*2234–2241* *Nucleic Acids Research, 2003, Vol. 31, No. 8*
*DOI: 10.1093/nar/gkg304*

# Revised *Escherichia coli* selenocysteine insertion requirements determined by *in vivo* screening of combinatorial libraries of SECIS variants

**Karen E. Sandman, Daniel F. Tardiff, Lori A. Neely and Christopher J. Noren***

New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

Received October 3, 2002; Revised January 19, 2003; Accepted February 12, 2003

## ABSTRACT

**To investigate the stringency of the *Escherichia coli* selenocysteine insertion sequence (SECIS) requirements, libraries of SECIS variants were screened via a novel method in which suppression of the selenocysteine (Sec) opal codon was coupled to bacteriophage plaque formation. The SECIS variant libraries were designed with a mostly paired lower stem, so that randomization could be focused on the upper stem and loop regions. We identified 19 functional non-native SECIS sequences that violated the expected pairing requirements for the SECIS upper stem. All of the SECIS variants were shown to permit Sec insertion in phage (by chemical modification of the Sec residue) and fused to lacZα (by β-galactosidase assay). The diminished pairing of the upper stem appears to be mitigated by the overall stem stability; a given upper stem variant has significantly higher readthrough in the context of a paired, rather than unpaired, lower stem. These results suggest an unexpected downstream sequence flexibility in prokaryotic selenoprotein expression.**

## INTRODUCTION

Since the determination that selenocysteine (Sec) is a genetically encoded, naturally occurring amino acid (1), there has been great progress in elucidating the mechanism of Sec insertion (2,3). In *Escherichia coli*, Sec is encoded by the UGA opal stop codon, and requires the constitutively expressed SelA, SelB, SelC and SelD gene products. Additionally, a downstream mRNA hairpin structure known as the Sec insertion sequence (SECIS) is required. The eukaryotic Sec insertion pathway differs from the prokaryotic mechanism (4) in both the sequence and position of the SECIS element. In *E.coli*, the 38-nt SECIS is positioned immediately downstream from the opal codon, whereas the eukaryotic SECIS can be located several hundred nucleotides downstream in the 3′ untranslated region of the gene. This difference in SECIS position, combined with the specific sequence requirements for the *E.coli* SECIS, presents significant barriers to selenoprotein expression in *E.coli*.

For many purposes, including heterologous expression, site-directed mutagenesis or site-specific protein labeling at a Sec residue, it would be desirable to generate selenoproteins in *E.coli.* Human thioredoxin reductase was successfully expressed in *E.coli* (5,6) using a near-native SECIS because Sec is the penultimate residue in the protein. In other cases, to engineer a Sec residue within a large protein sequence without having to change the amino acid sequence to fit the SECIS requirements, researchers have resorted to semi-synthetic methods such as native chemical ligation (7) or expressed protein ligation (8–10). To address the persistent challenges in selenoprotein expression, we decided to investigate the stringency of the *E.coli* SECIS sequence requirements.

Previous studies (11–15), summarized in Figure 1, have demonstrated that the SECIS can vary somewhat from that found in *E.coli* formate dehydrogenase (fdh). The lower stem, comprised of the nucleotides in the +4 to +14 and +32 to +41 positions, was found to have no pairing or sequence requirements, but there must be 11 nt between the UGA and the upper stem. It was also concluded that the upper stem, or nucleotides +15 to +19 and +28 to +31, had a required bulged U at position +17, and that the rest of the upper stem must be fully paired. Finally, the loop of nucleotides +20 to +27 has shown almost no flexibility in sequence requirements.

We recently (16,17) demonstrated that selenopeptides can be expressed as N-terminal fusions to M13 phage coat protein III. Because pIII is essential for phage infectivity and amplification, our system effectively couples phage production to suppression of the opal codon within these peptide-coding sequences. Since phage plaque formation is absolutely dependent on pIII expression, which is in turn dependent on opal suppression, this system provides a rapid, direct visual readout for Sec insertion requirements. Interestingly, this selection occurs *in vivo*, during phage morphogenesis, rather than *in vitro* as is more commonly the case with phage display experiments.

