Received: 28 April 2020 | Revised: 17 June 2020 | Accepted: 19 June 2020

DOI: 10.1002/pro.3907

THE PROTEIN SOCIETY WILEY

ACCELERATED COMMUNICATION

Harnessing synthetic biology to enhance heterologous protein expression

Shlomo Zarzhitsky[1] | **Alex Jiang**[2] | **Elizabeth E. Stanley**[3] | **Michael H. Hecht**[1]

[1]Department of Chemistry, Princeton University, Princeton, New Jersey, USA

[2]Department of Molecular Biology, Princeton University, Princeton, New Jersey, USA

[3]Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey, USA

Correspondence
Michael Hecht, Department of Chemistry, Princeton University, Princeton, NJ 08544.
Email: hecht@princeton.edu

Present address
Alex Jiang, NIDDK, NIH, Bethesda, MD, USA

Elizabeth E. Stanley, Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA

Abstract

The ability to express heterologous proteins in microbial hosts is crucial for many areas of research and technology. In most cases, however, successful expression and purification of the desired protein require fusion to another protein. To date, all fusion partners have been chosen from natural sequences, which evolved for other purposes, and may not be optimal fusion partners. However, the rise of synthetic biology and protein design make it possible to design and optimize fusion proteins using *novel* sequences that did not arise in nature. Here, we describe a series of De novo Expression Enhancer Proteins (DEEPs) that facilitate high-level expression and facile purification of heterologous proteins and peptides. To test the DEEP system, a de novo protein was fused to several target proteins covering a range of sizes and solubilities. In all cases, fusions to DEEP outperformed fusions to SUMO, a commonly used natural fusion partner. The availability of novel proteins that can be engineered for specific fusion applications could be beneficial to enhance the expression of a wide range of heterologous proteins.

KEYWORDS

de novo protein design, fusion tag, protein expression, purification tag

1 | INTRODUCTION

Expressing heterologous proteins in microbial hosts enables the production of diverse peptides and proteins for myriad scientific and commercial applications. In principle, any sequence can be expressed in a microbial organism. In practice, however, the encoded polypeptide often fails to express, suffers degradation, misfolds, or aggregates.[1,2] Scientists and biotechnologists have responded to these challenges by linking the protein-of-interest (POI) to some other natural protein (a fusion partner) to endow the POI with favorable expression properties.[3,4] Sometimes this strategy succeeds, but often it does not. We suggest that traditional fusion partners often fail because they are natural proteins that evolved to perform specific functions in their host organisms. They did *not* evolve to serve as fusion partners for protein expression; consequently, they often fail at this task. To provide a novel alternative to natural fusion proteins, we designed and developed *novel* sequences to serve as fusion partners for expression and purification. Here, we show that De novo Expression Enhancer Proteins (DEEPs) facilitate high-level expression and facile purification of several different heterologous proteins and peptides.

Toward the goal of enhancing the expression of heterologous proteins, we investigated fusions based on DEEP sequences derived from libraries of novel proteins designed to fold into four-helix bundles. In contrast to naturally evolved sequences, our designed proteins are idealized: Whereas natural proteins often have irregular structures, and occasionally place polar residues in the interior and nonpolar residues on the

surface, our de novo proteins were designed using a "binary code" to specify idealized α-helices with perfect partitioning between polar surfaces and nonpolar interiors.[5] High-resolution structures confirm that proteins from these libraries form idealized structures.[6–8] Moreover, proteins from binary patterned libraries possess biophysical properties that seem ideal relative to natural proteins. For example, protein S-824,[6,9] a 102-residue sequence from our second-generation library, expresses in *E. coli* at extremely high levels, remains soluble under a wide range of conditions, maintains a stable structure at high temperatures, and can be affinity purified without additional tags. Here, we demonstrate that S-824 (hereafter called DEEP1)—and variants of S-824—may also be ideally suited as fusion partners to facilitate the expression and purification of diverse heterologous proteins and peptides, including GFP, Trp Cage, the Alzheimer's peptide (Aβ42), and a membrane-spanning peptide. In all cases, yields were enhanced significantly relative to a well-studied natural fusion partner with a similar size to DEEP. Both enzymatic and chemical cleavages of DEEP1 fusions were optimized to liberate the authentic POIs. These findings demonstrate that at least in some cases, de novo proteins can surpass natural sequences as fusion tags for expression and purification.

2 | RESULTS

2.1 | De novo expression enhancer proteins

Before using DEEP1 as a fusion tag, we characterized its properties as an isolated protein. Expression of DEEP1 in *E. coli* produced >40 mg of purified protein per liter of cells grown in shaker flasks. Following lysis, the protein was purified in a single step using immobilized metal affinity chromatography (IMAC), without the addition of a His tag. No tag is required because the binary code places abundant histidine residues on protein surfaces. The purification and biophysical characterization of DEEP1 are summarized in Figure 1. Circular dichroism (CD) analysis of the protein shows that DEEP1 is predominantly α-helical, as indicated by the minima at 222 and 208 nm. Moreover, as shown in the inset of Figure 1c, the protein is thermally stable, with a T_m close to 100°C.

2.2 | DEEP fusions enhance expression of heterologous proteins

Next, we probed the ability of DEEP1 to serve as a de novo expression enhancer protein by fusing it to several different POIs. We chose POIs spanning a range of sizes, from short peptides (20 residues) to full-length proteins (238 residues), including both soluble and insoluble sequences. We also assessed the ability of the DEEP1 fusion to facilitate the expression of difficult-to-express and membrane-spanning sequences. In all cases, the DEEP1 fusions were compared side by side with fusions to SUMO, a commonly used natural fusion tag that is nearly the same length as DEEP1 (98 and 102 residues, respectively.)

