HHS Public Access

Author manuscript

Nat Biotechnol. Author manuscript; available in PMC 2018 December 04.

Published in final edited form as:

Nat Biotechnol. 2018 August ; 36(7): 624–631. doi:10.1038/nbt.4154.

Custom selenoprotein production enabled by laboratory evolution of recoded bacterial strains

Ross Thyer[*],
Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, United States

Raghav Shroff,
Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, United States

Dustin R. Klein,
Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States

Simon d'Oelsnitz,
Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, United States

Victoria C. Cotham,
Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States

Michelle Byrom,
Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, United States

Jennifer S. Brodbelt, and
Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States

Andrew D. Ellington[*]

[*]To whom correspondence should be addressed: Ross Thyer, ross.thyer@utexas.edu; Andrew D. Ellington, ellingtonlab@gmail.com.

Editors summary
Selenocysteine is efficiently incorporated into recombinant proteins using evolved *E. coli* strains.

Accession codes
Primary accessions:
β_UU3: Genbank Accession No. CP023749
Referenced accessions:
C321.ΔA: Genbank Accession No. CP006698.1

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. 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.

Supplementary information
Supporting note, figures and tables.

Competing financial 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. R.S., D.R.K, S.D., V.C., M.B. and J.S.B. have no competing financial interests.

Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, United States

Abstract

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 *E. 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 21st amino acid selenocysteine is a promising candidate for incorporation into engineered proteins by synthetic biologists. Unfortunately, the complex 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 Böck[1]), 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 machinery[2], and is outcompeted by release factor two for access to the UGA codon, resulting in premature termination of protein synthesis. Despite attempts to bypass these constraints on selenocysteine incorporation in bacteria[3, 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* $tRNA^{Sec}$ variant[5] that enabled efficient, site specific selenocysteine incorporation at amber stop codons and overcame sequence constraints imposed by the endogenous bacterial selenocysteine incorporation machinery. Using this system we produced recombinant proteins containing selenyl-sulfhydryl and diselenide bonds[5]. 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 (SI Fig. 1). We hypothesized that this might be because the recoded *E. coli* host (RTΔA[5], which is derived from C321.ΔA[6]), had not adapted to the effects of genomic recoding[6] 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 group[7], which resulted in complex genetic adaptations that supported the ability to metabolize a carbon source it couldn't previously use[8]. 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 $tRNA^{SecUX}$, 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 $bla_{NMC\text{-}A}$ gene (disulfide-dependent β-lactamase from *Enterobacter cloacae*) into the chromosome[9]. 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-sulfhydryl bond and a U69-U238 variant (UU) containing an essential diselenide 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 (Na_2SeO_3). A wild-type $bla_{NMC\text{-}A}$ 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 2500 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 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[10, 11]; (2) both conditions have previously been used therefore 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 $tRNA^{Sec}$ stability, which in turn decreases selenocysteine

incorporation[12], 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 three-fold higher culture density and spent 20–40% less time in lag phase (Fig. 1c and SI 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 dramatically improved (Fig. 1d and SI 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 Na_2SeO_3, showing that selenocysteine dependence was maintained throughout the 2500 generations. No populations acquired unconditional dependence on selenocysteine, which would necessitate Na_2SeO_3 supplementation in the absence of carbenicillin.

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 (SI 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 (SI 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 analyse 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 SNPs and no genomic rearrangements. WGS indicated that the selenocysteine biosynthesis and incorporation machinery was maintained throughout evolution, that there was no loss of TAG codons in the $bla_{\text{NMC-A}}$ gene and no contamination between the populations. We did not find any genomic suppressor tRNAs using WGS. We detected several new in-frame TAG codons in all sequenced genomes, including β_Δ and T_Δ populations which contain no suppressor tRNAs (SI Table 1). Mass spectrometry of a model protein (GFP) expressed in control strains confirmed that UAG was decoded as glutamine (SI Fig. 6), which is expected because $tRNA^{GlnCUG}$ is the most efficient near-cognate UAG suppressor[14].

Genome sequencing identified clusters of SNPs in genes conferring antibiotic resistance ($bla_{\text{NMC-A}}$, *ftsI* and *ompR*), 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. Interestingly, no mutations were observed in the *selA*, *pstK*, $tRNA^{SecUX}$ 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)[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 *yeeJ* to the wild-type sequence present in *E. coli* strain MG1655, the parental strain of C321.ΔA. The growth of MAGE-engineered strains was compared with that of RTΔA, and four mutations were identified that simultaneously decrease time spent in lag phase and increase the final culture density (Fig. 2b–e). These mutations included three reversions of SNPs that were acquired during construction of the C321.ΔA strain, and the *prfB* T246A mutation which repairs a defect in release factor 2 that is only present in *E. coli* K12 strains[16, 17]. Although most of the SNPs we identified either compensate for metabolic defects in the parental recoded *E. coli* or broadly improve cell growth, *cysK* and *pcnB* directly affect 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 Supporting Note 1, SI Figs. 7–9 and SI Tables 2–3.

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-$tRNA^{Sec}$, the immediate precursor in selenocysteine biosynthesis. Traditional genetic reporters for selenocysteine incorporation are available e.g. formate dehydrogenase, but are only qualitative, due to their narrow dynamic range, plus they require anaerobic growth conditions[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[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[20], so we hypothesized that a selenocysteine residue would support formation of a selenoether crosslink. To this end we replaced the essential cysteine residue in smURFP[20], an engineered autocatalytic phycobiliprotein, with selenocysteine (Fig. 3b). Upon comparison, a C52U variant of smURFP, designated seleno-smURFP, was >20-fold more fluorescent than a C52S variant (Fig. 3c–d). The difference in fluorescence between the C52 and C52U variants is likely due to differences in protein abundance, as canonical incorporation of

cysteine in response to its cognate sense codons (UGC in smURFP) is more efficient than selenocysteine incorporation in response to UAG. Omission of Na_2SeO_3 from the culture media abolished fluorescence of seleno-smURFP which confirmed chromophore ligation was dependent on selenocysteine. Serine incorporation is observed in model proteins expressed in the absence of selenite supplementation (SI Fig. 10). Combining the advantages of traditional fluorescent reporters with excellent specificity 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, SI Fig. 11). Transformants from populations dependent on selenocysteine incorporation (β_UC, β_UU, 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 (SI Fig. 11). No transformants from populations β_Δ 1–3 and T_Δ 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 5- to 7-fold increases in selenoprotein expression compared to the parental cells. Evolved β_UU populations show 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 β_UU strains achieved several fold higher cell density (as observed in Fig. 1c), increasing overall selenoprotein yield in some populations up to 20-fold.

