Custom selenoprotein production enabled by laboratory evolution of recoded bacterial strains

Ross Thyer1, Raghav Shroff1, Dustin R Klein2, Simon d’Oelsnitz1, Victoria C Cotham2, Michelle Byrom1, Jennifer S Brodbelt2 & Andrew D Ellington1

Incorporation of the rare amino acid selenocysteine to form diselenide bonds can improve stability and function of synthetic peptide therapeutics. However, application of this approach to recombinant proteins has been hampered by heterogeneous incorporation, low selenoprotein yields, and poor fitness of bacterial producer strains. We report the evolution of recoded *Escherichia coli* strains with improved fitness that are superior hosts for recombinant selenoprotein production. We apply an engineered β-lactamase containing an essential diselenide bond to enforce selenocysteine dependence during continuous evolution of recoded *E. coli* strains. Evolved strains maintain an expanded genetic code indefinitely. We engineer a fluorescent reporter to quantify selenocysteine incorporation *in vivo* and show complete decoding of UAG codons as selenocysteine. Replacement of native, labile disulfide bonds in antibody fragments with diselenide bonds vastly improves resistance to reducing conditions. Highly seleno-competent bacterial strains enable industrial-scale selenoprotein expression and unique diselenide architecture, advancing our ability to customize the selenoproteome.

Long exploited by nature for its unique biophysical properties, such as increased nucleophilicity and low pKₐ, the rare twenty-first amino acid selenocysteine is a promising candidate for incorporation into engineered proteins by synthetic biologists. Unfortunately, the complexity of the selenocysteine incorporation pathway, which requires multiple *cis*- and *trans*-acting protein and RNA factors to reassign the opal stop codon (reviewed in detail by Yoshizawa and Böck1), means that recombinant selenoprotein expression is difficult to engineer in common microbial hosts. Furthermore, the selenocysteine-charged tRNA is an inefficient substrate for the translational machinery2, and is outcompeted by release factor 2 for access to the UGA codon, resulting in premature termination of protein synthesis. Despite attempts to bypass these constraints on selenocysteine incorporation in bacteria3,4, inefficient translation, toxicity, and a lack of genetic tools mean that chemical synthesis remains the most reliable route to engineer new selenoproteins.

Previously, we evolved an *E. coli* tRNA^{Sec} variant5 that enabled efficient, site-specific selenocysteine incorporation at amber stop codons and overcame the sequence constraints imposed by the endogenous bacterial selenocysteine incorporation machinery. Using this system we produced recombinant proteins containing selenyl-sulfhydryl and diselenide bonds5. During our efforts to improve selenoprotein yield, we encountered two separate issues; we observed substantial toxicity from selenocysteine biosynthesis, and discovered that only a subset (<10%) of bacterial cells transformed with plasmids encoding a functional biosynthesis pathway acquired the ability to incorporate selenocysteine (Supplementary Fig. 1). We hypothesized that this might be because the recoded *E. coli* host (RTΔA5, which is derived from C321.ΔA6) had not adapted to the effects of genomic recoding4 on regulation and synthesis of the proteome, or because the new redox-active amino acid had attendant fitness burdens.

Whole genome evolution has been used to optimize bacterial fitness, most notably in the laboratory evolution of *E. coli* over 60,000 generations by the Lenksi group7, which resulted in complex genetic adaptations that supported the ability to metabolize a carbon source it couldn’t previously use8. To facilitate the adaptation of recoded *E. coli* RTΔA using genome evolution, we coupled cell survival to selenoprotein expression, by creating conditional dependence on an expanded genetic code. This enforces the identity of the UAG stop codon as selenocysteine and ensures cells always exceed a minimum threshold of selenocysteine incorporation required for survival. Whole genome evolution of selenocysteine-dependent lineages of recoded *E. coli* in parallel led to the isolation of a superior host that enabled production of improved yields of selenoproteins, including those with complex diselenide architectures.

**RESULTS**

**Establishing selenocysteine dependence in *E. coli***

We produced a selenocysteine-dependent host strain by transforming *E. coli* RTΔA cells with two plasmids encoding a synthetic selenocysteine biosynthesis pathway consisting of tRNA^{Sec3,5,7}X, selenocysteine synthase (SelA), selenophosphate synthase (SelD), and O-phosphoseryl-tRNA^{Sec} kinase (PstK)5. Strains with different degrees of selenocysteine dependence were established by integrating one of three different variants of the bla<sub>NMC-A</sub> gene (disulfide-dependent β-lactamase from *Enterobacter cloacae*) into the chromosome5.

---

1Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas, USA. 2Department of Chemistry, University of Texas at Austin, Austin, Texas, USA. Correspondence should be addressed to R.T. (ross.thyer@utexas.edu) or A.D.E. (ellingtonlab@gmail.com).

Received 5 October 2017; accepted 23 March 2018; published online 4 June 2018; doi:10.1038/nbt.4154
The three variants were a wild-type β-lactamase containing the native disulfide bond between C69 and C238 (CC), which did not enforce dependence on selenocysteine, a U69-C238 variant (UC) containing an essential selenyl-sulphydryl bond, and a U69-U238 variant (UU) containing an essential disulphide bond (Fig. 1a). Selenocysteine-dependent β-lactamase activity of the UC and UU variants was confirmed by sensitivity to carbenicillin when grown in selenium-free defined media, which could be rescued by the addition of 1 μM sodium selenite (Na2SeO3). A wild-type blaNMCA gene was also integrated into RTΔA cells containing empty plasmids (Δ strain) to provide a control to identify mutations that accrue independently of selenocysteine incorporation.

We selected for selenocysteine tolerance and improved fitness by serial passage of the four parental strains (CC, UC, UU, and Δ) for 2,500 cell doublings. All experiments were carried out in triplicate, under two different growth conditions that we hypothesized would elicit different adaptive responses. Growth of parental strains was either in an increasing concentration of a β-lactam antibiotic (β populations) or increasing temperature (T populations) (Fig. 1b). These conditions were chosen because: (1) they are mildly mutagenic, through induction of DNA polymerase IV, thereby increasing the chance of an adaptive response; (2) both conditions have previously been used, thus enabling us to compare our results with those already published; (3) both conditions impose stress on selenocysteine biosynthesis and incorporation, but in different ways.

Increasing β-lactam antibiotic concentration exerts selective pressure for increased production of NMC-A β-lactamase, which in turn requires an increase in selenocysteine incorporation. Elevated temperature reduces tRNAsecc stability, which in turn decreases selenocysteine incorporation12, and might result in evolution of more stable or more active selenocysteine incorporation machinery.