For our first test case, we fused GFP to the C-terminus of DEEP1. Specifically, we used the folding reporter version of GFP (frGFP) developed by Waldo et al.[10] Because the formation of the GFP fluorophore

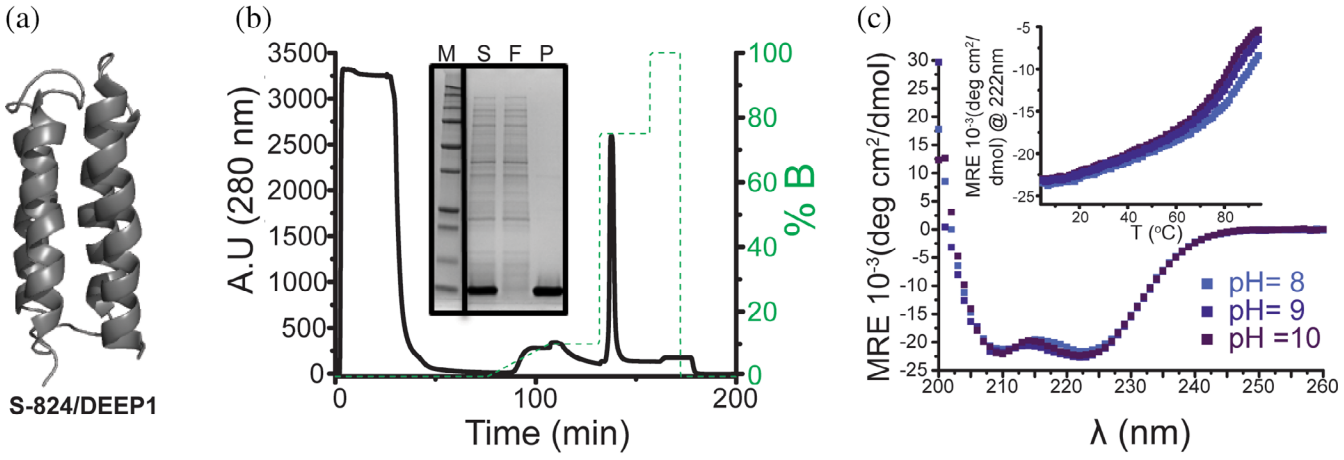

FIGURE 1 Purification and biophysical characteristics of DEEP1. (a) The structure of S-824/DEEP1 (PDB: 1p68). (b) Purification on a Ni-IMAC column. DEEP1 contains 12 surface-exposed histidine residues and readily binds to a Ni-IMAC column. The inset shows SDS-PAGE analysis of the purification process, M—marker, S—soluble fraction (clarified lysate), F—flow through, and P—elution peak. (c) Circular dichroism and thermal denaturation (inset) of the purified protein at several different pHs

requires the correct three-dimensional structure—and because frGFP is a relatively slow folder—the fluorescence of frGFP is often used to report the quantity, correct folding, and solubility of fusion proteins.[10] Figure 2a compares the fluorescence of frGFP fused to either DEEP1 or to SUMO. Fusions to DEEP1 reproducibly yield higher fluorescence, indicating a greater amount of correctly folded GFP. Furthermore, a hexahistidine tag was necessary to affinity purify SUMO-GFP, while DEEP1-GFP was readily purified without an additional tag.

Encouraged by these results, we set out to explore the efficacy of DEEP1 for producing more difficult-to-express POIs, including short, insoluble, and membrane-bound sequences. As an example of a short but soluble sequence, we chose TrpCage.[11] This 20-residue peptide has been described as the smallest fully folded miniprotein and is often used as a model for folding and dynamics. Although TrpCage can be synthesized by solid-phase peptide synthesis, for many applications—notably to prepare isotopically labeled material—expression in vivo is crucial. Yet, the small size of TrpCage can make this challenging.

TrpCage was fused to DEEP1 and to SUMO, the fusion proteins were expressed, and purified on Ni-IMAC columns. As was the case for the GFP fusions, the SUMO construct required an additional His tag, while the DEEP1 fusion was purified without additional tags. The eluted fractions were loaded on a C-18 reverse phase column, and protein concentration was calculated from the area under the peak. As shown in Figure 2b, the fusion of TrpCage to DEEP1 yields significantly more protein than fusion to SUMO.

To facilitate the release of TrpCage from the fusion, we introduced a methionine at the beginning of the TrpCage sequence: M[1]NLYIQWLKDGGPSSGRPPPS, where [1] indicates a cyanogen bromide (CNBr) cleavage site. The purified fusion was treated with CNBr,[12] aliquots were sampled at 10 and 16 hr, and assayed by SDS-PAGE. The identities of DEEP1-TrpCage and the cleaved TrpCage product were confirmed by HPLC/MS (-Figure S1). Expression of the DEEP1-TrpCage fusion in minimal media supplemented with ^{15}N-labeled ammonia, followed by CNBr cleavage readily produced ^{15}N labeled peptide for ongoing NMR studies (*in preparation*). These results demonstrate the efficacy of DEEP1 for expressing short soluble sequences.