A single clone from the β_UU3 population was isolated and we expressed a series of model proteins containing pairs of selenocysteine residues to calculate the specific selenocysteine incorporation efficiency. *Ec*DHFR U39-U85 and GFP U135-U177 were expressed at yields of 8 and 40 $mg.L^{-1}$ respectively and only intact masses corresponding to selenocysteine incorporation were detected by mass spectrometry (SI Fig. 12). Ultraviolet photodissociation (UVPD) of *Ec*DHFR 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[21], 1 $mg.L^{-1}$) containing two essential bonds, and human growth hormone (hGH, 9 $mg.L^{-1}$) which contains two bonds, one of which is essential. Mass spectrometry confirmed complete

selenocysteine incorporation and correct diselenide formation in both proteins (SI Fig. 13). Recent attempts to express recombinant selenoproteins in *E. coli* (summarized by Cheng and Arnèr[22]), including in genomically recoded *E. coli* strains, have reported yields in the low mg.L^{-1} range with 10–30% selenocysteine content at a single position[22, 23]. Our results represent a substantial advance in terms of yield, incorporation efficiency and number of selenocysteine residues per protein.

Diselenide bond formation reduces structural heterogeneity

Due to the lower redox potential of diselenide bonds (−381 mV compared to −180 mV for disulfide bonds)[24], 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[25]. This property allowed us to observe species with both oxidised 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^{-1}), 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, SI Fig. 14) and calculated a monoisotopic mass of 27172.05 Da, corresponding to a sample containing only reduced cysteine residues. 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^{-1}), an *E. coli* strain which maintains an oxidizing cytoplasm and constitutively expresses disulfide bond isomerase (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 27170.12 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_H domain contained a gap in the distribution of fragment ions, confirming the presence of only one disulfide bond (Fig. 4d, SI 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 β_UU3-T7 cells (0.8 mg.L^{-1}) were confined to 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-scFvs expressed in *E. coli* strain BL21(DE3), T7 Shuffle Express, and seleno-

Herceptin-scFv expressed in β_UU3-T7 cells further suggest protein populations that are 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, SI Fig. 14).

Taken together these data confirm 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[31]. This is well established in small peptides where bond formation is the primary driver of structure[32], 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[33], which contains two essential bonds, in both disulfide (T7 Shuffle Express, 1 $mg.L^{-1}$) and diselenide (β_UU3-T7, 1 $mg.L^{-1}$) configurations, confirmed by mass spectrometry (Fig. 5, SI Fig. 15). Proteins were incubated with varying concentrations of 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 $tRNA^{Sec}$ variant and an engineered β-lactamase containing an essential diselenide bond. Dependence on selenocysteine incorporation was maintained for more than 2500 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[34] that have been optimized using different methods. With recent advances in orthogonal translation systems[35] and protein design this type of approach should prove generalizable to other amino acids in other proteins[36–38].

Previous attempts to improve genomically recoded *E. coli* strains have focused exclusively on growth rate[34, 39] but have found less substantial improvements in the suppression of UAG

codons by non-canonical amino acids. Using our evolved strains, we have dramatically 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 lay the foundation for 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[22]. 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[40] and oxytocin[31], 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 into seleno-biobetters.

In addition to enabling the incorporation of diselenide bonds as a protein engineering tool[41], 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[42, 43], but progress has been hindered by inefficient protein expression in cysteine auxotrophs or by specialized protein ligation strategies[44]. In addition, these approaches have inherent limitations, requiring either the removal of cysteine residues or accepting indiscriminate selenocysteine incorporation, or a need for the truncated enzyme fragments to remain soluble.

Similarly, efforts to characterize the human selenoproteome (comprising 25 proteins of which half are uncharacterized), have relied on auxotrophic selenocysteine incorporation[45], or substitution 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[46]. Recent attempts to produce two of these proteins have been successful using complete chemical synthesis of the entire selenoprotein[47]. Expression in a bacterial host will enable easier analysis of these proteins.

Furthermore, selenocysteine dependent conjugation chemistries[48], 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.

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 the Supporting Information (SI 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 (SI Table 6) and pCDF-SelD-UX (SI 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 recombinase machinery induced by addition of L-arabinose to a final concentration of 0.3% and incubation for one hour 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 linear dsDNA encoding the $bla_{\text{NMC-A}}$ gene under the control of the *Serratia marcescens* $P_{bla\text{SME}-1}$ promoter with 50 bp homology arms complementary to the *E. coli ahpC* locus (SI Table 8). RTΔA cells containing pRSF-SelA-PstK and pCDF-SelD-UX were transformed with a wild-type $bla_{\text{NMC-A}}$ gene (C69-C238), a $bla_{\text{NMC-A}}$ gene containing a C69U mutation, or a $bla_{\text{NMC-A}}$ double mutant containing the C69U and C238U mutations. RTΔA cells containing the empty pRSF and pCDF vectors were transformed with the wild-type $bla_{\text{NMC-A}}$ gene only. Cells were plated on LB agar supplemented with 50 μg.mL kanamycin, 100 μg.mL spectinomycin and either 50 or 100 μg.mL ampicillin. Plates for the U69 and U69-U238 variants were supplemented with 10 μM Na_2SeO_3. No U69-U238 colonies developed on the 100 μg.mL plate and an extended incubation (48 hours) was required to develop colonies on the 50 μg.mL plate. Resulting ampicillin resistant colonies were screened by PCR to identify correct integrations, which were confirmed by Sanger sequencing. A single verified clone from the RTΔA cells containing the empty pRSF and pCDF vectors was designated the Δ parent strain. Single verified clones containing the pRSF-SelA-PstK and pCDF-SelD-UX plasmids and the three $bla_{\text{NMC-A}}$ gene variants (wild-type, U69 and U69-U238) were designated the CC, UC and UU parent strains respectively.