**Evolved E. coli populations bypass metabolic defects**

Following serial passaging, we observed a substantial increase in fitness of all evolved populations. Growth rates and culture densities increased in all β populations, with particularly large improvements observed for β_UC and β_UU populations, which reached two- to threefold higher culture density and spent 20–40% less time in lag phase (Fig. 1c and Supplementary Fig. 2). Similarly, growth of all evolved β populations in a defined selenium-free medium (MOPS EZ), which barely supported growth of the parental cells, was greatly improved (Fig. 1d and Supplementary Fig. 3). This indicates that populations were able to overcome serious defects in core metabolic processes despite being cultured exclusively in rich media. All UC and UU populations retained sensitivity to carbenicillin in MOPS EZ, which could be reversed by supplementation with Na2SeO3, showing that selenocysteine dependence was maintained throughout the 2,500 generations. No populations acquired unconditional dependence on selenocysteine, which would necessitate Na2SeO3 supplementation in the absence of carbenicillin.

---

**Figure 1** Evolution of selenocysteine-dependent *E. coli* strains improves fitness. (a) A recoded *E. coli* strain deficient in selenocysteine incorporation (RTΔA) was made conditionally dependent on selenocysteine by integrating NMC-A β-lactamase variants containing either an essential selenyl-sulphydryl (UC) or disulphide (UU) bond and supplying the biosynthesis and incorporation machinery *in trans*. Control strains contained a wild-type β-lactamase containing a disulfide bond, lacking either selenocysteine dependence (CC) or the entire incorporation trait (Δ). (b) The four populations were serially passaged, in triplicate, 200 times to saturation (∼2,500 generations) in two different environmental conditions: increasing β-lactam stress (β populations) or increasing temperature (T populations). Note, at 43.5 °C (∗) passaging technique was adjusted for the T populations. Curves are plotted with a line showing the mean and shading representing ± s.e.m., where n = 3 independent biological replicates. (c) Growth rate, culture density, and β-lactam resistance of an evolved population containing an essential disulphide bond (β_UU3) were substantially increased compared to the parent population (β_UU) at different concentrations of carbenicillin. (d) Metabolic defects that impaired growth in defined culture media (MOPS EZ) were overcome and strains remained dependent on selenocysteine incorporation for growth in the presence of β-lactam antibiotics. Carbenicillin and Na2SeO3 were added at concentrations of 100 µg/mL and 1 µM, respectively.
Supplementary Table 1

| 1   | 2   | 3   |
|-----|-----|-----|
| A   | B   | C   |

This table represents the values for each experiment.

Figure 2 Genes mutated during continuous evolution. (a) Mutations were identified by whole genome sequencing of all evolved populations after 2,500 generations. Genes that contained SNPs with >50% frequency in at least 4 of the 12 independent subpopulations are reported using a circular representation. From innermost to outermost, the rings represent the genome location with a scale in megabase pairs (Mb), the ∆ populations, the UC populations, and the UU populations. The yellow, orange, and red bars represent one, two, or three mutant alleles in a given set of triplicate subpopulations, respectively. The details of all detected mutations are reported in Supplementary Tables 2 and 3. (b–e) Growth curves of ftsA G124E, hemA P196L, yeeJ S1467P, and prfB T246A mutants compared to the parental RTΔA strain. Curves are plotted with a line showing the mean and shading representing ± s.e.m., where n = 3 independent biological replicates. Note, ftsA G124E and yeeJ S1467P represent reversions to wild-type MG1655 sequence, which were not directly observed and are discussed in Supplementary Note 1.

Populations evolved under thermal stress grew poorly at 45 °C during serial passaging, and most populations had inferior, although highly variable, growth rates and culture density when compared with parental cells grown at 37 °C in rich LB media (Supplementary Fig. 4). This is consistent with previous observations of E. coli MG1655 when it was evolved at elevated temperatures (45–48 °C)13. While growth was impaired in rich media, all evolved T populations outgrew parental strains in defined MOPS EZ medium at 37 °C (Supplementary Fig. 5), as was observed for the β populations. All T_UC and T_UU populations retained selenocysteine dependence for carbenicillin resistance.

Genetic analysis of evolved strains

In order to analyze the genetic basis of evolved growth phenotypes, we carried out whole genome sequencing on evolved populations and their parental strains, including RTΔA. Single clonal isolates were also sequenced from each of the β_UC and β_UU populations and were found to be highly representative of the evolved population sequences, containing the same fixed single-nucleotide polymorphisms (SNPs) and no genomic rearrangements. Whole genome sequencing indicated that the selenocysteine biosynthesis and incorporation machinery was maintained throughout evolution, that there was no loss of TAG codons in the blaNMC-A gene and no contamination between the populations. We did not find any genomic suppressor tRNAs via whole genome sequencing. We detected several new in-frame TAG codons in all sequenced genomes, including β_Δ and T_Δ populations, which contained no suppressor tRNAs (Supplementary Table 1). Mass spectrometry of a model protein (GFP) expressed in control strains confirmed that UAG was decoded as glutamine (Supplementary Fig. 6), which is expected because tRNA^{Glu}_{GAG} is the most efficient near-cognate UAG suppressor14.

Genome sequencing identified clusters of SNPs in genes conferring antibiotic resistance (blaNMC-A, ftsI, ompR, and selA), genes mediating oxidative stress or selenite resistance (oxyR and cysK), genes mutated during construction of C321.ΔA (ftsA, hemA, pta, and yeeJ), genes involved in plasmid replication (pcnB), and the prfB gene, which codes for release factor 2 (Fig. 2a). Genes enriched with SNPs were defined as having acquired SNPs with >50% frequency in four or more independent populations evolved under the same conditions. Notably, no mutations were observed in the selA, ptk, tRNA^{Sec}_{GUG}, or selD genes which form the synthetic selenocysteine biosynthesis pathway.
The impact of these mutations on fitness was evaluated by introducing a subset of the most enriched SNPs in genes not related to multidrug resistance back into the parental RTΔA strain, using multiplex automated genome engineering (MAGE)\textsuperscript{15}. SNPs that we engineered into the parental strain included oxyR A233T, cysK T69I, T73A and H153Y, prfB T246A, N276D, and K282R, and also reversion of the mutations in ftsA, hemA, pta, and yew\textsuperscript{1} to the wild-type sequence present in \textit{E. coli} strain MG1655, the parental strain of C321.AA. The growth of MAGE-engineered strains was compared with that of RTΔA, and four mutations were identified that simultaneously decreased time spent in lag phase and increased the final culture density (Fig. 2b–e). These mutations included three reversions of SNPs that were acquired during construction of the C321.AA strain, and the \textit{prfB} T246A mutation, which repairs a defect in release factor 2 that is only present in \textit{E. coli} K12 strains\textsuperscript{16,17}. Although most of the SNPs we identified either compensated for metabolic defects in the parental recoded \textit{E. coli} or broadly improved cell growth, cysK and \textit{pcnB} directly affected selenocysteine tolerance and incorporation. An extended discussion of the SNPs observed in the evolved populations and our efforts to characterize them can be found in Supplementary Note 1, Supplementary Figures 7–9, and Supplementary Tables 2 and 3.

