim W, George A, Evans M, Conticello VP. *ChemBioChem*. 2004; 5:928–936. [PubMed: 15239049] b) Kim W, McMillan RA, Snyder JP, Conticello VP. *J Am Chem Soc*. 2005; 127:18121–18132. [PubMed: 16366565] c) Kim W, Hardcastle KI, Conticello VP. *Angew Chem Int Ed Engl*. 2006; 45:8141–8145. [PubMed: 17109457]
20. a) Kirshenbaum K, Carrico IS, Tirrell DA. *ChemBioChem*. 2002; 3:235–237. [PubMed: 11921403] b) Zhang K, Diehl MR, Tirrell DA. *J Am Chem Soc*. 2005; 127:10136–10137. [PubMed: 16028902] c) Nowatzki PJ, Franck C, Maskarinec SA, Ravichandran G, Tirrell DA. *Macromolecules*. 2008; 41:1839–1845.
21. Teeuwen RL, van Berkel SS, van Dulmen TH, Schoffelen S, Meeuwissen SA, Zuillhof H, de Wolf FA, van Hest JC. *Chem Commun*. 2009:4022–4024.
22. a) Budisa N. *Angew Chem Int Ed Engl*. 2004; 43:6426–6463. [PubMed: 15578784] b) Link AJ, Tirrell DA. *Methods*. 2005; 36:291–298. [PubMed: 16076455]
23. Takasu A, Kondo S, Ito A, Furukawa Y, Higuchi M, Kinoshita T, Kwon I. *Biomacromolecules*. 2011; 12:3444–3452. [PubMed: 21823658]
24. Liu CC, Schultz PG. *Annu Rev Biochem*. 2010; 79:413–444. [PubMed: 20307192]
25. Furter R. *Protein Sci*. 1998; 7:419–426. [PubMed: 9521119]
26. a) Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, Sakamoto K. *Nucleic Acids Res*. 2010; 38:8188–8195. [PubMed: 20702426] b) Mukai T, Yanagisawa T, Ohtake K, Wakamori M, Adachi J, Hino N, Sato A, Kobayashi T, Hayashi A, Shirouzu M, Umehara T, Yokoyama S, Sakamoto K. *Biochem Biophys Res Commun*. 2011; 411:757–761. [PubMed: 21782790]
27. a) Johnson DBF, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, Xiang Z, Ecker JR, Briggs SP, Wang L. *Nat Chem Biol*. 2011; 7:779–786. [PubMed: 21926996] b) Johnson DB, Wang C, Xu