To construct a strain with improved characteristics for protein expression we integrated the T7 RNA polymerase gene into a clonal isolate of β_UU3. A cassette was assembled which encoded T7 RNAP under the control of the P_{LtetO} promoter and a constitutively expressed bicistronic sequence containing the *aac3* gene conferring gentamicin resistance and the tet repressor. This cassette was flanked by 500 bp homology arms corresponding to the *E. coli lac* locus (SI Table 9) and the entire fragment cloned into an R6K plasmid derived from

pKD4. The T7 RNAP cassette including the *lac* homology arms was amplified by PCR and integrated into a clone of β_UU3 containing pKD78 as described for bla_{NMC-A}. Gentamicin resistant colonies were screened by PCR and the sequence of the T7 RNAP and tet repressor confirmed by Sanger sequencing. The resulting strain was designated β_UU3-T7.

NMC-A β-lactamase Assay

To assess the toxicity of selenocysteine biosynthesis and identify a possible threshold effect for selenocysteine incorporation, RTΔA cells were transformed with p15A-NMC-A (U69-U238) (SI Table 5) and one of two pRSF-UX-SelA variants; one encoding a wild-type *selA* gene (SI Table 4) and one encoding a *selA* gene truncated by a stop codon. Cells were recovered in SOC for one hour and then diluted into 20 mL LB containing 50 μg.mL kanamycin, 12.5 μg.mL tetracycline and 10 μM Na_2SeO_3 and incubated overnight. 10-fold dilutions were performed from the saturated cultures and plated on LB agar containing 50 μg.mL kanamycin, 10 μM Na_2SeO_3 and either 12.5 μg.mL tetracycline or 100 μg.mL carbenicillin.

Bacterial Passaging

5mL cultures of LB supplemented with 25 μg.mL kanamycin, 50 μg.mL spectinomycin, 10 μM Na_2SeO_3 and 100 μg.mL carbenicillin were inoculated in triplicate with the Δ, CC, UC and UU parent strains. Following growth to stationary phase, cells were diluted 5000-fold into fresh media resulting in ~ 12.5 doublings every passage. Glycerol stocks were prepared from all cultures every five passages and the selection stringency was increased every 10 passages. During the β-lactam resistance experiment, stringency was adjusted by increasing the carbenicillin concentration by 100 μg.mL. This continued evenly until 2500 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 three hours 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.

For the competition experiment between wild-type RTΔA cells and those carrying mutant *cysK* alleles, strains were initially grown to saturation, diluted to OD_{600} 0.1 and mixed in a 1:1 ratio. 5 mL cultures of LB supplemented with 33 μg.mL zeocin and ± 10 μM Na_2SeO_3 were inoculated in triplicate with 1 μL of the RTΔA:mutant cell mix. Cultures were incubated at 37 °C and serially passaged to saturation 10 times (125 generations).

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 approximately 5×10^9 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-paired-end reads at the

Genome Sequencing and Analysis Facility (University of Texas at Austin). Across all sequenced bacterial lines, an average coverage of 144.4X was obtained with a standard deviation 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 breseq (v0.27.0) using the assembled RTΔA genome and plasmid sequences as references. Mutations occurring at or above 20 percent 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 were diluted into three 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 one hour. One mL of the induced culture was removed and centrifuged at 8000 × *g* for one minute 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 hour recovery, 10-fold serial dilutions were plated on LB supplemented with 33 μg.mL zeocin to obtain single colonies. Mutants were identified using 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 Na_2SeO_3 were inoculated from glycerol stocks and grown to saturation. Aliquots of each culture were diluted to OD_{600} 1.0 and 1 μL used to inoculate 100 μL cultures, in triplicate, comprising of 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 Na_2SeO_3 and the four carbenicillin concentrations were 0, 100, 1000, or 10000 μg.mL. Plates were sealed with an optically clear gas permeable membrane and incubated with constant orbital agitation (amplitude of 3 mm) at 37 °C. OD_{600} measurements were taken at five minute intervals for 24 hours (Tecan Infinite M200 Pro). Populations evolved with β-lactam stress were assayed using all four carbenicillin concentrations. Populations evolved with thermal 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 ± SEM represented as ribbon.

For characterization in defined media, cultures were started in MOPS EZ containing 0.5 μg.mL D-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 OD_{600} 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 Na_2SeO_3 and media containing both carbenicillin and Na_2SeO_3. 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 D-biotin. Na_2SeO_3 was supplemented at 10 μM in LB and 1 μM in MOPS EZ. Growth curve data is representative from two or three repeated experiments.

qPCR and Mutant Allele Detection

RTDA 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 four hours with constant agitation, where OD_{600} was monitored. Following incubation the plates were removed and put on ice, where a 2 uL aliquot from each well was added to a qPCR reaction mix containing EvaGreen double stranded DNA dye, and qPCR primers specific for either the *aphA1* 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 OD_{600} 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 rpm agitation. For samples passaged with selenite, the medium was supplemented with 10 μM Na_2SeO_3 for any further growth. Saturated cultures were diluted 1:50 in LB and incubated for 3 hours until reaching mid-log phase. 1.5 mL aliquots of each culture were normalized to the highest OD_{600} (~ 0.4) and centrifuged at 3500 × *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[49]. 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 2X Fast EvaGreen qPCR Master Mix (Biotium, Inc., Fremont, CA), and two μL of cell supernatant from each sample. Reactions were run on the Roche LightCycler 96 and analyzed using the manufacturer's software. qPCR data shown is representative from three experiments.

smURFP Optimization and Fluorescent Reporter Assays

To make plasmid pUC-BAD-smURFP (SI Table 10), a fragment containing the *araC* gene, the P_{BAD} promoter and the bicistronic smURFP-*pbsA1* was amplified from pBAD-smURFP-RBS-HO-1 (Addgene plasmid # 80341) and cloned into a Cm^R derivative of pUC19. The TAG stop codon in *pbsA1* (encoding *Synechocystis* sp. PCC 6803 heme oxidase) was changed to TAA during this process. Expression of smURFP and heme oxidase

in *E. coli* DH10B from either pUC-BAD-smURFP or pBAD-smURFP-RBS-HO-1 resulted in green cultures or colonies which were highly fluorescent in the red spectrum. In contrast, RTΔA cells or β_UU3 cells transformed with pUC-BAD-smURFP and induced with 0.33% L-arabinose were not visibly green and were only weakly fluorescent. This was assumed to be due to a lack of the biliverdin chromophore, likely a result of defects in the heme biosynthesis pathway limiting the pool of intracellular heme. Both strains contain a L196P mutation in HemA which catalyzes the initial and rate limiting step of heme biosynthesis in *E. coli*, a mutation inherited from C321.ΔA known to deleteriously affect cell growth. To overcome this problem, 5-aminolevulinic acid (5-ALA) was added to bypass HemA and increase flux through the heme biosynthesis pathway. Fluorescence assays were performed with varying concentrations of 5-ALA and a final concentration of 100 μg.mL was found to result in cells which were visibly green and highly fluorescent.