**Seleno-smURFP: a selenocysteine-specific genetic reporter**

Although serial passage improved the fitness of our various populations, it was equally important to assess how well evolved strains produced selenoproteins. While quantitative fluorescent reporter proteins, such as GFP, are used to measure UAG suppression, they cannot discriminate between selenocysteine and competing serine incorporation from Ser-tRNA\textsubscript{sec}, the immediate precursor in selenocysteine biosynthesis. Traditional genetic reporters for selenocysteine incorporation are available, such as, formate dehydrogenase, but are only qualitative, due to their narrow dynamic range, and, in addition, they require anaerobic growth conditions\textsuperscript{18}.

Therefore, we needed to develop a quantitative selenocysteine-dependent fluorescent reporter. In order to differentiate selenocysteine from serine, we co-opted the process of chromophore ligation in phycobiliproteins, which occurs by formation of a thioether bond between a nucleophilic cysteine residue in the protein and a proximal vinyl group in the tetrapyrrole chromophore\textsuperscript{19} (Fig. 3a). This reaction is limited by the nucleophilicity of the cysteine thiol at physiological pH and is not supported by a serine hydroxyl group\textsuperscript{20}, so we hypothesized that a selenocysteine residue would support formation of a selenoether cross-link. To this end we replaced the essential cysteine residue in \textit{E. coli} from serine, an engineered autocatalytic phycobiliprotein, with seleno-

![Graphs represent the average fluorescence of four independent clones normalized to cell density ± s.e.m.](image)

**Figure 3** Selenocysteine incorporation in evolved strains. (a) Homology model of seleno-smURFP (US2) bound to biliverdin (modeled from PDB: 4po5) with the selenoether bond shown. (b) Schematic representation of discrimination between selenocysteine and serine by smURFP-X52. Both Ser-tRNA\textsubscript{sec} and Sec-tRNA\textsubscript{sec} are substrates for EF-Tu but only incorporation of selenocysteine at position 52 enables covalent tethering of the biliverdin chromophore that results in fluorescence. Non-covalent association of the chromophore still occurs when serine is incorporated. (c) Plate assay of smURFP mutants with varying concentrations of Na\textsubscript{2}SeO\textsubscript{3} confirms fluorescence is dependent on selenocysteine. Plate is representative of six biologically independent grids. (d) Assay in liquid media shows greater than 20-fold difference between seleno-smURFP and smURFP-S52. (e) Plate assays show biologically independent transformants (\(n = 100\)) and are representative of two experiments. (f) Populations evolved with β-lactam stress screened for seleno-smURFP expression. Incorporation of Sec was measured with our fluorescent reporter, seleno-smURFP. Graphs represent the average fluorescence of four independent clones normalized to cell density ± s.e.m.
for selenocysteine, seleno-smURFP is a robust tool for optimizing selenocysteine biosynthesis and incorporation.

We used the seleno-smURFP reporter to measure whether selenocysteine dependence affected retention of the selenocysteine incorporation trait. Cell populations were transformed with a plasmid expressing seleno-smURFP, and 100 transformants were screened for red fluorescence (Fig. 3e and Supplementary Fig. 11). Transformants from populations dependent on selenocysteine incorporation (β_UC, βUU, T_UC, and T_UU) were almost universally fluorescent, confirming that the ability to incorporate selenocysteine had been retained, independent of the selective pressure encountered during evolution. In contrast, while all three β_CC populations retained the ability to incorporate selenocysteine, the T_CC populations had highly variable rates of selenocysteine incorporation; <10%, 75%, and >90% for T_CC 1–3, respectively (Supplementary Fig. 11). No transformants from populations β_A 1–3 and T_A 1–3, which lacked the selenocysteine incorporation trait, were fluorescent. Taken together these data indicate that the engineered dependence on selenocysteine was successful and that selenocysteine incorporation was maintained throughout serial passaging.

Evolved strains are superior hosts for selenoprotein synthesis

To measure the fold-improvement in selenoprotein expression, transformants from each of the parental strains and the evolved β-lactam populations were assayed for fluorescence in liquid media (Fig. 3f). Evolved β_CC and β_UC populations showed five- to sevenfold increases in selenoprotein expression compared to the parental cells. Evolved βUU populations showed less change from the parental cells, which had much higher fluorescence than either the β_CC or β_UC parental cells. In addition to the increased fluorescence (normalized to cell density), the evolved β_UC and βUU strains achieved several fold higher cell density (Fig. 1c), increasing overall selenoprotein yield in some populations up to 20-fold.

A single clone from the βUU3 population was isolated, and we expressed a series of model proteins containing pairs of selenocysteine residues to calculate the specific selenocysteine incorporation efficiency. ExDHFR U39-U85 and GFP U135-U177 were expressed at yields of 8 and 40 mg/L, respectively, and only intact masses corresponding to selenocysteine incorporation were detected by mass spectrometry (Supplementary Fig. 12). Ultraviolet photodissociation (UVPD) of ExDHFR U39-U85 and GFP U135-U177 confirmed complete formation of a diselenide bond between the selenocysteine residues.

To validate diselenide bond formation in clinically relevant protein families with complex bond architectures, we expressed a model antibody fragment (anti-MS2 scFv) containing two essential bonds, and human growth hormone (hGH, 9 mg/L), which contains two bonds, one of which is essential. Mass spectrometry confirmed complete selenocysteine incorporation and correct...
Diselenide bond formation reduces structural heterogeneity

Due to the lower redox potential of diselenide bonds (~381 mV compared to ~180 mV for disulfide bonds)\(^1\), diselenide bond formation is more favorable under reducing conditions than disulfide bond formation, but the consequences of bond replacement on protein structure are unknown. To investigate the effect of diselenide bonds on protein structure, we expressed the Trastuzumab-scFv (Herceptin-scFv), a highly engineered antibody fragment which retains the strictly conserved disulfide bonds in the variable domains, without requiring bond formation to fold\(^2\). This property allowed us to observe species with both oxidized and reduced cysteine thiols without compromising solubility. As a baseline, wild-type Herceptin-scFv was expressed under reducing conditions in the cytoplasm of \(E.\ coli\) strain BL21(DE3) (12 mg/L), a common host for recombinant protein expression. Using top-down mass spectrometry, we identified a series of intact protein ions normally distributed around the 29+ charge state (Fig. 4a and Supplementary Fig. 14) and calculated a monoisotopic mass of 27,172.05 Da, corresponding to a sample containing only one disulfide bond. UVPD of the 29+ charge state showed an even distribution of fragment ions throughout the protein sequence, confirming a lack of disulfide bonds (Fig. 4b).