- J, Schultz MD, Schmitz RJ, Ecker JR, Wang L. *ACS Chem Biol*. 2012; 7:1337–1344. [PubMed: 22662873]
28. a) Ryden SM, Isaksson LA. *Mol Gen Genet*. 1984; 193:38–45. [PubMed: 6419024] b) Zhang S, Ryden-Aulin M, Kirsebom LA, Isaksson LA. *J Mol Biol*. 1994; 242:614–618. [PubMed: 7932718] c) Kaczanowska M, Ryden-Aulin M. *J Bacteriol*. 2004; 186:3046–3055. [PubMed: 15126466]
 29. a) Longstaff DG, Larue RC, Faust JE, Mahapatra A, Zhang L, Green-Church KB, Krzycki JA. *Proc Natl Acad Sci U S A*. 2007; 104:1021–1026. [PubMed: 17204561] b) Namy O, Zhou Y, Gundllapalli S, Polycarpo CR, Denise A, Rousset JP, Soll D, Ambrogelly A. *FEBS Lett*. 2007; 581:5282–5288. [PubMed: 17967457]
 30. a) Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, Zhao G, Kang PT, Green-Church KB, Chan MK, Krzycki JA. *Nature*. 2004; 431:333–335. [PubMed: 15329732] b) Polycarpo C, Ambrogelly A, Berube A, Winbush SM, McCloskey JA, Crain PF, Wood JL, Soll D. *Proc Natl Acad Sci U S A*. 2004; 101:12450–12454. [PubMed: 15314242]
 31. a) Kavran JM, Gundllapalli S, O'Donoghue P, Englert M, Soll D, Steitz TA. *Proc Natl Acad Sci U S A*. 2007; 104:11268–11273. [PubMed: 17592110] b) Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S. *J Mol Biol*. 2008; 378:634–652. [PubMed: 18387634] c) Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S. *Chem Biol*. 2008; 15:1187–1197. [PubMed: 19022179]
 32. a) Santoro SW, Wang L, Herberich B, King DS, Schultz PG. *Nat Biotechnol*. 2002; 20:1044–1048. [PubMed: 12244330] b) Young TS, Ahmad I, Yin JA, Schultz PG. *J Mol Biol*. 2010; 395:361–374. [PubMed: 19852970]
 33. a) Wang L, Schultz PG. *Chem Biol*. 2001; 8:883–890. [PubMed: 11564556] b) Guo J, Melancon CE 3rd, Lee HS, Groff D, Schultz PG. *Angew Chem Int Ed Engl*. 2009; 48:9148–9151. [PubMed: 19856359]
 34. a) Chin JW, Martin AB, King DS, Wang L, Schultz PG. *Proc Natl Acad Sci U S A*. 2002; 99:11020–11024. [PubMed: 12154230] b) Chin JW, Schultz PG. *ChemBioChem*. 2002; 3:1135–1137. [PubMed: 12404640] c) Umanah G, Huang LY, Schultz PG, Naider F, Becker JM. *Adv Exp Med Biol*. 2009; 611:333–335. [PubMed: 19400216]
 35. Stokes AL, Miyake-Stoner SJ, Peeler JC, Nguyen DP, Hammer RP, Mehl RA. *Mol BioSyst*. 2009; 5:1032–1038. [PubMed: 19668869]
 36. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. *J Am Chem Soc*. 2002; 124:9026–9027. [PubMed: 12148987]
 37. Young DD, Young TS, Jahnz M, Ahmad I, Spraggon G, Schultz PG. *Biochemistry*. 2011; 50:1894–1900. [PubMed: 21280675]
 38. Gabriel K, McClain WH. *J Mol Biol*. 2001; 310:543–548. [PubMed: 11439022]
 39. Lutz R, Bujard H. *Nucleic Acids Res*. 1997; 25:1203–1210. [PubMed: 9092630]
 40. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. *Nat Biotechnol*. 2006; 24:79–88. [PubMed: 16369541]
 41. Abedi MR, Caponigro G, Kamb A. *Nucleic Acids Res*. 1998; 26:623–630. [PubMed: 9421525]
 42. Urry, DW.; Luan, C-H.; Harris, CM.; Parker, TM. *Protein-Based Materials*. Kaplan, D.; McGrath, K., editors. Birkhauser; Boston: 1997. p. 133-177.
 43. Conticello, VP.; Carpenter Desai, HE. *Polymer Science: A Comprehensive Reference*. Matyjaszewski, K.; Möller, M., editors. Vol. 9. Elsevier; 2012. p. 71-103.
 44. a) Almine JF, Bax DV, Mithieux SM, Nivison-Smith L, Rnjak J, Waterhouse A, Wise SG, Weiss AS. *Chem Soc Rev*. 2010; 39:3371–3379. [PubMed: 20449520] b) Kim W, Chaikof EL. *Adv Drug Deliv Rev*. 2010; 62:1468–1478. [PubMed: 20441783] c) MacEwan SR, Chilkoti A. *Biopolymers*. 2010; 94:60–77. [PubMed: 20091871] d) Daamen WF, Veerkamp JH, van Hest JC, van Kuppevelt TH. *Biomaterials*. 2007; 28:4378–4398. [PubMed: 17631957]
 45. Padgett KA, Sorge JA. *Gene*. 1996; 168:31–35. [PubMed: 8626061]
 46. a) Goeden-Wood NL, Conticello VP, Muller SJ, Keasling JD. *Biomacromolecules*. 2002; 3:874–879. [PubMed: 12099837] b) McMillan RA, Lee TAT, Conticello VP. *Macromolecules*. 1999; 32:3643–3648.
 47. a) Feinstein SI, Altman S. *Genetics*. 1978; 88:201–219. [PubMed: 346437] b) Bossi L. *J Mol Biol*. 1983; 164:73–87. [PubMed: 6188841] c) Pedersen WT, Curran JF. *J Mol Biol*. 1991; 219:231–