Cells transformed with pUC-BAD-smURFP variants (C52, S52, U52, TAA52) were grown to saturation in LB supplemented with 25 μg.mL kanamycin, 50 μg.mL spectinomycin, 100 μg.mL carbenicillin, 33 μg.mL chloramphenicol and 10 μM Na_2SeO_3. Cultures were diluted 1/45 into 900 μL fresh LB containing antibiotics and 0, 10 or 25 μM Na_2SeO_3 and incubated at 37 °C with agitation in a 2 mL 96-well plate for three hours. Cells were induced by the addition of 100 μL of media containing 3.3% L-arabinose to achieve a final concentration of 0.33%. Cultures were incubated for 16 hours and the cells recovered by centrifugation. Cell pellets were resuspended in PBS and fluorescence was measured using excitation and emission wavelengths of 633 and 695 nm respectively. Fluorescence was normalized to OD_{600}. For smURFP assays on solid media, aliquots were taken from saturated overnight cultures, diluted to OD_{600} 0.1 and 3 μL dotted onto agar plates containing appropriate antibiotics, 0.33% L-arabinose and 0, 10 or 25 μM of Na_2SeO_3. Plates were imaged with a Chemi Doc MP imaging system (Bio-Rad) using a manual exposure of 0.1 seconds and analyzed using Image Lab 5.2.1.

Protein Purification

Coding sequences for DHFR, GFP variants and anti-MS2 scFv were cloned into a p15A-PcaU plasmid (SI Table 11) with a C-terminal His6-tag. Plasmids were transformed into electrocompetent β_UU3 cells and single transformants grown to saturation in LB supplemented with 1000 μg.mL carbenicillin, 12.5 μg.mL tetracycline and 10 μM Na_2SeO_3. Cultures were diluted 1/250 in terrific broth supplemented with antibiotics and 25 μM Na_2SeO_3 in baffled flasks and incubated at 37 °C with agitation (250 rpm) 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 10x 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 8000 × *g* for 10 minutes and the cell pellets resuspended in 25 mL of wash buffer (50 mM K_2HPO_4, 300 mM NaCl and 10% glycerol at pH 8.0) with protease inhibitor cocktail (cOmplete, mini EDTA free, Roche) and lysozyme (0.5 $mg.mL^{-1}$). Cells were incubated for 20 min at 4 °C with gentle agitation and lysed by sonication (Model 500, Fisher Scientific). Lysate was repeatedly clarified by centrifugation (35000 × *g* for 30 min), filtered through a 0.2 μm membrane and protein recovered by 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 FPLC. DHFR was dialysed into 50 mM NH_4Ac pH 6.5, GFP variants were dialysed into 50 mM potassium phosphate pH 7.0 and hGH and all scFvs were dialysed into TBS pH 7.5.

Human growth hormone (hGH), seleno-Herceptin scFv and anti-RCA scFvs were expressed from either p15A-P_{T7} or pUC-P_{T7} plasmids (SI Table 12 and 13). Proteins were expressed and purified as described above except using β_UU3-T7 cells which 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 Shuffle express. Protein yield following 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 $mg.L^{-1}$ depending on conditions. The anti-MS2 scFv was expressed in three batches with final yields of ≤1 $mg.L^{-1}$, although substantial precipitation was observed during purification and storage. Seleno-hGH was expressed in three batches with yields ranging from 5–9 $mg.L^{-1}$. Yield information for other proteins is based on a single batch purification from a total culture volume of 1 L 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 nano-electrospray 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[50]. 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[50]. 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 μscans 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 C_α-C, C-N, and N-C_α 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 approximately one Da difference per selenocysteine residue. To account for this shift during mass error (ppm) 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 one 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 full mass spectra and gaps in sequence coverage of the UVPD spectra were used to identified the presence of disulfide or diselenide bonds.

ELISA

Wild-type and seleno-anti-RCA scFv were diluted to 1 μM in TBS pH 7.5 with 0, 1, 10 or 50 mM DTT and incubated for 16 hours 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 two hours. 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 two hours 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 two hours at 25 °C. Plates were incubated with TMB for 10 minutes at 25 °C with agitation and color development terminated by the addition of 2M H_2SO_4. Absorbance was immediately measured at 450 nm. All binding, blocking and enzymatic steps were performed on an orbital shaker at 450 rpm, and all washes were performed with PBS or PBS with 0.05% Tween 20. ELISA data shown is representative from three experiments.

Statistical Analysis and Reproducibility

All data in the manuscript is displayed as mean ± standard error 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. Additional information can be found in the accompanying Life Science Reporting Summary.

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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 ARO (SP0036191-PROJ0009952) and the NIH NCI (5K99CA207870-02 to R.T.) is acknowledged.