Using this system, we confirmed that the 9 nt downstream of the UGA can be fully randomized with no deleterious effect on

*To whom correspondence should be addressed. Tel: +1 978 927 5054; Fax: +1 978 921 1350; Email: noren@neb.com
Present addresses:
Karen E. Sandman, Vertex Pharmaceuticals, 130 Waverly Street, Cambridge, MA 02139, USA
Daniel F. Tardiff, Department of Biology, Brandeis University, Waltham, MA 02454, USA
Lori A. Neely, U.S. Genomics, 6H Gill Street, Woburn, MA 01801, USA

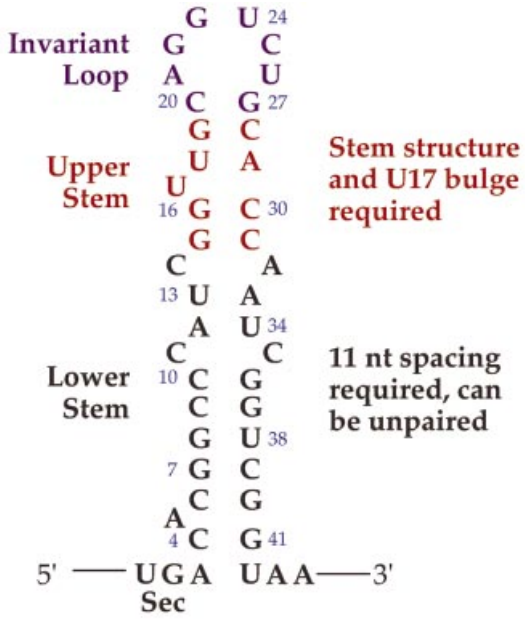


**Figure 1.** The *E.coli* fdh SECIS, with permissible mutations as reported by Liu *et al*. (11).

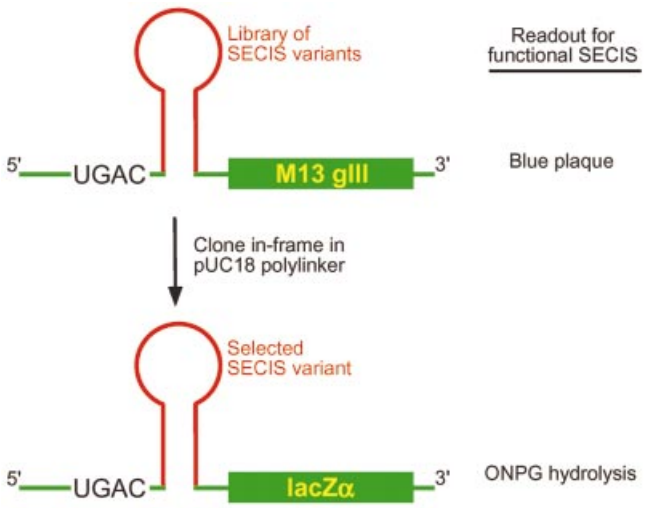


**Figure 2.** Experimental design. In-frame genetic fusion of a library of SECIS variants to an essential phage coat protein couples opal (UGA) codon suppression to phage production, providing a strong *in vivo* selection for functional SECIS clones. Sec insertion was verified for each selected clone by media selenium dependence and selenol-specific chemical modification. Relative Sec incorporation efficiencies were determined by fusing each selected SECIS variant upstream from the *lacZ* α-fragment, and assaying expressed β-galactosidase levels by measuring ONPG hydrolysis in cell extracts.

Sec insertion (16). Moreover, our results (17) demonstrated that the immediate downstream nucleotide from the UGA determines the identity of the inserted amino acid, with purines or CUG allowing for endogenous Trp-inserting opal suppression (18). A downstream U or C (not CUG) exclusively directs Sec insertion.