To assess the effect of disulfide bond formation, we expressed wild-type Herceptin-scFv in T7 Shuffle Express (15 mg/L), an \(E.\ coli\) strain which maintains an oxidizing cytoplasm and constitutively expresses DsbC to facilitate bond formation.\(^26\) The intact protein ions observed during top-down mass spectrometry spanned a broad range of lower charge states, suggesting a mix of reduced and oxidized protein species\(^27\)–\(^29\) (Fig. 4c). We calculated a monoisotopic mass of 27,170.12 Da, corresponding to a species containing only a single disulfide bond. Considering the broad charge state distribution, it is possible that the calculated monoisotopic mass additionally reflects an averaged mass of both fully reduced and fully oxidized species. UVPD fragmentation of the 23+ charge state showed only the \(V_L\) domain contained a gap in the distribution of fragment ions, confirming the presence of only one disulfide bond (Fig. 4d and Supplementary Fig. 14). The identity of the 29+ charge state, potentially a fully reduced species, was unable to be confirmed by UVPD due to the low peak intensity.

In contrast, protein ions from seleno-Herceptin-scFv expressed in \(\beta\_\text{UU3-T7}\) cells (0.8 mg/L) were confined to a narrow distribution of lower charge states, consistent with a highly bonded species, and the monoisotopic mass matched a species containing two diselenide bonds (Fig. 4e). Due to reduced conformational flexibility in the gas phase, disulfide/diselenide-bonded proteins are expected to be observed in lower charge states than their disulfide/diselenide-reduced counterparts\(^30\). Comparison of the charge state distributions of Herceptin-scFv s expressed in \(E.\ coli\) strain BL21(DE3), T7 Shuffle Express, and seleno-Herceptin-scFv expressed in \(\beta\_\text{UU3-T7}\) cells further suggest protein populations that were fully reduced, a mixture of reduced and oxidized, and fully oxidized, respectively. UVPD fragment analysis of the 20+ charge state identified gaps in sequence coverage between the pairs of selenocysteine residues in both the \(V_L\) and \(V_H\) domains, confirming correct formation of both diselenide bonds (Fig. 4f and Supplementary Fig. 14).

Taken together these data confirm that diselenide bonds can reduce sample heterogeneity while ensuring formation of the correct covalent architecture, and suggest that the elevated redox potential of the cytoplasm in T7 Shuffle Express cells is insufficient to completely oxidize cysteine residues, resulting in incomplete disulfide bond formation.

Diselenide bonds confer resistance to reducing conditions

Although diselenide bonds do not provide complete immunity to reduction, the kinetics of bond reformation are faster than disulfide bonds, which has implications for protein stability under reducing conditions. Due to reduced conformational flexibility in the gas phase, disulfide/diselenide-bonded proteins are expected to be observed in lower charge states than their disulfide/diselenide-reduced counterparts. Comparison of the charge state distributions of Herceptin-scFv expressed in \(E.\ coli\) strain BL21(DE3), T7 Shuffle Express, and seleno-Herceptin-scFv expressed in \(\beta\_\text{UU3-T7}\) cells further suggest protein populations that were fully reduced, a mixture of reduced and oxidized, and fully oxidized, respectively. UVPD fragment analysis of the 20+ charge state identified gaps in sequence coverage between the pairs of selenocysteine residues in both the \(V_L\) and \(V_H\) domains, confirming correct formation of both diselenide bonds (Fig. 4f and Supplementary Fig. 14).
conditions. This is well established in small peptides where bond formation is the primary driver of structure, but not in proteins, which have independent secondary and tertiary structure. To determine if diselenide bonds could improve protein stability under reducing conditions, we expressed the anti-RCA (ricin A-chain) scFv, which contains two essential bonds, in both disulfide (T7 Shuffle Express, 1 mg/L) and diselenide (β_UU3-T7, 1 mg/L) configurations, as confirmed by mass spectrometry (Fig. 5 and Supplementary Fig. 15). Proteins were incubated with varying concentrations of dithiothreitol (DTT), which we hypothesized would reduce both disulfide and diselenide bonds, but prevent reformation of only the disulfide bonds. DTT treatment impaired binding of the disulfide-bonded scFv at all concentrations tested (1, 10, and 50 mM), whereas the diselenide-bonded scFv was only mildly affected by 50 mM DTT, the highest concentration tested (Fig. 5c,f). Given that the disulfide bonds in antibody fragments are buried in the core of the variable domains, further increased stabilization might be observed in proteins with more solvent-accessible bonds, such as EGF-family proteins.

**DISCUSSION**

We engineered an improved bacterial host for selenoprotein expression by endowing *E. coli* strains with conditional dependence for incorporation of selenocysteine using an amber suppressor tRNA and an engineered β-lactamase containing an essential diselenide bond. Dependence on selenocysteine incorporation was maintained for more than 2,500 generations, allowing cell populations to evolve tolerance to constitutive selenocysteine biosynthesis and acquire several mutations that improved cell fitness. This is the first demonstration, to our knowledge, of enforcing an expanded genetic code to facilitate host adaptation and improve recombinant protein production. We anticipate that selected mutations identified in our study could be integrated into recoded *E. coli* strains that have been optimized using different methods. With recent advances in orthogonal translation systems and protein design, this type of approach should prove generalizable to other amino acids in other proteins.

Previous attempts to improve genomically recoded *E. coli* strains have focused exclusively on growth rate but have found less substantial improvements in the suppression of UAG codons by non-canonical amino acids. Using our evolved strains, we have greatly expanded the number of selenocysteine residues that can be site-specifically incorporated into recombinant or endogenous proteins, enabling the production of selenoproteins containing multiple diselenide bonds, such as seleno-antibody fragments. These strains can be used in future efforts to engineer and customize the selenoproteome and, along with the development of new tools such as diselenide bond isomerases, may enable protein engineers to access complex nonsequential diselenide bond configurations.

We developed a fluorescent reporter, seleno-smURFP, that enabled us to confirm that our evolved *E. coli* populations were superior hosts for recombinant selenoprotein expression. Seleno-smURFP enables quantitative measurement of selenocysteine incorporation in vivo and will facilitate rapid optimization of the selenocysteine biosynthesis machinery. Using our evolved *E. coli* host (strain β_UU3) we expressed a diverse set of recombinant selenoproteins containing diselenide bonds with yields, selenocysteine incorporation efficiency, and diselenide bond complexity that are superior to previous attempts. Furthermore, we report improved formation of diselenide bonds compared with disulfide bonds in the bacterial cytoplasm, and confirm that diselenide bonds in recombinant proteins can confer resistance to reducing conditions.

Diselenide bonds have been used to improve the stability and biological half-life of therapeutic peptides, such as insulin, oxytocin, which are manufactured using solid-phase synthesis. The ability to efficiently introduce diselenide bonds into recombinant proteins extends this stabilizing motif to previously inaccessible classes of protein therapeutics, and provides a high-throughput approach to improve existing therapeutics.

In addition to enabling the incorporation of diselenide bonds as a protein engineering tool, a reliable host for recombinant selenoprotein expression will find broad utility among the wider protein biology community. Replacing catalytic cysteine residues in enzymes with selenocysteine has enabled advances in mechanistic enzymology, but progress has been hindered by inefficient protein expression in cysteine auxotrophs or by specialized protein ligation strategies. In addition, these approaches have inherent limitations, requiring either the removal of native cysteine residues or accepting indiscriminate selenocysteine incorporation, or maintaining the solubility of truncated enzyme fragments.