References

1. Yoshizawa S, Bock A. The many levels of control on bacterial selenoprotein synthesis. Biochim Biophys Acta. 2009; 11:27.
2. Suppmann S, Persson BC, Bock A. Dynamics and efficiency in vivo of UGA-directed selenocysteine insertion at the ribosome. Embo J. 1999; 18:2284–2293. [PubMed: 10205181]
3. Arner ES, Sarioglu H, Lottspeich F, Holmgren A, Bock A. High-level expression in Escherichia coli of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the selA, selB and selC genes. J Mol Biol. 1999; 292:1003–1016. [PubMed: 10512699]
4. Thyer R, Filipovska A, Rackham O. Engineered rRNA enhances the efficiency of selenocysteine incorporation during translation. J Am Chem Soc. 2013; 135:2–5. [PubMed: 23256865]
5. Thyer R, Robotham SA, Brodbelt JS, Ellington AD. Evolving tRNA(Sec) for efficient canonical incorporation of selenocysteine. J Am Chem Soc. 2015; 137:46–49. [PubMed: 25521771]
6. Lajoie MJ, et al. Genomically recoded organisms expand biological functions. Science. 2013; 342:357–360. [PubMed: 24136966]
7. Blount ZD, Barrick JE, Davidson CJ, Lenski RE. Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature. 2012; 489:513–518. [PubMed: 22992527]
8. Quandt EM, et al. Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. Elife. 2015; 14:09696.
9. Majiduddin FK, 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. 2003; 47:1062–1067. [PubMed: 12604542]
10. Gutierrez A, et al. beta-Lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. Nat Commun. 2013; 4
11. Layton JC, Foster PL. Error-prone DNA polymerase IV is regulated by the heat shock chaperone GroE in Escherichia coli. J Bacteriol. 2005; 187:449–457. [PubMed: 15629916]
12. Ishii TM, Kotlova N, Tapsoba F, Steinberg SV. The long D-stem of the selenocysteine tRNA provides resilience at the expense of maximal function. J Biol Chem. 2013; 288:13337–13344. [PubMed: 23525102]
13. Rudolph B, Gebendorfer KM, Buchner J, Winter J. Evolution of Escherichia coli for growth at high temperatures. J Biol Chem. 2010; 285:19029–19034. [PubMed: 20406805]
14. Nilsson M, Ryden-Aulin M. Glutamine is incorporated at the nonsense codons UAG and UAA in a suppressor-free Escherichia coli strain. Biochim Biophys Acta. 2003; 13:1–6.
15. Wang HH, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature. 2009; 460:894–898. [PubMed: 19633652]
16. Baggett NE, Zhang Y, Gross CA. Global analysis of translation termination in E. coli. PLoS Genet. 2017; 13
17. Mora L, Heurgue-Hamard V, de Zamaroczy M, Kervestin S, Buckingham RH. Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. J Biol Chem. 2007; 282:35638–35645. [PubMed: 17932046]

18. Lacourciere GM, Levine RL, Stadtman TC. Direct detection of potential selenium delivery proteins by using an Escherichia coli strain unable to incorporate selenium from selenite into proteins. Proc Natl Acad Sci U S A. 2002; 99:9150–9153. [PubMed: 12084818]
19. Scheer H, Zhao KH. Biliprotein maturation: the chromophore attachment. Mol Microbiol. 2008; 68:263–276. [PubMed: 18284595]
20. Rodriguez EA, et al. A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. Nat Methods. 2016; 13:763–769. [PubMed: 27479328]
21. Miklos AE, et al. Structure-based design of supercharged, highly thermoresistant antibodies. Chem Biol. 2012; 19:449–455. [PubMed: 22520751]
22. Cheng Q, Arnér ESJ. Selenocysteine Insertion at a Predefined UAG Codon in a Release Factor 1 (RF1)-depleted Escherichia coli Host Strain Bypasses Species Barriers in Recombinant Selenoprotein Translation. Journal of Biological Chemistry. 2017; 292:5476–5487. [PubMed: 28193838]
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 Synthetic Biology. 2017
24. Walewska A, et al. Integrated oxidative folding of cysteine/selenocysteine containing peptides: improving chemical synthesis of conotoxins. Angew Chem Int Ed Engl. 2009; 48:2221–2224. [PubMed: 19206132]
25. Gaciarz A, et al. Systematic screening of soluble expression of antibody fragments in the cytoplasm of E. coli. Microb Cell Fact. 2016; 15:016–0419.
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. 2012; 11:1475–2859.
27. Zhang Y, Cui W, Zhang H, Dewald HD, Chen H. Electrochemistry-assisted top-down characterization of disulfide-containing proteins. Anal Chem. 2012; 84:3838–3842. [PubMed: 22448817]
28. Zhao DS, Gregorich ZR, Ge Y. High throughput screening of disulfide-containing proteins in a complex mixture. Proteomics. 2013; 13:3256–3260. [PubMed: 24030959]
29. Kraj A, Brouwer HJ, Reinhoud N, Chervet JP. A novel electrochemical method for efficient reduction of disulfide bonds in peptides and proteins prior to MS detection. Anal Bioanal Chem. 2013; 405:9311–9320. [PubMed: 24077854]
30. Valentine SJ, Anderson JG, Ellington AD, Clemmer DE. Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. The Journal of Physical Chemistry B. 1997; 101:3891–3900.
31. Muttenthaler M, et al. Modulating oxytocin activity and plasma stability by disulfide bond engineering. J Med Chem. 2010; 53:8585–8596. [PubMed: 21117646]
32. Armishaw CJ, et al. Alpha-selenoconotoxins, a new class of potent alpha7 neuronal nicotinic receptor antagonists. J Biol Chem. 2006; 281:14136–14143. [PubMed: 16500898]
33. Pelat T, et al. Isolation of a human-like antibody fragment (scFv) that neutralizes ricin biological activity. BMC Biotechnol. 2009; 9:1472–6750.
34. Kuznetsov G, et al. Optimizing complex phenotypes through model-guided multiplex genome engineering. Genome Biol. 2017; 18:017–1217.
35. Amiram M, et al. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. Nature biotechnology. 2015; 33:1272–1279.
36. Rovner AJ, et al. Recoded organisms engineered to depend on synthetic amino acids. Nature. 2015; 518:89–93. [PubMed: 25607356]
37. Mandell DJ, et al. Biocontainment of genetically modified organisms by synthetic protein design. Nature. 2015; 518:55–60. [PubMed: 25607366]
38. Tack DS, et al. Addicting diverse bacteria to a noncanonical amino acid. Nat Chem Biol. 2016; 12:138–140. [PubMed: 26780407]
39. Wannier TM, et al. Long-term adaptive evolution of genomically recoded Escherichia coli. bioRxiv. 2017
40. Arai K, et al. Preparation of Selenoinsulin as a Long-Lasting Insulin Analogue. Angew Chem Int Ed Engl. 2017; 56:5522–5526. [PubMed: 28394477]