We have again employed phage display to study Sec insertion requirements by generating randomized libraries of SECIS mutants for cloning into M13 phage (Fig. 2). By fixing the immediate downstream nucleotides as CAC, we were able to ensure that any opal suppression, and thus phage production, would be due to Sec and not Trp insertion. Our method identified functional SECIS variants by sequencing the DNA of selenium-dependent plaque-forming phage. Sec insertion was then confirmed by cloning each SECIS variant upstream of lacZα and measuring opal suppression in β-galactosidase assays (Fig. 2).

## MATERIALS AND METHODS

All enzymes, western detection reagents, dNTPs and bacterial strains were from New England Biolabs (Beverly, MA) unless otherwise indicated. Sodium selenite was from Aldrich (Milwaukee, WI). Iodoacetamido-LC-biotin (I-Bt) was from Pierce (Rockford, IL). *o*-Nitrophenylgalactopyranoside (ONPG) was from Sigma (St Louis, MO). Oligodeoxyribonucleotides were synthesized by the Organic Synthesis Division of New England Biolabs using the phosphoramidite method.

### Cloning of *selABC* genes

*Escherichia coli* genomic DNA was prepared as described previously (19). PCR primers used for amplification of the *selC* open reading frame and the *selAB* operon were as follows: *selC* forward 5′-AGACGTGGATCCGGGGGATGTAGAAACTC-3′, *selC* reverse 5′-TATCTGGATCCACAGTATTACAAGGGG-3′, *selAB* forward, 5′-GGTGTGTCGACCGCCCACAC-3′, and *selAB* reverse, 5′-CACTGGTCGACTTAATTAATCATTTCCTTATTTTTCCGG-3′. The *selC* tRNA structural gene, together with 70 bp of 5′ flanking sequence and 52 bp of 3′ flanking sequence was PCR amplified using *Taq* DNA polymerase (Perkin-Elmer). The resulting 240 bp product was digested with *Bam*HI, gel-purified using β-agarase, and ligated to gel-purified, *Bam*HI linearized, pBR322 vector. Following electroporation into *E.coli* ER2738 cells, insert-containing clones were identified by restriction mapping, and the presence of *selC* was confirmed by DNA sequencing. The resulting plasmid will be subsequently referred to as pSelC.

The *selAB* operon, together with 186 bp of 5′ flanking sequence, was amplified from *E.coli* genomic DNA (Sigma) using Vent DNA polymerase, and the 3.5 kb product was gel-purified using β-agarase. The resulting blunt-ended PCR product was TOPO® cloned into the pCR4Blunt-Topo vector (Invitrogen). One microliter of the topoisomerase reaction was transformed into the chemically competent Topo One-Shot strain (Invitrogen). An insert-containing clone was then digested with *Apa*LI, *Hin*dIII and *Xba*I to liberate the 3.5 kb *selAB* operon from the Topo vector. The gel-purified fragment was then ligated to gel-purified *Hin*dIII and *Nhe*I-digested pSelC. The resulting vector, pSelABC, contains the *selC* and *selAB* genes under the transcriptional control of their endogenous promoters.

### Cloning of SECIS inserts

Phage libraries and single clones were constructed in the pentavalent phage vector M13KE (20) as described previously (17). Briefly, SECIS inserts were prepared by annealing the