241. [PubMed: 2038055] d) Phillips-Jones MK, Watson FJ, Martin R. J Mol Biol. 1993; 233:1–6. [PubMed: 8377179] e) Pavlov MY, Freistroffer DV, Dincbas V, MacDougall J, Buckingham RH, Ehrenberg M. J Mol Biol. 1998; 284:579–590. [PubMed: 9826500] f) Zhang S, Ryden-Aulin M, Isaksson LA. FEBS Lett. 1999; 455:355–358. [PubMed: 10437803]
48. a) Furgeson DY, Dreher MR, Chilkoti A. J Control Release. 2006; 110:362–369. [PubMed: 16303202] b) McPherson DT, Xu J, Urry DW. Protein Expr Purif. 1996; 7:51–57. [PubMed: 9172783]
49. a) Kauer JC, Erickson-Viitanen S, Wolfe HR Jr, DeGrado WF. J Biol Chem. 1986; 261:10695–10700. [PubMed: 3733726] b) Dorman G, Prestwich GD. Biochemistry. 1994; 33:5661–5673. [PubMed: 8180191]
50. a) Schwyzer R, Caviezel M. Helv Chim Acta. 1971; 54:1395–1400. [PubMed: 5121750] b) Fischli W, Caviezel M, Eberle A, Escher E, Schwyzer R. Helv Chim Acta. 1976; 59:878–879. [PubMed: 1262224] c) Escher E, Schwyzer R. FEBS Lett. 1974; 46:347–350. [PubMed: 4424934]
51. a) Kobayashi T, Yanagisawa T, Sakamoto K, Yokoyama S. J Mol Biol. 2009; 385:1352–1360. [PubMed: 19100747] b) Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A, Chin JW. J Am Chem Soc. 2009; 131:8720–8721. [PubMed: 19514718] c) Fekner T, Li X, Lee MM, Chan MK. Angew Chem Int Ed Engl. 2009; 48:1633–1635. [PubMed: 19156778] d) Li X, Fekner T, Ottesen JJ, Chan MK. Angew Chem Int Ed Engl. 2009; 48:9184–9187. [PubMed: 19882608] e) Li X, Fekner T, Chan MK. Chem Asian J. 2010; 5:1765–1769. [PubMed: 20544791] f) Huang Y, Wan W, Russell WK, Pai PJ, Wang Z, Russell DH, Liu W. Bioorg Med Chem Lett. 2010; 20:878–880. [PubMed: 20074948] g) Wang YS, Russell WK, Wang Z, Wan W, Dodd LE, Pai PJ, Russell DH, Liu WR. Mol Biosyst. 2011; 7:714–717. [PubMed: 21234492] h) Takimoto JK, Dellas N, Noel JP, Wang L. ACS Chem Biol. 2011; 6:733–743. [PubMed: 21545173] i) Ou W, Uno T, Chiu HP, Grunewald J, Cellitti SE, Crossgrove T, Hao X, Fan Q, Quinn LL, Patterson P, Okach L, Jones DH, Lesley SA, Brock A, Geierstanger BH. Proc Natl Acad Sci U S A. 2011; 108:10437–10442. [PubMed: 21670250] j) Virdee S, Kapadnis PB, Elliott T, Lang K, Madrzak J, Nguyen DP, Riechmann L, Chin JW. J Am Chem Soc. 2011; 133:10708–10711. [PubMed: 21710965] k) Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW. J Am Chem Soc. 2011; 133:11418–11421. [PubMed: 21736333]
52. Buechter DD, Paolella DN, Leslie BS, Brown MS, Mehos KA, Gruskin EA. J Biol Chem. 2003; 278:645–650. [PubMed: 12399455]
53. Pinkas DM, Ding S, Raines RT, Barron AE. ACS Chem Biol. 2011; 6:320–324. [PubMed: 21210682]
54. Lim S, Kim KR, Choi YS, Kim DK, Hwang D, Cha HJ. Biotechnol Prog. 2011; 27:1390–1396. [PubMed: 21732552]