41. Metanis N, Hilvert D. Strategic use of non-native diselenide bridges to steer oxidative protein folding. Angew Chem Int Ed Engl. 2012; 51:5585–5588. [PubMed: 22454335]
42. Onderko EL, Silakov A, Yosca TH, Green MT. Characterization of a selenocysteine-ligated P450 compound I reveals direct link between electron donation and reactivity. Nat Chem. 2017; 9:623–628. [PubMed: 28644466]
43. Vandemeulebroucke A, Aldag C, Stiebritz MT, Reiher M, Hilvert D. Kinetic consequences of introducing a proximal selenocysteine ligand into cytochrome P450cam. Biochemistry. 2015; 54:6692–6703. [PubMed: 26460790]
44. Hondal RJ. Using chemical approaches to study selenoproteins-focus on thioredoxin reductases. Biochim Biophys Acta. 2009; 11:4.
45. Yu Y, et al. Characterization and structural analysis of human selenium-dependent glutathione peroxidase 4 mutant expressed in Escherichia coli. Free Radic Biol Med. 2014; 71:332–338. [PubMed: 24681209]
46. Mannes AM, Seiler A, Bosello V, Maiorino M, Conrad M. Cysteine mutant of mammalian GPx4 rescues cell death induced by disruption of the wild-type selenoenzyme. Faseb J. 2011; 25:2135–2144. [PubMed: 21402720]
47. Dery L, et al. Accessing human selenoproteins through chemical protein synthesis. Chem Sci. 2017; 8:1922–1926. [PubMed: 28451306]
48. Cohen DT, Zhang C, Pentelute BL, Buchwald SL. An Umpolung Approach for the Chemoselective Arylation of Selenocysteine in Unprotected Peptides. Journal of the American Chemical Society. 2015; 137:9784–9787. [PubMed: 26225900]
49. Byrom M, Bhadra S, Jiang YS, Ellington AD. Exquisite allele discrimination by toehold hairpin primers. Nucleic Acids Res. 2014; 42:2.
50. Shaw JB, et al. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. J Am Chem Soc. 2013; 135:12646–12651. [PubMed: 23697802]

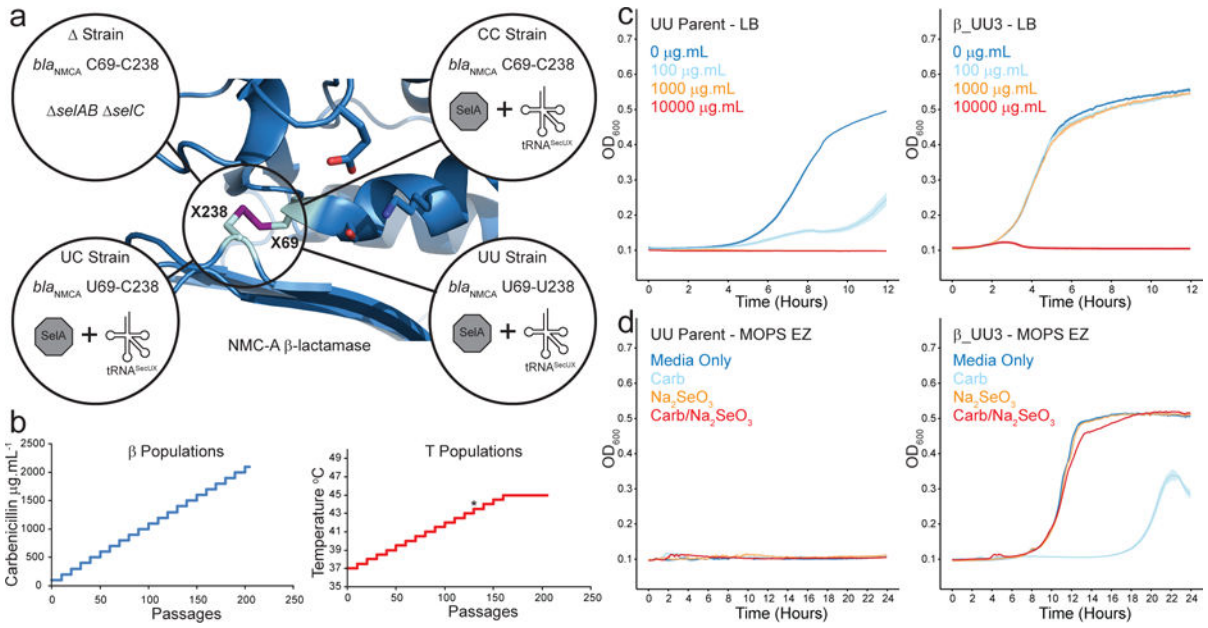

Figure 1. Evolution of selenocysteine dependent *E. coli* strains improves fitness

Carbenicillin concentrations are shown in the labels, and curves are plotted with a ribbon representing the mean ± S.E.M. where n = three independent biological replicates. (**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-sulfhydryl (UC) or diselenide (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 (~2500 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, details can be found in the **Online Methods**. (**c**) Growth rate, culture density and β-lactam resistance of an evolved population containing an essential diselenide bond (β_UU3) are substantially increased compared to the parent population (β_UU) at different concentrations of carbenicillin. (**d**) Metabolic defects which 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 Na_2SeO_3 were added at concentrations of 100 μg.mL and 1 μM respectively.