Next, we challenged the DEEP1 system to express a POI that is insoluble. For these studies, we chose the Alzheimer's peptide, Aβ42, which readily forms amyloid fibrils. When pursuing experiments with Aβ42, it is tempting to bypass the difficulties associated with

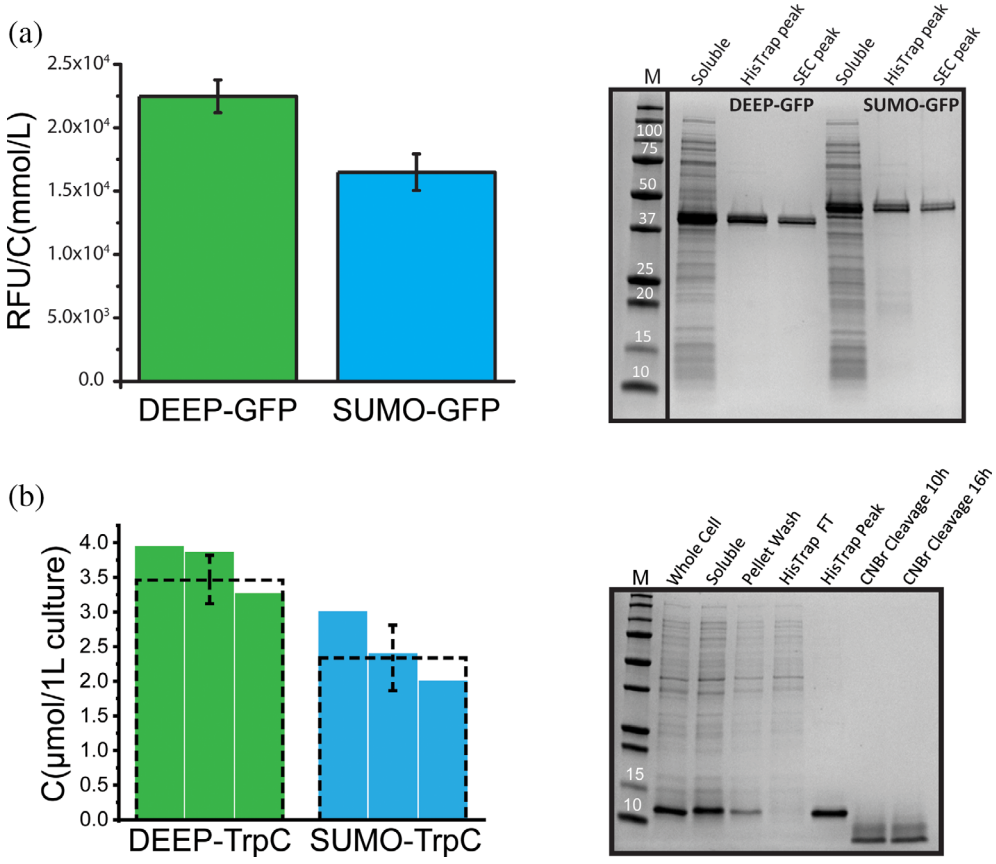

FIGURE 2 Quantified expression and purification of fusions to soluble proteins, GFP, and TrpCage: (a) Fluorescence of DEEP1-GFP and SUMO-GFP normalized to total protein concentration. On the right, SDS-PAGE comparing DEEP1 and SUMO fusion proteins at different stages of purification. (b) Comparison of DEEP1-TrpCage and SUMO-TrpCage expression and purification yields (in triplicate). On the right, SDS-PAGE of DEEP1-TrpC purification followed by CNBr cleavage. Protein concentrations were determined from the area under the peaks for samples run on reverse phase high pressure liquid chromatography (RP-HPLC) immediately after Ni-IMAC. Protein mass was confirmed using LCMS-ESI-TOF

expressing an insoluble peptide in vivo by relying on material from solid-phase synthesis. However, synthetic 42-mers are expensive. Moreover, a recent study showed that chemically synthesized Aβ42 does not have the same aggregation propensity or neurotoxicity as the same sequence expressed recombinantly, and the authors suggested that recombinant Aβ42 is a more appropriate material for studies related to Alzheimer disease.[13]

To assess the efficacy of DEEP1 for expressing Aβ42, we compared fusions to DEEP1 to fusions to SUMO, which as noted above, is nearly the same length as DEEP1. As expected, expression of DEEP-Aβ42 produced inclusion bodies. In contrast, and somewhat surprisingly, SUMO-Aβ42 was spread over soluble and insoluble fractions. Moreover, the SUMO-Aβ42 in the soluble fraction was substantially degraded (Figure 3a, Figure S2). In contrast, sequestration of DEEP-Aβ42 into inclusion bodies protected the peptide from proteolytic degradation and led to substantially higher yields.

To purify the desired product, we washed the DEEP1-Aβ42 inclusion bodies, solubilized them in 8 M urea, and loaded this material onto a Ni-IMAC column. Under these denaturing conditions, the DEEP1 fusion bound to the resin and was readily eluted and purified (under denaturing conditions) by the addition of imidazole (Figure S3).