Similarly, efforts to characterize the human selenoproteome (comprising 25 proteins of which half are uncharacterized) have relied on auxotrophic selenocysteine incorporation, or replacement of catalytic selenocysteine residues with cysteine to overcome the difficulty of expression at the cost of producing proteins with low activity and unknown biological relevance. Recent attempts to produce two of these proteins have been successful using complete chemical synthesis of entire selenoproteins. Expression in a bacterial host will enable easier analysis of these proteins.

Furthermore, selenocysteine-dependent conjugation chemistries methods currently employed by peptide chemists, can now be expanded to recombinant proteins using our strains, making orthogonal drug conjugation easier by removing the need to eliminate other reactive surface residues.

We envision that the tools and host strains for highly efficient site-specific selenocysteine incorporation that we report here will serve as a platform for exploring the potential of the selenoproteome and seleno-stabilized therapeutics.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

Funding from the Welch Foundation (F-1654 to A.D.E. and F-1155 to J.S.B.), the NSSEFF (FA9550-10-1-0169 to A.D.E.), the NSF (CHE-1402753 to J.S.B.), the NIH (5K99CA207870-02 to R.T.) is acknowledged.

**AUTHOR CONTRIBUTIONS**

R.T. designed the experiments and performed the strain evolution, protein purification, and protein characterization. Bioinformatic analysis was done by R.S. MAGe was conducted by R.T. and R.S. The fluorescent reporter was developed by R.T. and S.d’O. Mass spectrometry was performed by D.R.K., V.C.C., and J.S.B. qPCR was performed by M.B. The manuscript was written by R.T. with support from A.D.E. R.T. and A.D.E. supervised all aspects of the study.

**COMPETING INTERESTS**

R.T. and A.D.E. have equity in GRO Biosciences, a company developing protein therapeutics, and receive royalties from licensing material described in this manuscript.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
1. Yoshizawa, S. & Böck, A. The many levels of control on bacterial selenoprotein synthesis. *Biochim. Biophys. Acta* 1790, 1404–1414 (2009).

2. Suppmann, S., Persson, B.C. & Böck, A. Dynamics and efficiency in vivo of UGA-directed selenocysteine insertion at the ribosome. *EMBO J.* 18, 2284–2293 (1999).

3. Arai, K. et al. Fine-tuning citrate synthase flux potentials and refines metabolic innovation in the Lenski evolution experiment. eLife 4, e09636 (2015).

4. Majiduddin, F.K. & Palzkill, T. Amino acid sequence requirements at residues 69 and 238 for the SME-1 beta-lactamase to confer resistance to beta-lactam antibiotics. *Antimicrob. Agents Chemother.* 47, 1062–1067 (2003).

5. Gutierrez, A. et al. Preparation of selenoinsulin as a long-lasting insulin analogue. *Angew. Chem. Int. Edn Engl.* 56, 5558–5588 (2017).

6. Baggett, N.E., Zhang, Y. & Gross, C.A. Global analysis of translation termination in *E. coli*. PLoS Genet. 13, e1006676 (2017).

7. Blount, Z.D., Barrick, J.E., Davidson, C.J. & Lenski, R.E. Genomic analysis of a metabolic innovation in the Lenski evolution experiment. *Science* 342, 357–360 (2013).

8. Quandt, E.M. Selenocysteine insertion at a predefined UAG codon in *Escherichia coli*. *J. Biol. Chem.* 285, 19029–19034 (2010).

9. Wang, H.H. et al. Multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* 34, 623–628 (2017).

10. Blount, Z.D., Barrick, J.E., Davidson, C.J. & Lenski, R.E. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489, 513–518 (2012).

11. Rudolph, B., Gebendorfer, K.M., Buchner, J. & Winter, J. Evolution of selenocysteine tRNA provides resilience at the expense of maximal function. *J. Biol. Chem.* 288, 13337–13344 (2013).

12. Ishii, T.M., Kottova, N., Tapsova, F. & Steinberg, S.V. The long D-stem of the selenocysteine tRNA provides resilience at the expense of maximal function. *J. Biol. Chem.* 288, 13337–13344 (2013).

13. Tack, D.S. et al. Addicting diverse bacteria to a noncanonical amino acid. *Nat. Chem.* 12, 138–140 (2016).

14. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898 (2009).

15. Wannier, T.M., Kuznetsov, G. & Green, M.T. Characterization of a far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. *Nat. Methods* 8, 1922–1926 (2011).

16. Dery, L. Characterization and structural analysis of human selenium-dependent glutathione peroxidase 4 mutant expressed in *Escherichia coli*. *J. Biol. Chem.* 281, 14316–14314 (2006).

17. Dery, L. Evolution of translation machinery in recoded bacteria enables top-down characterization of disulfide-containing proteins. *Nat. Biotechnol.* 33, 1272–1277 (2015).

18. Mandell, D.J. et al. Biocatalysis of genetically modified organisms by synthetic protein design. *Nature* 518, 55–60 (2015).

19. Israel, A. et al. Accessing human selenoproteins through chemical protein synthesis. *Nat. Chem.* 5, 09696 (2015).

20. Cohen, D.T., Zhang, C., Pentelute, B.L. & Buchwald, S.L. An umpolung approach to modulating oxytocin activity and plasma stability by disulfide bond engineering. *J. Med. Chem.* 58, 8585–8596 (2015).

21. Arnér, E.S., Sarioglu, H., Lottspeich, F., Holmgren, A. & Böck, A. High-level efficient canonical incorporation of selenocysteine. *J. Am. Chem. Soc.* 135, 2–5 (2013).

22. Cheng, Q. & Arnér, E.S.J. Selenocysteine insertion at a predefined UAG codon in *Escherichia coli*. *J. Biol. Chem.* 282, 16270–16279 (2007).

23. Fan, Z., Song, J., Guan, T., Lv, X. & Wei, J. Efficient expression of glutathione peroxidase with chimeric tRNA in Amber-less *Escherichia coli*. *ACS Synth. Biol.* 7, 249–257 (2018).

24. Wallewska, A. et al. Integrated oxidative folding of cysteine/selenocysteine containing peptides: improving chemical synthesis of toxins. *Angew. Chem. Int. Edn Engl.* 48, 2221–2224 (2009).

25. Gaciarz, A. et al. Systematic screening of soluble expression of antibody fragments in the cytoplasm of *E. coli*. *Microb. Cell Fact.* 15, 22 (2016).

26. Lobstein, J. et al. Shuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* 11, 753 (2012).

27. Zhang, Y., Cui, W., Zhang, H., Dewald, H.D. & Chen, H. Electrochemistry-assisted top-down characterization of disulfide-containing proteins. *Anal. Chem.* 84, 3838–3842 (2012).

28. Zhao, D.S., Gregorich, Z.R. & Ge, Y. High throughput screening of disulfide-containing proteins in a complex mixture. *Proteomics* 13, 3256–3260 (2013).