Figure 1.
Non-canonical amino acids employed as substrates for co-translational incorporation experiments.

Figure 2.

Flow cytometric analysis of representative sfGFP expression cultures supplemented with the indicated non-canonical amino acids. (Legend: grey dot, *E. coli* MRA30 with wild-type sfGFP (positive control); black, *E. coli* MRA30 with co-expression of synthetase/tRNA in the presence of non-canonical amino acid (0.5 mM); grey, *E. coli* DH10B with co-expression of synthetase/tRNA in the presence of non-canonical amino acid (0.5 mM); black thin line, *E. coli* MRA30 without co-expression of the synthetase/tRNA pair (negative control); and black dash, *E. coli* MRA30 without non-canonical amino acid (negative control)).

Figure 3.

Deconvoluted ESI-mass spectra of sfGFP derivatives isolated from expression cultures of *E. coli* strain MRA30 in the presence of the respective orthogonal synthetase/tRNA pair. The masses are consistent with substitution of the indicated non-canonical amino acids at amber codon sites encoded within the genetic template. Note that the lower mass species in the tBoc derivative result from partial hydrolysis of the *tert*-butoxycarbonate group in the gas-phase.

Figure 4.

A) Repeat sequence for the elastin-mimetic polypeptides, elastin-(TAG)₁₂ and elastin-(TAG)₂₂, indicating the position within the coding sequence of the amber codon. Agarose gel electrophoretogram derived from restriction analysis of the plasmid-based concatemer clones, indicating the relative size of the coding sequences for the elastin concatemers. B) Western blot analyses of SDS-PAGE of whole cell lysates of elastin expression cultures (M, molecular weight standards; Lanes 1–9, pyrrolysine analogues; Lanes 10–13, tyrosine analogues). Lanes 1 and 6, negative controls (no amino acid supplementation); Lane 2, DH10B/elastin-(TAG)₁₂ with Aloc; Lanes 3–5, MRA30/elastin-(TAG)₁₂ with Aloc, tBoc, and Cbz, respectively; Lanes 7–9, MRA30/elastin-(TAG)₂₂ with Aloc, tBoc, and Cbz, respectively; Lanes 10 and 11, MRA330/elastin-(TAG)₁₂ with Bpa and AzF, respectively; Lanes 12 and 13, MRA30/elastin-(TAG)₂₂ with Bpa and AzF, respectively. Target protein bands were visualized through recognition of the His-tag sequence with the anti-(His)₅ antibody.

Figure 5. MALDI-TOF mass spectra of Elastin-(Aloc)₁₂ (A) and Elastin-(Aloc)₂₂ (B). Note the smaller peaks within the respective mass spectra that correspond to truncation products that result from termination.

Figure 6. Tandem ESI-MS/MS analysis of thermolysin cleavage fragments resulting proteolysis of elastin-(Aloc)₁₂ (**A**) and elastin-(AzF)₁₂ (**B**), respectively. The spectra were obtained by selecting the species $[M+H]^+$ having m/z values of 950 and 616, respectively.