To cleave Aβ42 from the fusion, we considered two issues: (a) The sequence of Aβ42 contains methionine, so CNBr could not be used to liberate the intact peptide. (b) The limited solubility of DEEP-Aβ42 would hamper enzymatic leverage in standard buffers. We overcame these challenges by inserting a recognition site for a protease that maintains activity under conditions that also maintain the solubility of the fusion. Specifically, we inserted the recognition site for Factor Xa (Ile-Glu-Gly-Arg) at the fusion junction. Because Factor Xa cleaves C-terminal to the arginine residue, this would produce a clean product without a "scar" from the recognition sequence. We optimized buffers to enable Factor Xa cleavage under conditions that maintain the solubility of DEEP-Aβ42. This was achieved by using a non-denaturing buffer supplemented with sarkosyl (wt/vol = 0.5%). Under these conditions, cleavage was complete in 4 hr (Figure S4a) and the product was confirmed by MS (Figure S4b).

The results described above demonstrate the utility of DEEP1 fusions to produce POIs that are either long or short (238 vs. 42 residues), and soluble or insoluble. Next,

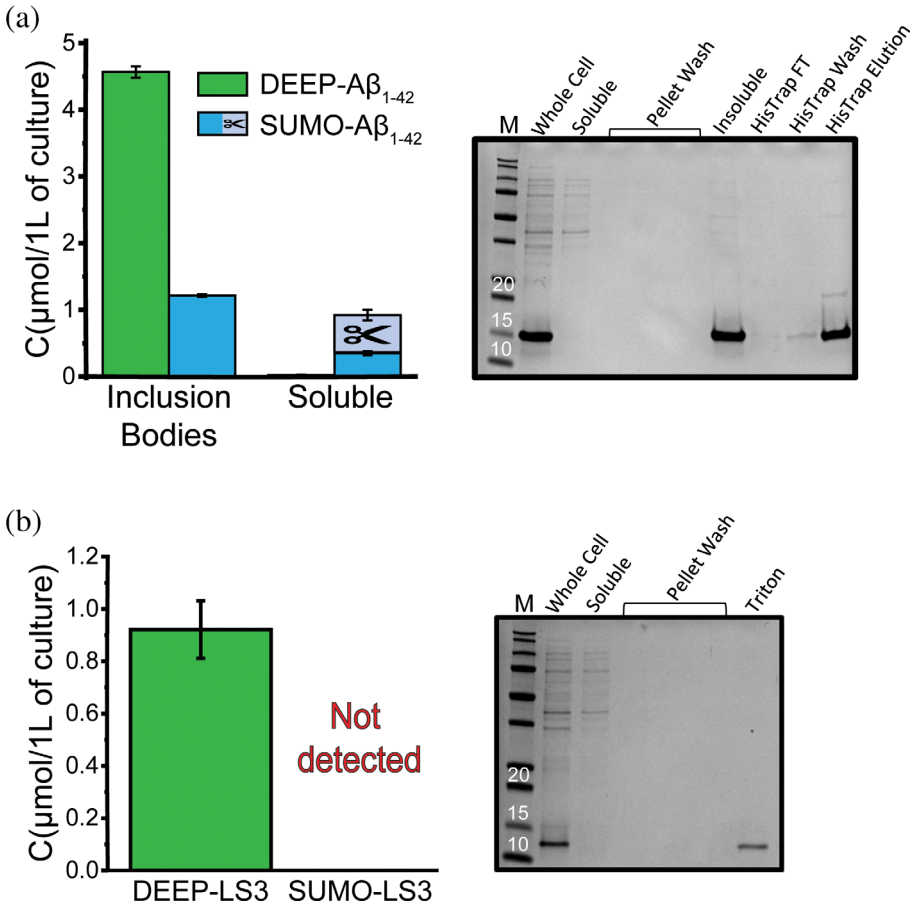

FIGURE 3 Expression and purification of fusions to insoluble peptides Aβ42 and LS3. (a) Quantity of DEEP1-Aβ42 and SUMO-Aβ42 in soluble and insoluble fractions. (Scissor indicates extensive proteolytic degradation of soluble SUMO-Aβ42). On the right, SDS-PAGE of the lysis, inclusion bodies wash, and Ni-IMAC purification of DEEP1-Aβ42 under denaturing conditions. (b) The concentration of DEEP1-LS3 in 2% Triton X-100 as determined by high pressure liquid chromatography (HPLC). SUMO-LS3 peaks could not be detected. On the right, SDS-PAGE of DEEP1-LS3 fractions following processing of inclusion bodies

we challenged DEEP1 to express a sequence targeted to membranes, and insoluble in native aqueous buffers. For this test, we chose the sequence $(LSLLLSL)_3$ designed by DeGrado and coworkers to model transmembrane pores.[14] $(LSLLLSL)_3$ is extremely hydrophobic and was shown to assemble as trimers/tetramers in model membranes.

LS3 fused to SUMO produced no observable expression (Figure 3b). In contrast, DEEP1-LS3 expressed at high levels. The DEEP1-LS3 material was found exclusively in the insoluble fraction. Initially, we assumed this insoluble material contained inclusion bodies; however, the addition of 2% Triton X-100 brought DEEP1-LS3 into the soluble phase (Figure S7) [More typical inclusion bodies are not solubilized by Triton X-100; indeed, this detergent is often used to wash membrane proteins away from inclusion bodies[15]]. The DEEP1-LS3 solubilized in 2% Triton X-100 was then purified on the Ni-IMAC column and digested with CNBr to produce the final LS3 product (Figures S5–S7).