29. Kraj, A., Brouwer, H.J., Reinhold, N. & Chervert, J.P. A novel electrochemical method for efficient reduction of disulfide bonds in peptides and proteins prior to MS detection. *Anal. Bioanal. Chem.* 405, 9311–9320 (2013).

30. Sun, G.C., Jiang, J., Onderko, E.L., Silakov, A., Yosca, T.H. & Green, M.T. Characterization of a selenium-dependent glutathione peroxidase with chimeric tRNA in Amber-less *Escherichia coli*. *J. Mol. Biol.* 292, 1003–1016 (1999).

31. 47. Dery, L. Characterization and structural analysis of human selenium-dependent glutathione peroxidase 4 mutant expressed in *Escherichia coli*. *J. Biol. Chem.* 281, 14316–14314 (2006).

32. Pelat, T. et al. Isolation of a human-like antibody fragment (scFv) that neutralizes ricin biological activity. *BMC Biotechnol.* 9, 60 (2009).

33. Rudolph, B., Gebendorfer, K.M., Buchner, J. & Winter, J. Evolution of selenocysteine tRNA for growth at high temperatures. *J. Biol. Chem.* 285, 19029–19034 (2010).

34. Nilsson, M. & Rydén-Aulin, M. Glutamine is incorporated at the nonsense codons for the SME-1 beta-lactamase to confer resistance to beta-lactam antibiotics. *Antimicrob. Agents Chemother.* 47, 1062–1067 (2003).

35. Nilsson, M. & Rydén-Aulin, M. Glutamine is incorporated at the nonsense codons for the SME-1 beta-lactamase to confer resistance to beta-lactam antibiotics. *Antimicrob. Agents Chemother.* 47, 1062–1067 (2003).

36. Rodriguez, E.A. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898 (2009).

37. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898 (2009).

38. Tack, D.S. et al. Addicting diverse bacteria to a noncanonical amino acid. *Nat. Chem.* 12, 138–140 (2016).

39. Wannier, T.M. et al. Long-term adaptive evolution of genomically recoded *Escherichia coli*. Preprint @ https://www.biorxiv.org/content/early/2017/07/12/162834 (2017).

40. Ara, K. et al. Preparation of selenoinsulin as a long-lasting insulin analogue. *Angew. Chem. Int. Edn Engl.* 56, 5522–5526 (2017).

41. Metanis, N. & Hilvert, D. Strategic use of non-native diselenide bridges to steer oxidative protein folding. *Angew. Chem. Int. Edn Engl.* 51, 5585–5588 (2012).

42. Onderko, E.L., Silakov, A., Yosca, T.H. & Green, M.T. Characterization of a selenium-containing ligated P450 compound I reveals direct link between electron donation and reactivity. *Nat. Chem.* 9, 623–628 (2017).

43. Dery, L. et al. Accessing human selenoproteins through chemical protein synthesis. *Chem. Sci. (Camb.)* 8, 1922–1925 (2017).

44. Cohen, D.T., Zhang, C., Penultimate, B.L. & Buchwald, S.L. An umpolung approach for the chemoselective arylation of selenocysteine in unprotected peptides. *J. Am. Chem. Soc.* 137, 9784–9787 (2015).
ONLINE METHODS

Molecular biology. The plasmids described in this work were constructed using Gibson assembly and standard molecular biology techniques. Synthetic genes were obtained as gblocks from IDT. Relevant DNA sequences including those of representative plasmids can be found in Supplementary Tables 4–13.

To construct the parental E. coli strains for the passaging experiment, RTΔA cells were transformed with pKD78 and either pRSF-SelA-PstK (Supplementary Table 6) or pCDF-SelD-UX (Supplementary Table 7) or empty pRSF and pCDF vectors, containing only the origin of replication and the antibiotic resistance marker. RTΔA cells containing the three plasmids were grown to mid-log phase at 30 °C and the lambda red recombination machinery induced by addition of L-arabinose to a final concentration of 0.3% and incubation for 1 h at 37 °C. Electrocompetent cells were prepared from the induced cultures by washing cells three times in 10% glycerol. Cells were transformed with 200 ng double-stranded DNA encoding the bla

strains were characterized by Parental and evolved strains. The plasmids described in this work were constructed using trimmomatic (v0.32). The sequence of the RTΔA genome was assembled using the De novo Assembly Module in Geneious. Variant detection was performed using bresq (v0.27.0), with the assembled RTΔA genome and plasmid sequences as references. Mutations occurring at or above 20% in the Δ, CC, UC, and UU parental strains were removed in their respective evolved populations. Genes enriched with SNPs were defined as having acquired SNPs with >50% frequency in four or more independent populations evolved under the same conditions.

Whole genome sequencing and bioinformatic analysis. Genomic DNA from RTΔA cells, the parental Δ, CC, UC, and UU stocks and each evolved bacterial population was extracted from ~5 × 10^6 cells using a Zymo Research Fungal/Bacterial DNA Kit according to the manufacturer’s instructions. DNA was prepared for sequencing with a 300 base-pair target insert size using standard methods. Samples were sequenced on an Illumina HiSeq 2500 system using 125 base-pair-end reads at the Genome Sequencing and Analysis Facility (University of Texas at Austin). Across all sequenced bacterial lines, an average coverage of 144× was obtained with a s.d. of 43.0. Raw sequencing reads were processed by trimming and removing adapters using trimmomatic (v0.32). The sequence of the RTΔA genome was assembled using the De novo Assembly Module in Geneious. Variant detection was performed using bresq (v0.27.0), with the assembled RTΔA genome and plasmid sequences as references. Mutations occurring at or above 20% in the Δ, CC, UC, and UU parental strains were removed in their respective evolved populations. Genes enriched with SNPs were defined as having acquired SNPs with >50% frequency in four or more independent populations evolved under the same conditions.

MAGE. 100 µL of RTΔA cells containing pKD78 from a saturated culture was diluted into 3 mL of LB supplemented with 33 µg/mL chloramphenicol and grown to mid-log phase at 30 °C. To induce the lambda red-recombination machinery, 100 µL of 10% w/v L-arabinose was added to give a final concentration of 0.3% and the cells transferred to 37 °C and incubated for 1 h. One mL of the induced culture was removed and centrifuged at 8,000g for 1 min to pellet the cells. Cell pellets were resuspended in 10% glycerol and washed three times to prepare electrocompetent cells. Mutagenic oligonucleotides were added to a final concentration of 1 µM each. Cells were electroporated at 1.8 kV, 25 µF capacitance and 200 ohms (Bio-Rad E. coli Pulser) and recovered in three mL LB supplemented with chloramphenicol. Cells were grown to mid-log phase and an additional two MAGE cycles were performed as described above. Following a final 3-h recovery, tenfold serial dilutions were plated on LB supplemented with 33 µg/mL zeocin to obtain single colonies. Mutants were identified using multiplex allele-specific colony (MASC)-PCR, after which the target genes were amplified by PCR and mutations confirmed by Sanger sequencing.