insert extension primer 5′-CATGCCCGGGTACCTTTCT-ATTCTC-3′ (NEB product E8101S) to each of the following oligonucleotides, extending with Klenow fragment, digesting with *Acc*65I and *Eag*I, and ligating the gel-purified product to *Acc*65I/*Eag*I-digested M13KE. For each oligonucleotide, *Acc*65I and *Eag*I restriction sites are shown in bold, and the TGA or TGG codon and immediate downstream nucleotide in italics (antisense strand shown). SECIS Library I: underlined positions were randomized with 85% of the native nucleotide at each position and 5% each of the other three (e.g. T = 85% T, 5% A, 5% C, 5% G). 5′-CATGTTT**CGGCCG**A-TTGGTGCAGACCTGCAACCGATGGGCCGT*GTCA*GAC-ACGAGCGCTAGAGTGAGAATAGAAA**GGTACC**CGGG-3′. SECIS Library II: underlined positions were randomized with 70% of the native nucleotide and 10% each of the other three. 5′-CATGTTT**CGGCCG**ATTGGTGCAGAC-CTGCAACCGATGGGCCGT*GTCA*GACACGAGCGCTAG-AGTGAGAATAGAAA**GGTACC**CGGG-3′. TGAGwt: 5′-CATGTTT**CGGCCG**ATTGGTGCAGACCTGCAACCGA-TGGGCCGT*CTCA*GACACGAGCGCTAGAGTGAGAAT-AGAAA**GGTACC**CGGG-3′. TGACbad: 5′-CATGTTT-**CGGCCG**ATTGGTCCAGACCTCCATCCGATGGGCCG-T*GTCA*GACACGAGCGCTAGAGTGAGAATAGAAA**GG-TACC**CGGG-3′. TGAGbad: 5′-CATGTTT**CGGCCG**ATT-GGTCCAGACCTCCATCCGATGGGCCGT*CTCA*GACAC-GAGCGCTAGAGTGAGAATAGAAA**GGTACC**CGGG-3′. TGGCwt: 5′-CATGTTT**CGGCCG**ATTGGTGCAGACCTG-CAACCGATGGGCCGT*GCCA*GACACGAGCGCTAGAG-TGAGAATAGAAA**GGTACC**CGGG-3′. SV8*: 5′-CATGT-TT**CGGCCG**ATTTGTGCAGACCTGCAACTGAATGAC-TCC*ATCA*GACACGAGCGCTAGAGTGAGAATAGAAA-**GGTACC**CGGG-3′. SV15*: 5′-CATGTTT**CGGCCG**ATT-GGTGCAGACCTGCACACGTATGACTCC*ATCA*GACAC-GAGCGCTAGAGTGAGAATAGAAA**GGTACC**CGGG-3′.

Ligations were electroporated into *E.coli* ER2738 [F′ *proA*$^{+}$*B*$^{+}$ *lacI*$^{q}$ Δ(*lacZ*)*M15 zzf::Tn10* (*Tet*$^{R}$)/*fhuA2 glnV thi-1* Δ(*lac-proAB*) Δ(*hsdS-mcrB*)*5*] harboring pSelABC, and phage was produced as described previously (17). All phage cultures and plating media contained 100 μg/ml ampicillin and 2 μM sodium selenite unless otherwise indicated. Individual clones were amplified, titered and sequenced as described previously (17).

### Biotin modification and immunoblotting

Stocks of individual phage clones were prepared by amplification in ER2738/pSelABC. The phage supernatants were precipitated twice with one-sixth volume of 20% polyethylene glycol-8000/2.5 M sodium chloride and resuspended in 50 mM Tris–HCl pH 7.4, 150 mM sodium chloride. Final phage concentration was determined by plating serial dilutions of phage with a mid-log ER2738/pSelABC culture on LB agar plates with 100 μg/ml ampicillin, 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 0.1 mM Xgal. Sequences of amplified phage were confirmed by automated DNA sequencing. Biotin modification and immunoblotting were carried out as described (17).

### Cloning selected SECIS variants in pUC18

Individual SECIS variants were amplified by PCR from ~100 ng of single stranded phage DNA, or from M13KE ligation reactions if phage was not isolated, using the duplex extension primer 5′-CATGCCCGGGTACCTTTCTATT-CTC-3′ (NEB product E8101S) and the *Bam*HI primer 5′-GACGTTAGTAAATGAATTTTCTGGATCCGATTTTGC-3′. The amplified inserts were digested with *Acc*65I and *Bam*HI, gel-purified, and ligated into *Acc*65I/*Bam*HI-digested pUC18. The heat-killed ligation products were electroporated into ER2738 and plated on LB agar + ampicillin. Individual clones were analyzed by restriction mapping and DNA sequencing.