2.3 | Design of alternative DEEP proteins

Because the sequences of novel proteins are not constrained by evolutionary history, they can readily be designed and optimized for specific technological applications. In particular, the DEEP proteins, which are based on combinatorial libraries of binary patterned sequences, can be manipulated to enhance the expression and/or purification of particular POIs. For example, a simple strategy to liberate a POI from its fusion partner is cleavage by trypsin. (Indeed, trypsin is used in the production of insulin, one of the most important biotechnological products.) Because trypsin cleaves after lysine and arginine, it would be advantageous to have a DEEP protein devoid of these basic residues. To create such a sequence, we mutated all eight lysines in DEEP1 to histidines. The resulting protein, called DEEP2, retains arginine at its C-terminus as a trypsin cleavage site to liberate the POI. For other applications, it may be advantageous to have a DEEP protein that is even more stable than the original DEEP1. To achieve this goal, we mutated Leu5 and Leu99 to cysteines, thereby enabling the formation of a disulfide bond linking the N- and C-termini of the four-helix bundle. This protein is called DEEP3. Finally, for some applications, it may be desirable to increase the affinity of DEEP fusions to Ni-IMAC resins. This would allow column washings with higher concentrations of imidazole, thereby enhancing the removal of contaminating endogenous proteins. To achieve this goal, four residues in the loop connecting α-helices 3 and 4 were mutated to histidine. The resulting protein, called DEEP4, contains eight histidines in the stretch of 11 residues spanning from the end of α-helix 3 into the beginning of α-helix 4.

The sequences of the redesigned DEEPs are shown in Figure 4a. The variants were purified to >95% purity using Ni-IMAC and SEC (Figure S8). Biophysical characterization of the purified proteins demonstrates that the redesigned DEEPs retain the expected α-helical structure, and all are stable (Figure 4b). Moreover, as expected from the design, the disulfide-bonded DEEP3 is slightly more stable than the DEEP1 parental protein (Figure 4b). The sequence of DEEP2, which is devoid of lysine residues is indeed resistant to trypsin digestion, as shown in Figure 4c. Finally, the histidine enriched sequence of DEEP4 remains bound to Ni-IMAC resins at significantly higher concentrations of imidazole; and this is true irrespective of whether the purification is done in native buffer or in 8 M urea (Figure 4d,e). These results confirm the flexibility of the DEEP platform for the design of novel fusions partners customized for particular applications.

3 | DISCUSSION

In principle, an optimal fusion tag should possess the following properties:

1. It should be relatively small. Longer sequences would squander cellular resources to produce a fusion tag that is ultimately discarded.
2. At the same time, the fusion tag should be long enough to fold into a stable globular structure that resists degradation by cellular proteases. Considerations (1) and (2) suggest a length between 60 and 150 resides.
3. The mRNA and protein sequence should express at high levels in the desired microbial host.
4. The fusion protein should fold into a structure that is stable under a range of conditions. This allows manipulation of growth environments in vivo (e.g., temperature), and purification conditions in vitro (e.g., temperature, pH, salt, detergent) to suit particular POIs.
5. Ideally, the fusion protein should also serve as an affinity tag for purification.
6. Optimally, the affinity properties of the fusion protein should be present in both its native folded state and under denaturing conditions (e.g. 8 M urea.)

Protein S-824, from a second-generation library of binary patterned sequences, possesses all these properties. S-824, which we renamed DEEP1, is a relatively small protein containing 102 residues. It expresses well