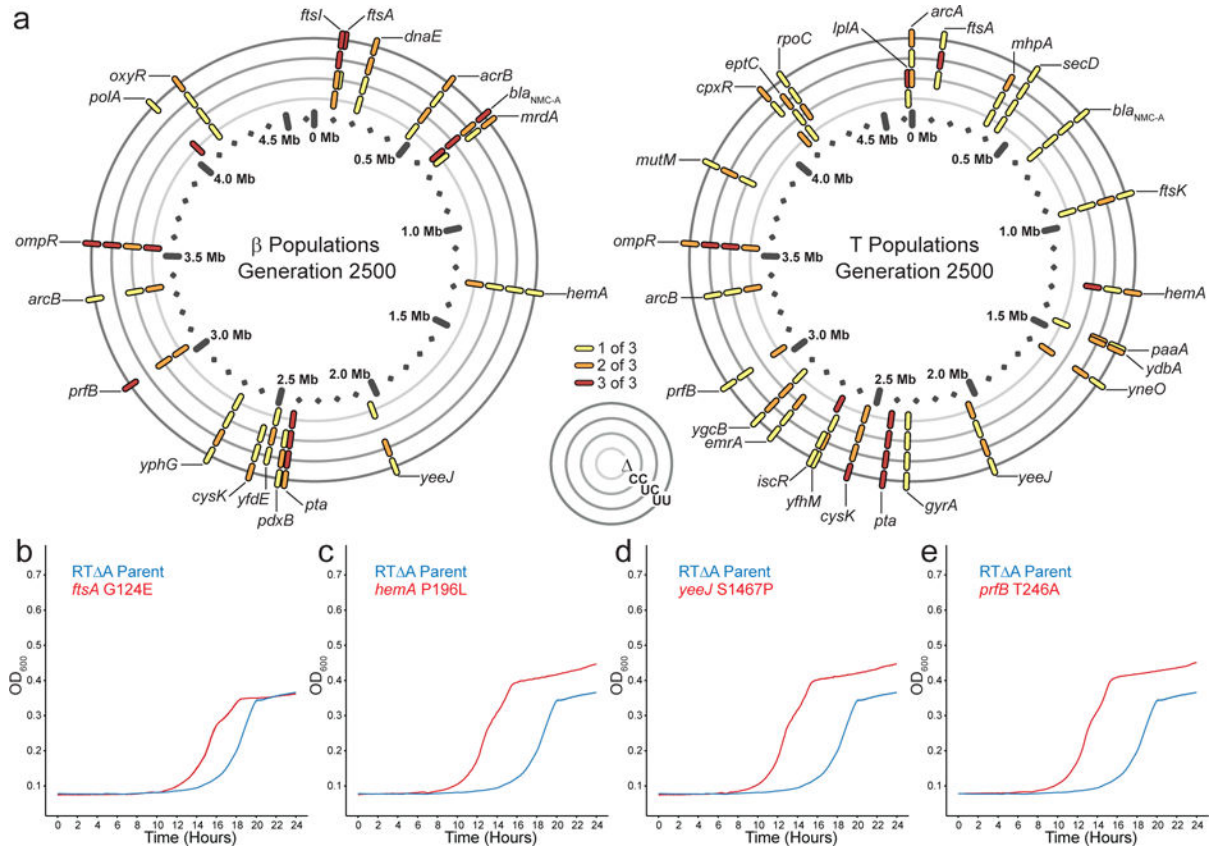

Figure 2. Genes mutated during continuous evolution

(**a**) Mutations were identified by whole genome sequencing of all evolved populations after 2500 generations. Genes that contained SNPs with >50% frequency in at least four of the twelve 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 Δ populations, the CC 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 SI Tables 2–3. (**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 ribbon representing the mean ± S.E.M. where n = three 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 **Supporting Note 1**.

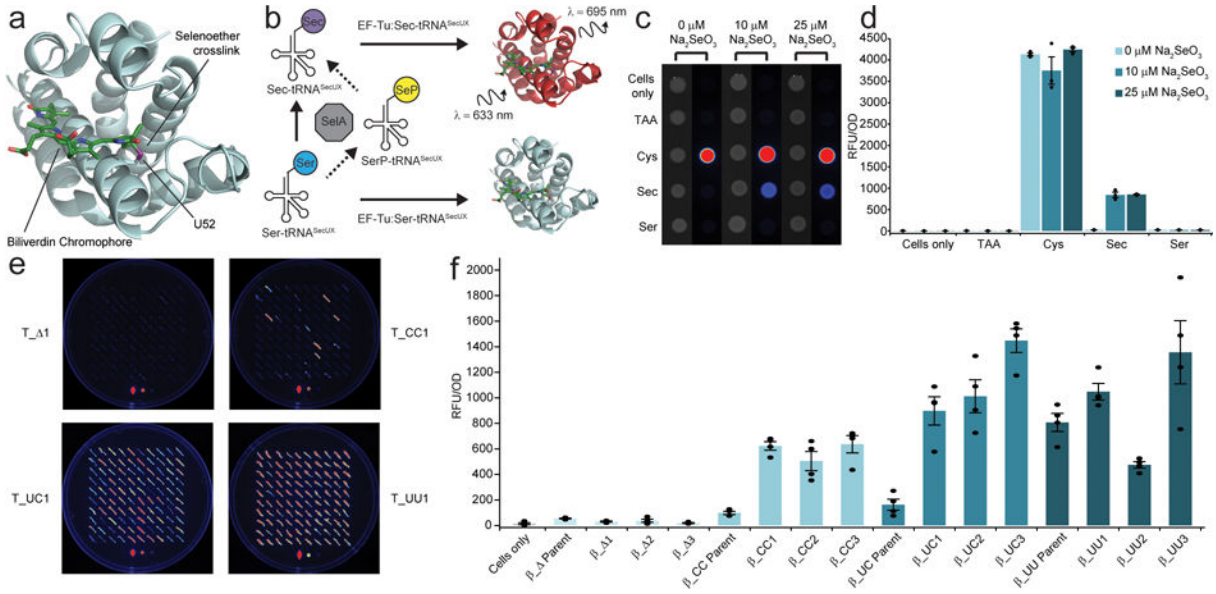

Figure 3. Selenocysteine incorporation in evolved strains

Incorporation of Sec was measured with our fluorescent reporter, seleno-smURFP. Graphs represents the average fluorescence of four independent clones normalized to cell density ± SEM. (**a**) Homology model of seleno-smURFP (U52) bound to biliverdin (modelled from PDB: 4po5) with the selenoether bond shown. (**b**) Schematic representation of discrimination between selenocysteine and serine by smURFP-X52. Both Ser-tRNA^{SecUX} and Sec-tRNA^{SecUX} are substrates for EF-Tu but only incorporation of selenocysteine at position 52 enables covalent tethering of the biliverdin chromophore which 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_2SeO_3 confirms fluorescence is dependent on selenocysteine. Plate is representative from six biologically independent grids. (**d**) Assay in liquid media shows greater than 20-fold difference between seleno-smURFP and smURFP-S52. The difference between the C52 and C52U variants is likely due to differences in protein abundance, since even in the absence of RF1, recoding UAG codons with selenocysteine is less efficient than canonical cysteine incorporation. (**e**) Evolved populations (T_Δ1, T_CC1, T_UC1 and T_UU1) have different rates of retention of the selenocysteine incorporation trait. The selenocysteine dependent UC and UU populations show nearly complete retention, in contrast, retention in CC populations was highly variable. Plates show biologically independent transformants (n = 100) and are representative from two experiments. (**f**) Populations evolved with β-lactam stress were screened for seleno-smURFP expression. The normalized fluorescence of most selenocysteine incorporating populations was several fold higher than their respective parental strains. The β_UU parental strain had comparatively high fluorescence, but extremely poor growth as shown in Figure 1c.