### β-Galactosidase assays for TGA suppression

Freshly transformed *E.coli* ER2738 colonies were grown aerobically at 37°C in LB medium prepared with Milli-Q deionized water, with or without 2 μM sodium selenite. When an $OD_{600}$ of ~0.5 was reached, 0.3 μM IPTG was added and the culture was shaken at 37°C for an additional 2 h. Cells (1.5 ml) were then centrifuged, the supernatant discarded, and the pellet stored overnight at –20°C. The cell pellet was resuspended in 1 ml Z buffer (100 mM sodium phosphate pH 7.0, 10 mM potassium chloride, 1 mM magnesium sulfate, 50 mM 2-mercaptoethanol). β-Galactosidase activity was assayed by adding 200 μl of 13 mM ONPG, incubating at 28°C for 1 h, and stopping the reaction with 0.5 ml of 1 M $Na_2CO_3$. Reactions were then centrifuged to pellet cell debris. An $A_{420}$ was obtained from the supernatant and β-galactosidase activities determined as described previously (21).

### Computer modeling of SECIS variants

Secondary RNA structures of SECIS variants were modeled using the online RNA folding program mfold version 3.1 (22,23). Nine nucleotides upstream, and 50 nt downstream, of the Sec UGA were included in structure determination and free energy calculations. No constraints were placed on the folding of the SECIS variants.

## RESULTS

### Phage production from non-native SECIS elements

SECIS Library I was designed with an 85:5:5:5 mutagenesis scheme for the nucleotides in positions +13 to +33, which span the loop and upper stem, as well as 4 nt from the lower stem. Preliminary attempts to isolate novel SECIS variants using ER2738 *E.coli* with endogenous levels of SelA, SelB and SelC resulted in few TGA sequences with a non-native downstream SECIS, probably because of the growth advantage afforded to phage containing point mutations in the TGA codon. All of the presently reported phage experiments were performed in ER2738 transformed with pSelABC, which overexpresses three of the four genes involved in Sec incorporation, and greatly enhances the genetic stability of phage containing the TGA codon (L.A.Neely, K.E.Sandman and C.J.Noren, unpublished results). From Library I, sequencing data was obtained for 39 phage plaques. Of the resulting sequences, 24 contained the native fdh SECIS, eight had mutations in the lower stem and seven included upper stem mutations.

Because there were no mutations discovered within the eight-membered loop, the loop was fixed as wild-type (wt) in Library II, with a greater (70:10:10:10) randomization of the stem. Forty-four phage plaques were sequenced from Library II; three clones had a point mutation within the TGA codon. Of