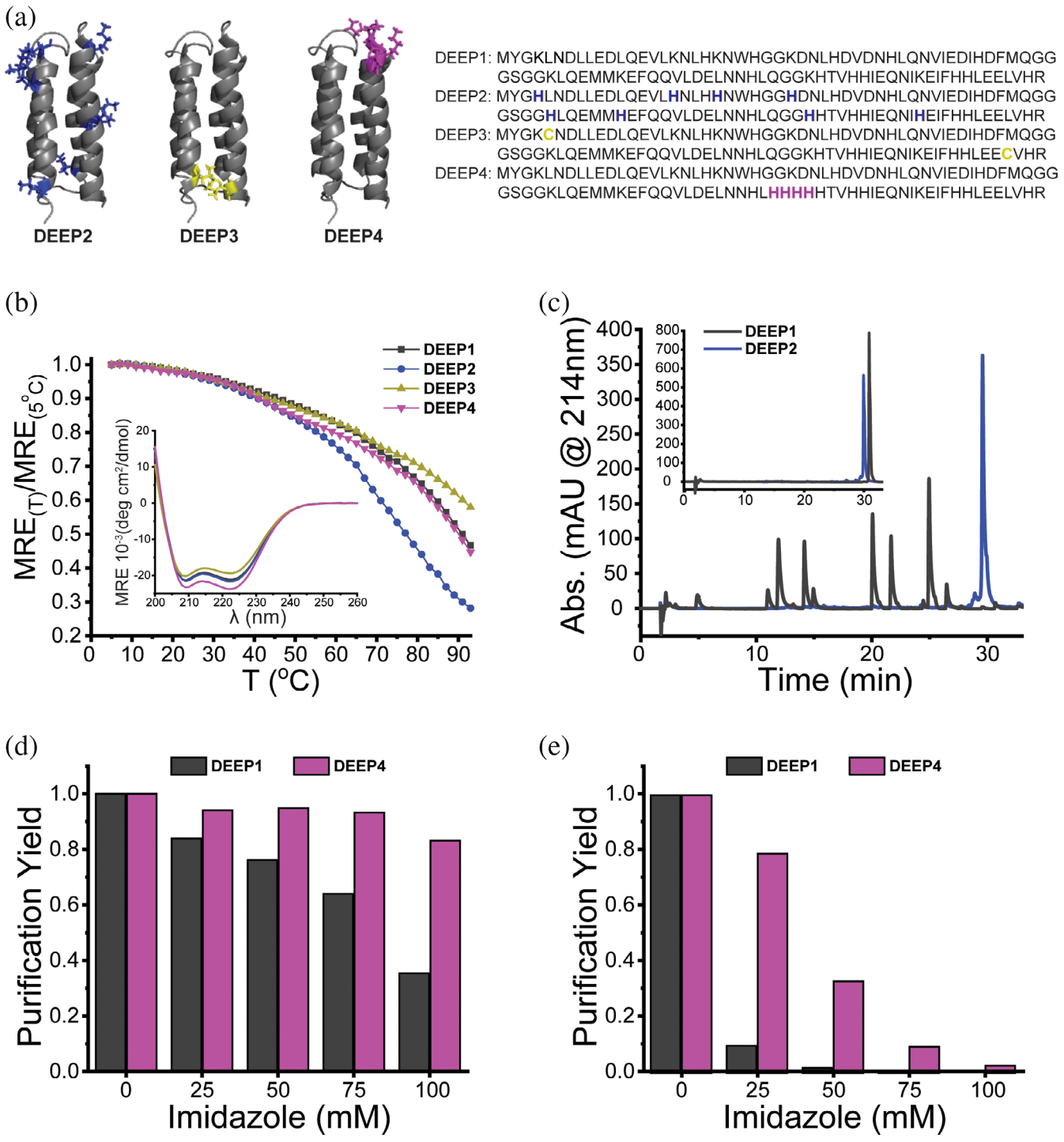

FIGURE 4 Amino acid sequences and characterization of modified DEEP proteins. In DEEP2, eight lysines were mutated to histidines. DEEP3 contains a disulfide bond connecting the first and fourth α-helices. DEEP4 contains additional histidine residues in an inter-helical turn. (a) Sequences and structural locations of the mutated amino acids. (b) Thermal denaturation and CD (inset) of DEEP1, 2, 3, and 4. (c) Reverse phase High Pressure Liquid Chromatography comparing the trypsin digestion of DEEP1 and DEEP2 (inset shows traces of the same samples before the addition of trypsin). (d, e) Purification yield of DEEP1 and DEEP4 as a function of imidazole concentration in the pre-elution wash step. Figure (d) shows purification in native buffer, and (e) shows purification in a buffer containing 8 M urea. Purification yield was calculated assuming 100% recovery of the DEEP protein in elutions at 500 mM imidazole following a wash step in 0 mM imidazole (for details see Figure S9)

under a range of conditions. It folds into a very stable four-helix bundle, with a denaturation temperature close to the boiling point of water. Moreover, because the sequence of DEEP1 contains an abundance of histidine residues, the protein binds to Ni-IMAC resins in native buffers or in 8 M urea. Finally, because the N- and C-

termini of S-824 are close together in its three-dimensional structure, POIs can be fused to both ends to facilitate co-assembly of two different peptides (Zarzhitsky et al., in preparation.)

While these properties are important, adherence to these properties per se cannot predict that a novel fusion partner will actually facilitate the expression and purification of a POI. Therefore, we chose four POIs spanning a range of sizes and solubilities as test cases for DEEP1. From the smallest to the largest (TrpCage is 20 residues and GFP is 238) the tested POIs differ in size by an order of magnitude. [We have not tested the efficacy of DEEP for very large multidomain proteins, such as antibodies.] In some cases the DEEP1 fusions expressed as soluble material, while in other cases they expressed as inclusion bodies or other forms of insoluble material. Nonetheless, in all four cases, DEEP1 fusions performed as well as, or better than, fusions to SUMO, a natural fusion partner of similar size, which has been used for a variety of applications in academia and industry.

Our results with DEEP1 represent the first example of a sequence that did not evolve in nature being used to enhance the expression and purification of heterologous natural proteins. However, DEEP1 may not be unique: Because the binary code specifies only the polar/nonpolar patterning of a sequence, but not the exact sequence, the binary code strategy enables the design and construction of vast libraries of related—but different—de novo sequences. Since each member of such libraries has a different sequence, the resulting proteins will have different biological and biophysical properties. Expression levels, solubilities, stabilities, affinities to resins, and other properties can span a range of values. In the current work, we began to explore the potential of this diversity by constructing variant DEEP proteins that were devoid of lysine, rich in histidine, or able to form intramolecular disulfide bonds. All these variants produced stable α-helical proteins with desired chemical and biophysical properties. These results demonstrate that binary patterned libraries can provide a rich source of de novo proteins from which to choose optimal fusion partners for expressing and purifying a range of important and valuable proteins.

4 | MATERIALS AND METHODS

4.1 | Genes, chemicals, columns, and solvents

DNA was ordered from IDT as single gBlocks. Cloning primers and other chemicals were purchased from Sigma. High pressure liquid chromatography (HPLC) columns were Zorbax C-18 (Agilent) analytical or semi-prep (SP). HPLC solvents, water, and acetonitrile were supplemented with 0.1% trifluoroacetic acid.