Growth assays. Parental and evolved E. coli strains were characterized by growth assays in both rich and defined media. For characterization in rich media, 5-mL cultures of LB containing 25 µg/mL kanamycin, 50 µg/mL spectinomycin, and 10 µM Na₂SeO₃ were inoculated from glycerol stocks and grown to saturation. A liloquots of each culture were diluted to OD₅₆₀ 1.0 and 1 µL used to inoculate 100-µL cultures, in triplicate, comprising four different carbenicillin concentrations in a 96-well plate. The growth assay media consisted of LB containing 25 µg/mL kanamycin, 50 µg/mL spectinomycin, 10 µM Na₂SeO₃, and the four carbenicillin concentrations were 0, 100, 1,000, or 10,000 µg/mL. Plates were seeded with an optically clear, gas-permeable membrane and incubated with constant orbital agitation (amplitude of 3 mm) at 37 °C. OD₅₆₀ measurements were taken at 5-min intervals for 24 h (Tecan Infinite M200 Pro). Populations evolved with β-lactam stress were assayed using all four carbenicillin concentrations. Populations evolved with thermal conditions even up to 2,500 cell doublings had occurred. For the thermal tolerance experiment, the temperature was increased by 0.5 °C (and the carbenicillin concentration kept constant at 100 µg/mL). At an incubation temperature of 43.5 °C, some cultures could no longer be passaged reliably when the freshly diluted cells were immediately incubated at 43.5 °C. To overcome this problem, all cultures in the thermal experiment were pre-incubated at 43 °C for 3 h and then the incubation temperature elevated to the correct level. At 45.5 °C several cultures had poor viability even with a 43 °C pre-incubation, and all remaining passages were performed at 45 °C.
stress were only assayed at 100 µg/mL carbenicillin. All growth curves are plotted as the mean of three biological replicates performed in technical triplicate ± s.e.m. represented as ribbon.

For characterization in defined media, cultures were started in MOPS EZ containing 0.5 µg/mL n-biotin, 25 µg/mL kanamycin, and 50 µg/mL spectinomycin. Parental strains were observed to grow poorly in MOPS EZ, and cultures were diluted to OD₆₀₀ 0.1 rather than 1.0. Growth assays were performed as previously described using four variations of the MOPS EZ media used for overnight growth: media only, media containing 100 µg/mL carbenicillin, media containing 1 µM Na₂SeO₃ and media containing both carbenicillin and Na₂SeO₃. Mutant strains generated by MAGE were assayed as described above in either LB supplemented with 33 µg/mL zeocin or MOPS EZ containing 33 µg/mL zeocin and 0.5 µg/mL n-biotin. Na₂SeO₃ was supplemented at 10 µM in LB and 1 µM in MOPS EZ. Growth curve data are representative of two or three repeated experiments.

qPCR and mutant allele detection. RTAA cells along with the polA and pcnB mutants were transformed with the pRSF and pCDF empty plasmids. Replicates of each of the three clones for each mutant/plasmid pair were grown to saturation and then diluted 1:100 in a 96-well plate. Plates were incubated in a plate reader (BioTek Cytation 5) at 37 °C for 4 h with constant agitation, where OD₆₀₀ was monitored. Following incubation the plates were removed and put on ice, where a 2-µL aliquot from each well was added to a qPCR reaction mix containing EvaGreen double-stranded DNA dye, and qPCR primers specific for either the aphaI or aadA genes from pRSF and pCDF, respectively. Purified plasmid DNA was quantified, diluted, and used as a standard on each qPCR plate for quantification. Fluorescence was read and analyzed using a Roche Lightcycler 96. Absolute quantifications of each sample were normalized to the OD₆₀₀ of the well corresponding to the sample.

For detection of mutant cysK alleles, cultures were inoculated from glycerol stocks taken from the final serial passage for each sample and incubated at 37 °C with 225 r.p.m. agitation. For samples passaged with selenite, the medium was supplemented with 10 µM Na₂SeO₃ for any further growth. Saturated cultures were diluted 1:50 in LB and incubated for 3 h until reaching mid-log phase. 1.5-µL aliquots of each culture were normalized to the highest OD₆₀₀ (~0.4) and centrifuged at 3,500 x g for 5 min, then resuspended in 100 µL of LB. Cells were then boiled for 15 min at 95 °C to prevent PCR inhibition. Cell debris was pelleted by centrifugation and the supernatant was recovered. Primers specific for wild-type and mutant alleles at T69, T73, and H153 were designed as previously described. Oligos were purchased from IDT (Coralville, IA). Triplicate qPCR reactions (20 µL) were set up using 500 nM Forward and Reverse primer, 10 µL of 2× Fast EvaGreen qPCR Master Mix (Biottium, Inc., Fremont, CA), and 2 µL of cDNA supernatant from each sample. Reactions were run on the Roche Lightcycler 96 and analyzed using the manufacturer’s software. qPCR data shown are representative of three experiments.

**smURFP optimization and fluorescent reporter assays.** To make plasmid pUC-BAD-smURFP (Supplementary Table 10), a fragment containing the araC gene, the P₃BAD promoter and the bicistronic smURFP-pbsA₁ was amplified from pBAD-smURFP-RBS-HO-1 (Addgene plasmid # 80341) and cloned into a CmiI derivative of pUC19. The TAG stop codon in pbsA₁ (encoding Synchocystis sp. PCC 6803 heme oxidase) was changed to TAA during this process. Expression of smURFP and heme oxidase in E. coli DH10B from Supplementary Table 11 with a C-terminal His6-tag. Plasmids were transformed into electrocompetent β₃UU3 cells and single transformants grown to saturation in LB supplemented with 1,000 µg/mL carbenicillin, 12.5 µg/mL tetracycline, and 10 µg/mL Na₂SeO₃. Cultures were diluted 1/250 in terrific broth supplemented with antibiotics and 25 µg Na₂SeO₃ in baffled flasks and incubated at 37 °C with agitation (250 r.p.m.) until reaching mid-log phase. Protein expression was induced by addition of 3.4-dihydroxybenzoic acid to achieve a final concentration of 1 mM. 3,4-dihydroxybenzoic acid was added from a 10× stock, which was freshly prepared in culture media. The optimal induction period and temperature were established for each protein individually. Cells were harvested by centrifugation at 8,000 g for 10 min and the cells resuspended in 25 mL of wash buffer (50 mM K₂HPO₄, 300 mM NaCl, and 10% glycerol at pH 8.0) with protease inhibitor cocktail (cOmplete, mini EDTA free, Roche) and lysed by sonication (Model 500, Fisher Scientific). Lysate was repeatedly clarified by centrifugation (35,000 g for 30 min), filtered through a 0.2-µm membrane and protein recovered by immobilized metal ion affinity chromatography (IMAC) using Ni-NTA resin and gravity flow columns. Eluate was concentrated and dialyzed into the appropriate buffer followed by purification to apparent homogeneity by size exclusion fast protein liquid chromatography (FPLC). DHFR was dialyzed into 50 mM NH₄Ac pH 6.5, GFP variant were dialyzed into 50 mM potassium phosphate pH 7.0 and hGH and all scFvs were dialyzed into TBS pH 7.5.