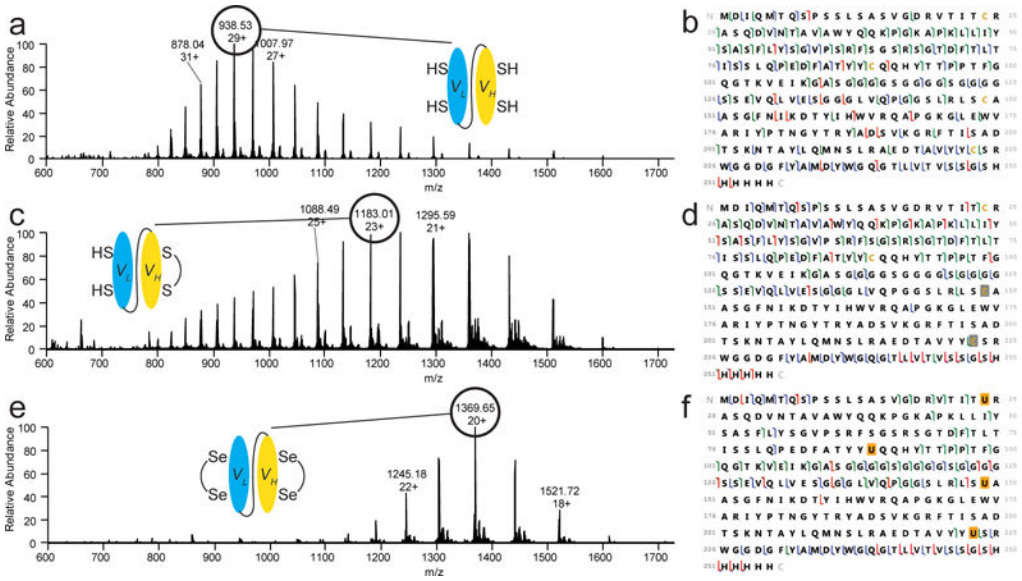

Figure 4. Recombinant seleno-scFvs containing diselenide bonds

(**a**) MS-trace of intact wild-type Herceptin-scFv produced under reducing conditions. (**b**) The 29+ charge state was analyzed by ultraviolet photodissociation (UVPD) and determined to represent a species containing no disulfide bonds. (**c**) MS analysis of intact wild-type Herceptin-scFv produced under oxidizing conditions. (**d**) The 23+ charge state corresponded to a species containing a single disulfide bond in the V_H domain. The presence of numerous higher charge states suggests a completely reduced species is also present. UVPD analysis of the 20+ charge state did not resolve a species containing both disulfide bonds. (**e**) MS analysis of intact seleno-Herceptin-scFv expressed in β_UU3-T7 cells. (**f**) UVPD analysis of the 20+ charge state identified a species containing a full complement of diselenide bonds. All spectra are representative from a minimum of two technical replicates.

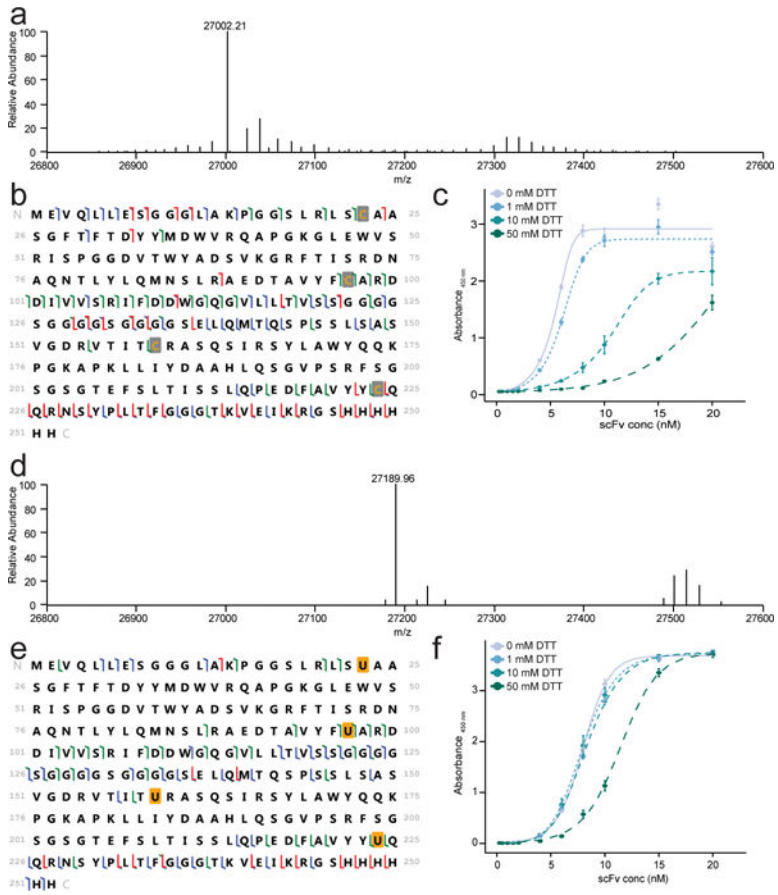

Figure 5. MS and ELISA curves for anti-RCA scFvs

(**a**) Deconvoluted MS of wild-type anti-RCA scFv. The experimentally determined monoisotopic mass (27002.21 Da) matches the theoretical masses (27002.17 Da) of wild-type anti-RCA scFv containing two disulfide bonds. (**b**) UVPD fragment map showing a lack of detectable fragments between C23 and C97 in the V_H domain, and C159 and C224 in the V_L domain, which confirms correct formation of both disulfide bonds. (**c**) ELISA of wild-type anti-RCA scFv treated with DTT. EC_{50} values (nM) of 5.47 ± 0.15, 6.16 ± 0.11, 10.8 ± 0.38 were determined for 0, 1, and 10 mM DTT, respectively. Points represent the mean ± the S.E.M. where n = four replicates. (**d**) Deconvoluted mass spectrum of seleno-anti-RCA. The experimentally determined monoisotopic mass (27193.94 Da) matches the theoretical mass (27193.96 Da) of seleno-anti-RCA containing two diselenide bonds. (**e**) UVPD fragment map showing a lack of detectable fragments between U23 and U97 in the V_H domain, and U159 and U224 in the V_L domain, which confirms correct formation of both diselenide bonds. (**f**) ELISA of seleno-anti-RCA scFv treated with DTT. EC_{50} values (nM) of 7.95 ± 0.06, 7.91 ± 0.10, 8.25 ± 0.05 and 11.4 ± 0.13 were determined for 0, 1, 10 and 50 mM DTT respectively. Points represent the mean ± the S.E.M. where n = four replicates. All spectra are representative from a minimum of two technical replicates.