4.2 | Plasmid construction and protein overexpression

Using standard techniques, DNA encoding fusion proteins were cloned into a pET30 vector carrying a kanamycin resistance gene. The plasmid was transformed into *E. coli*, DH5α, mini-prepped, and sequenced (Genewiz). Plasmids carrying the correct sequences were transformed into BL21(DE3) *E. coli* cells and plated on LB agar plates supplemented with 30 mg/L kanamycin. A single colony was picked and inoculated into 5 ml LB supplemented with 30 mg/L kanamycin. This starter culture was incubated overnight in a shaking incubator at 37°C. The following day, 1 L of LB containing 30 mg/L kanamycin was inoculated with 4 ml of the overnight culture. At $OD_{600} = 0.5$, the soluble constructs (S-824, GFP and Trp Cage) were supplemented with isopropyl β- d-1-thiogalactopyranoside (IPTG) to final concentration of 0.1 mM. The temperature was lowered to 18°C and the cells were allowed to grow for an additional 16 hr. For the insoluble constructs (Aβ42 and LS3), IPTG was added at $OD_{600} = 0.8$ to a final concentration of 0.5 mM, and cells were grown for an additional 3 hr at 37°C. All expression experiments were performed in at least triplicate. Cells were pelleted by centrifugation for 30 min at 5,000*g* and stored at −80°C until further use.

4.3 | Protein purification

Cell pellets were resuspended in buffer A (50 mM Tris, 300 mM NaCl pH = 8) and lysis was initiated using either 15 sonication pulses (Branson 550) at 10 s On and 50 s Off, or five passages through an Emulsiflex C3 homogenizer operating at 15,000 psi. Lysed cells were centrifuged for 30 min at 15,000*g*.

S-824, *GFP*, and *Trp Cage constructs*: The soluble constructs were purified on FPLC (Akta Pure, GE) equipped with a 5 ml Ni-IMAC column (HisTrap HP, GE). Clarified lysates were loaded on a preequilibrated column in buffer A. Bound protein was washed with 10 column volumes (CV) of buffer A and 10 CV of 5% buffer B (50 mM Tris, 300 mM NaCl, 500 mM Imidazole at pH = 8). Proteins were eluted using 75% B. S-824 was further purified on a size-exclusion column (HiLoad 26/600, Superdex 75 pg, GE) using buffer A, and lyophilized after HPLC SP C-18 purification. Protein

concentration was determined using an adaptation of the Beer–Lambert equation

$$n[\text{mol}] = \frac{\text{AUC}\,[sec] \times F\left[\frac{\text{L}}{\text{sec}}\right]}{\varepsilon\left[\frac{\text{L}}{\text{mol} \times \text{cm}}\right] \times d\,[\text{cm}]}$$

where n is the moles of protein in the peak, F is the flow rate, ε is the molar extinction coefficient, and d is the path length.

Aβ42 and *LS3 constructs*: Soluble fractions of the Aβ42 and LS3 constructs were stored at −80°C until further use. The insoluble fractions of Aβ42 were washed sequentially with 1% Triton X-100 in buffer A and twice with water by manual resuspension of inclusion bodies and centrifugation for 15 min at 10,000*g*. The washed pellets for Aβ42 constructs were then resuspended in 6 M guanidine hydrochloride (GdnHCl) in buffer A and incubated overnight at 4°C. The following day, the resuspended inclusion bodies were centrifuged for 30 min at 30,000*g*. Supernatants and soluble fractions of Aβ42, acidified with 5% TFA in MeCN, were analyzed via Reverse Phase High Pressure Liquid Chromatography (RP-HPLC) equipped with analytical C-18 column, equilibrated with 97% water and 3% acetonitrile. Proteins were eluted using a 30 min increasing linear gradient to 100% acetonitrile.

The insoluble fractions for LS3 constructs were washed thrice with buffer A by resuspension and centrifugation each time for 15 min at 10,000*g*. These washed pellets were then resuspended in 2% Triton in buffer A and sonicated with six sonication pulses (Branson 550) at 5 s On and 20 s Off. The homogenized solution was incubated at room temperature for 2 hr and then centrifuged for 30 min at 30,000*g*. The supernatants were subjected to a buffer exchange to 8 M urea in buffer A using PD-10 columns. Both the supernatants in urea and the soluble fractions of LS3 constructs, acidified with 5% TFA in MeCN, were analyzed by RP-HPLC with an analytical C-18 column, equilibrated in 97% water and 3% acetonitrile. Proteins were eluted using a linear gradient of acetonitrile immediately followed by a gradient of isopropyl alcohol.

4.4 | GFP fluorescence

Fluorescence was measured on a Varioskan plate reader (Thermo) with excitation set to 490 nm and emission collected as a spectrum from 500 to 600 nm. Protein concentration was determined using absorbance at 280 nm.

4.5 | Circular dichroism

CD spectra of purified S-824 were collected on a Chirascan CD spectrometer (Applied Photophysics) using a 1 mm path length quartz cuvette. Lyophilized powder of S-824 was dissolved in either 20 mM Tris at pH = 8/9 or 20 mM CAPS at pH = 10, to a final protein concentration of ∼40 μM. Measurements were performed from 260 to 200 nm, averaging five scans for every sample. For thermal denaturation, ellipticity at 222 nm was monitored across a temperature range of 5–94°C with a temperature increase of 1°C/min. For mean residue ellipticity calculation protein concentration was estimated using absorbance at 280 nm.