Human growth hormone (hGH), seleno-Herceptin-scFv and anti-RCA scFvs were expressed from either p15A-P₇₇ or pUC-P₇₇ plasmids (Supplementary Tables 12 and 13). Proteins were expressed and purified as described above except using β₃UU3-T₇ cells that were induced by the addition of anhydrous tetracycline to a final concentration of 200 ng/mL. Wild-type Herceptin-scFv was cloned into PET-11a and expressed in E. coli strains BL21(DE3) and T7 Shuttle expess. Protein yield following size-exclusion chromatography (SEC) was measured by BCA assay (Pierce BCA Protein Assay Kit, Thermo Scientific) using bovine serum albumin as a standard. GFP, aMS2, and hGH were expressed in multiple batches, and the reported yield corresponds to the batch for which mass spectra are presented. GFP U135-U177 was used to validate selenoprotein expression using a number of different expression plasmids, media compositions and other culture parameters. Yields ranged from 10–40 mg/L depending on conditions. The anti-MS2 scFv was expressed in three batches with final yields of ≤1 mg/L, although substantial precipitation was observed during purification and storage. Seleno-hGH was expressed in three batches with yields ranging from 5–9 mg/L. Yield information for other proteins is based on a single-batch purification from a total culture volume of 1 liter under the conditions described above.

**Mass spectrometry.** Samples were desalted using either a Millipore 10-kDa molecular weight cutoff filter (Darmstadt, Germany) or a Bio-Rad Micro
Bio-Spin size exclusion spin column (Hercules, CA) and diluted to a final concentration of 10 µM in 49.5/49.5/1 (water/acetonitrile/formic acid). Water and acetonitrile were purchased from Millipore (Darmstadt, Germany) and formic acid was purchased from Sigma (St. Louis, MO). Samples were loaded in conductive metal-coated pulled tip emitters (prepared in-house) for static nanoelectrospray infusion with an applied voltage of 1.0–1.5 kV. All experiments were performed on a Thermo Orbitrap Elite mass spectrometer (Bremen, Germany) equipped with a Coherent Excistar XS 193 nm excimer laser (Santa Cruz, CA) for photodissociation. All spectra were collected at a resolving power of 240,000 (at m/z 400) with the higher-energy collisional dissociation (HCD) cell pressure reduced to ~ 5 mTorr as previously described. For ultraviolet photodissociation (UVPD), isolated precursor ions were irradiated with 1–2 laser pulses with 1–2.5 mJ per pulse in the HCD cell and transferred to the Orbitrap analyzer for high-resolution mass analysis. Up to 100 scans were averaged for each spectrum with 10 µs per scan. All spectra were deconvoluted using Thermo Xtract algorithm with a S/N threshold of 3 and analyzed with ProSight Lite equipped with UVPD search capabilities. Protein backbone fragments generated during UVPD are represented in sequence maps as green, blue, and red flags with each indicating cleavage of Cα-C, C-N, and N-Cα backbone bonds, respectively. For each disulfide or diselenide bond, the mass of one hydrogen was subtracted from each residue. In sequence maps, cysteines engaged in disulfide bonds and selenocysteines engaged in diselenide bonds are highlighted with gray and yellow boxes, respectively. Additionally, in sequence maps of GFP, the tyrosine in the mature chromophore is represented as a yellow box. For selenocysteine-containing proteins, intact protein mass measurements from deconvoluted mass spectra and fragment ions containing selenocysteine from deconvoluted UVPD spectra were mass-shifted because the Xtract algorithm does not effectively account for the unusual isotope distribution of selenium. This mass shift corresponded to ~1 Da difference per selenocysteine residue. To account for this shift during mass error (p.p.m.) calculations of intact proteins, 1 Da was added to the deconvoluted mass for each selenocysteine. To account for the mass shift when searching UVPD (MS/MS) data, selenocysteine sites were each manually adjusted in ProSight Lite by subtracting the mass of 1 Da from each selenocysteine (in addition to one hydrogen loss to from each selenocysteine to account for diselenide formation) such that theoretically generated fragment ions could be correctly matched to the observed fragment ions containing selenocysteine. Both mass shifts associated with disulfide or diselenide bond formation in the ESI mass spectrum and gaps in sequence coverage of the UVPD spectra were used to identify the presence of disulfide or diselenide bonds.

Enzyme-linked immunosorbent assay (ELISA). Wild-type and seleno-antitoxin A-chain scFvs were diluted to 1 µM in TBS pH 7.5 with 0, 1, 10, or 50 mM DTT and incubated for 16 h at 37 °C. Ricin A-chain (Sigma L9514) was diluted to 10 µg/mL in PBS and bound to a 96-well plate overnight at 4 °C. Control plates were bound with PBS containing 5% w/v skim milk. Plates were washed with PBS and blocked with PBS 5% w/v skim milk for 2 h. Following incubation with DTT, scFvs were diluted in PBS with 5% w/v skim milk and 0.05% Tween 20 to 0, 0.25, 0.5, 1, 1.5, 2, 4, 6, 8, 10, 15, and 20 nM and bound in quadruplicate for 2 h at 25 °C. Secondary antibody (anti-polyhistidine-HRP, Sigma A7058) was diluted 1/10,000 in PBS with 5% w/v skim milk and 0.05% Tween 20 and bound for 2 h at 25 °C. Plates were incubated with 3,3′,5,5′-Tetramethylbenzidine (TMB) for 10 min at 25 °C with agitation and color development terminated by the addition of 2 M H₂SO₄. Absorbance was immediately measured at 450 nm. All binding, blocking, and enzymatic steps were performed on an orbital shaker at 450 r.p.m., and all washes were performed with PBS or PBS with 0.05% Tween 20. ELISA data shown are representative of three experiments.

Statistical analysis and reproducibility. All data in the manuscript are displayed as mean ± s.e.m. unless specifically indicated. Bacterial growth curves and bar graphs were plotted in R 3.1.2 using the package ggplot2. The ELISA curves were estimated in R using a general asymmetric five parameter logistic model with the package drm and plotted using ggplot2.

Life Sciences Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The whole genome sequence of E. coli strain β_UU3 is available on GenBank. Bacterial strains and all other materials described in the manuscript are available from the authors upon request.

Accession codes. Primary accessions: β_UU3: GenBank accession number CP023749. Referenced accessions: C321.ΔA: GenBank Accession number CP006698.1.

49. Byrom, M., Bhadra, S., Jiang, Y.S. & Ellington, A.D. Exquisite allele discrimination by toehold hairpin primers. *Nucleic Acids Res.* **42**, e120 (2014).
50. Shaw, J.B. *et al.* Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J. Am. Chem. Soc.* **135**, 12646–12651 (2013).