Table 1

Superfolder GFP expression and ESI-mass spectrometry data.

sfGFP	Yield (mg/L)	Theoretical m/z (u)	Experimental m/z (u)	(WT) _{exp} (u)	(WT) _{calc} (u)
WT	60.8	28791.3	28789	-	-
Alac	20.5	29428.1	29427	638	637
tBOC	21.0	29476.2	29477	688	685
CBZ	5.5	29578.2	29576	787	787
Bpa	4.1	29545.2	29543	765	754
AzF	13.4	29355.9	29354	563	565

[a] Yield was determined using the Bradford assay on purified protein solutions.

[b] Theoretical molar mass was calculated based on the assumption that *E. coli* methionyl-aminopeptidase results in proteolytic cleavage of the N-terminal methionine to afford the (sfGFP-Met) derivatives.

[c] The difference in experimental molar masses between the substituted sfGFP-(Xaa)₃ and wild-type sfGFP.

[d] Calculated molar mass for three equivalents of the indicated non-canonical amino acid as a peptidic unit (calcd molar mass: Alac, C₁₀H₁₆N₂O₃, 212.25 u; tBoc, C₁₁H₂₀N₂O₃, 228.29 u; Cbz, C₁₄H₁₈N₂O₃, 262.31 u; Bpa, C₁₆H₁₃NO₂, 251.28 u; AzF, 188.19 u.)

Table 2

Summary of elastin protein expression and MALDI-TOF mass spectrometry data.

Elastin derivatives ^[a]	Yield [mg/L] ^{[a], [b]}	Theoretical m/z ^{[c], [d]}	Experimental m/z
(Aloc) ₁₂	10	28540	28539
(Aloc) ₂₂	12	50147	50153
(tBoc) ₁₂	14	28733	28731
(tBoc) ₂₂	27	n.d. ^[e]	n.d.
(Cbz) ₁₂	5.0	29141	29132
(Bpa) ₁₂	8.3	28984	28964
(Bpa) ₂₂	6.4	n.d.	n.d.
(AzF) ₁₂	16.8	28228	28240
(AzF) ₂₂	10.0	n.d.	n.d.

^[a]Yield was determined using the Bradford assay on purified protein solutions and normalized to the volume of the expression culture.

^[b]Elastin proteins containing the *N*-alkoxycarbonyl-lysine derivatives (Aloc, tBoc, and Cbz) were purified via inverse temperature cycling. Elastin proteins containing the Bpa and AzF residues were purified via immobilized metal affinity chromatography.

^[c]Molar masses for the elastin derivatives were calculated based on complete substitution with the non-canonical amino acid at the amber-encoded sites. The calculated molar mass assumes proteolytic cleavage of the *N*-terminal methionine residue as a consequence of the endogenous activity of the *E. coli* methionyl-aminopeptidase.

^[d]Theoretical molar mass determinations were based on the amino acid sequence of the elastin polypeptides:

[(VPGVG)₂VPGXG(VPGVG)₂]_nVPGVGGSDDDDKGH₁₀ as [M+Na]⁺ species for the AlocLys, tBocLys, CbzLys derivatives and as [M+H]⁺ species for the Bpa and AzF derivatives.

^[e]Not determined.

Table 3

High-resolution ESI-mass spectrometric analysis of accurate mass of proteolytic fragments resulting from thermolysin cleavage of elastin-mimetic proteins.

Species	Accurate mass (u)	Elemental composition	Error (ppm)
[VPG(Aloc)KGV+H] ⁺	640.3665	C ₂₉ H ₅₀ O ₇ N ₉	0.07
[VGVPG(CBZ)K+H] ⁺	690.3820	C ₃₃ H ₅₂ O ₇ N ₉	-0.08
[VGVPG(Aloc)KGVPG+H] ⁺	950.5301	C ₄₃ H ₇₂ O ₁₃ N ₁₁	-0.46
[VGVPG(Bpa)GVPG+H] ⁺	989.5218	C ₄₉ H ₆₉ O ₁₂ N ₁₀	-0.28
[VGVPG(AzF)+H] ⁺	616.3217	C ₂₈ H ₄₂ O ₇ N ₉	2.0