4.6 | Cleavage of peptides

Trp Cage: CNBr cleavage was performed following a previously reported method.[12] Briefly, the peak eluted from the Ni-IMAC column was supplemented with 6 M guanidine-HCl, the pH was lowered using concentrated HCl, and CNBr dissolved in MeCN was added. The cleavage reaction was dried using a SpeedVac (Savant, SC110). Dried samples were reconstituted in water and further purified on HPLC (Agilent 1100) equipped with semi-preparative Zorbax 300SB-C18, 5 μm, 9.4 × 250 mm column using the following protocol: 20 min at 20% B followed by a 20 min gradient of 20–34% B at a flow rate of 3 ml/min (A-Water and B-MeCN, both supplemented with 0.1% trifluoroacetic acid). The identity of the collected peak was verified using LC/MS (Agilent 6220 Accurate-Mass Time-of-Flight).

Aβ42: Washed IBs of DEEP-Aβ42 were dissolved in 20 mM Tris, 120 mM NaCl and 0.5% (wt/vol) Sarkosyl pH = 8. The solution was briefly sonicated (three pulses, 10 s on and 10 s off, 50% amplitude) and incubated at 65°C for 10 min. Samples were spun down at 22,000*g* for 30 min. To perform a cleavage test, 100 μl of protein solution was diluted with 400 μl of 20 mM Tris, 120 mM NaCl and 0.5% (wt/vol) Sarkosyl and supplemented with 25 μl of 1 mg/ml Factor Xa protease (NEB). The cleavage reaction was further purified on a semi-preparative C-18 column (Vydac 218TP510) using the following protocol: 20 min at 20% B followed by 80 min gradient 20–60% B at a flow rate of 2 ml/min (A-Water and B-MeCN, both supplemented with 0.1% trifluoroacetic acid). The identity of the collected peak was verified using LC/MS (Agilent 6220 Accurate-Mass Time-of-Flight).

LS3: Cleavage of DEEP-LS3 was performed following a previously reported method (see Analytical Biochemistry 407 (2010) 144–146). Aliquots of the cleavage solution

were loaded on a semi-preparative C-18 column (Vydac 218TP510) and purified with the following method: 10 min at 10% B followed by 80 min gradient 50–100% B at a flow rate of 2 ml/min (A-80% water, 20% IPA and B-70% IPA, 20% MeCN and 10% water, both supplemented with 0.1% trifluoroacetic acid). The identity of the collected peak was verified using LC/MS (Agilent 6220 Accurate-Mass Time-of-Flight).

AUTHOR CONTRIBUTIONS

Shlomo Zarzhitsky: Conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing-original draft; writing-review and editing. **Alex Jiang:** Formal analysis; investigation; methodology. **Elizabeth Stanley:** Formal analysis; investigation; methodology. **Michael Hecht:** Supervision; writing-original draft; writing-review and editing.

ORCID

Shlomo Zarzhitsky https://orcid.org/0000-0003-2635-5224
Michael H. Hecht https://orcid.org/0000-0002-5538-9813

REFERENCES

1. Ahmad I, Nawaz N, Darwesh NM, et al. Overcoming challenges for amplified expression of recombinant proteins using *Escherichia coli*. Protein Express Purif. 2018;144:12–18.
2. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: Advances and challenges. Front Microbiol. 2014;5:172.
3. Wingfield PT. Overview of the purification of recombinant proteins. Curr Protoc Prot Sci. 2015;80:6.1.1–6.1.35.
4. Bernier SC, Cantin L, Salesse C. Systematic analysis of the expression, solubility and purification of a passenger protein in fusion with different tags. Protein Express Purif. 2018;152:92–106.
5. Kamtekar S, Schiffer J, Xiong H, Babik J, Hecht M. Protein design by binary patterning of polar and nonpolar amino-acids. Science. 1993;262:1680–1685.
6. Wei Y, Kim S, Fela D, Baum J, Hecht MH. Solution structure of a de novo protein from a designed combinatorial library. Proc Natl Acad Sci U S A. 2003;100:13270–13273.
7. Go A, Kim S, Baum J, Hecht MH. Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. Protein Sci. 2008;17:821–832.
8. Arai R, Kobayashi N, Kimura A, et al. Domain-swapped dimeric structure of a stable and functional de novo four-helix bundle protein, WA20. J Phys Chem B. 2012;116:6789–6797.
9. Wei Y, Liu T, Sazinsky S, Moffet D, Pelczer I, Hecht M. Stably folded de novo proteins from a designed combinatorial library. Protein Sci. 2003;12:92–102.
10. Waldo GS, Standish BM, Berendzen J, Terwilliger TC. Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol. 1999;17:691–695.
11. Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. Nat Struct Biol. 2002;9:425–430.
12. Andreev YA, Kozlov SA, Vassilevski AA, Grishin EV. Cyanogen bromide cleavage of proteins in salt and buffer solutions. Anal Biochem. 2010;407:144–146.
13. Finder VH, Vodopivec I, Nitsch RM, Glockshuber R. The recombinant amyloid-β peptide Aβ1–42 aggregates faster and is more neurotoxic than synthetic Aβ1–42. J Mol Biol. 2010;396:9–18.
14. Lear JD, Wasserman ZR, DeGrado WF. Synthetic amphiphilic peptide models for protein ion channels. Science. 1988;240:1177–1181.
15. Palmer I, Wingfield PT. Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. Curr Protoc Prot Sci. 2004. Chapter 6, Unit 6.3.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zarzhitsky S, Jiang A, E. Stanley E, H. Hecht M. Harnessing synthetic biology to enhance heterologous protein expression. *Protein Science*. 2020;29:1698–1706. https://doi.org/10.1002/pro.3907