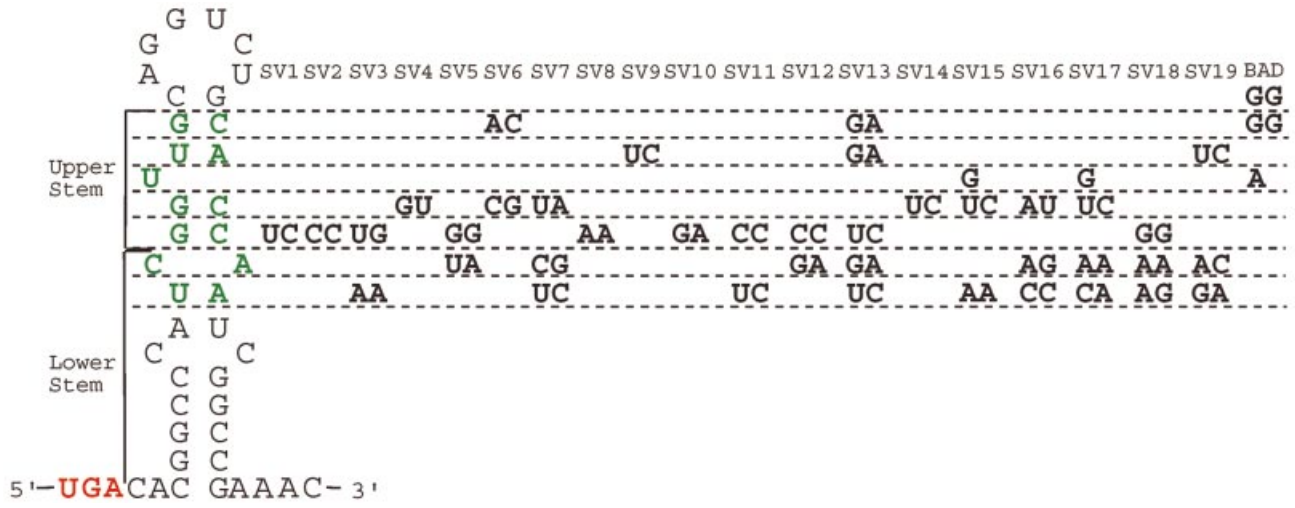


**Figure 3.** Sequences of functional SECIS variants and the 'bad' non-functional SECIS studied in this work. The stem positions randomized from the wt fdh SECIS are indicated in green, with the upper stem region bracketed. Sequence variations are shown in the position of the corresponding nucleotide in the wt fdh SECIS, although the pairings may differ from the native structure.

the 41 clones with an unmutated TGA codon, 10 had the wt SECIS, 12 had mutations in the lower stem and 19 carried mutations in the upper stem. All of the phage bearing upper stem mutations (19 unique sequences from both libraries), and one clone of wt fdh SECIS (TGACwt), were re-amplified for further study. Phage with the wt fdh SECIS and all 19 SECIS variants formed plaques only when plated with supplemental sodium selenite, while the phage plaque diameter of the TGAGwt SECIS was selenium-enhanced (17).

In addition, several control phage samples were prepared using synthetic oligonucleotide inserts. The TGGCwt contained the wt SECIS but a Trp TGG codon rather than the Sec TGA. The TGAGwt contained the fdh SECIS, but with the immediate downstream nucleotide changed from C to G to allow for Trp-inserting opal suppression (17,18). TGACbad and TGAGbad sequences had either downstream nucleotide followed by a SECIS mutant designed to be non-functional based on results reported previously (12). As expected, attempts to clone the TGACbad and TGAGbad inserts into phage were unsuccessful; resultant phage plaques contained a variety of mutations, frameshifts and deletions. Figure 3 shows the sequences of all SECIS variants SV1–SV19 studied in this work, as well as the non-functional (bad) SECIS.

### Confirmation of Sec insertion by chemical modification of phage

To confirm that the SV1–SV19 phage clones were indeed inserting Sec in the N-terminal pIII fusions, each clone was assayed for characteristic Sec reactivity. Because the pKa of Sec is lower than that of Cys (5.2 versus 8.1), Sec is preferentially modified by nucleophilic substitution at low pH (16). Each clone was amplified in ER2738/pSelABC, and a doubly PEG-precipitated phage stock of known titer was prepared. The phage DNA was sequenced again to ensure that the clones were stable upon repeated amplification. Control sequences prepared in identical fashion included the TGACwt sequence as a positive control, and a clone containing a single cysteine (Cys1) with nearly identical sequence to the TGACwt (SARVLCNH versus SARVSecHGP) as a negative control. Additionally, a library of random heptapeptides displayed on phage (PhD-7) was included to further rule out modification elsewhere on the phage. Equivalent amounts of each phage clone were treated with I-Bt at pH 2.5, denatured and subjected to SDS–PAGE. The gels were blotted and probed with an anti-biotin antibody, which recognizes biotin-modified pIII, and an anti-pIII monoclonal, which measures pIII independent of modification, to control for differences in phage loading. Figure 4 shows immunoblots of the SECIS variants and controls, after reaction with I-Bt. All of the clones showed specific biotinylation compared with the single cysteine (Cys1) and random peptide library (PhD7) controls.

### Quantitation of opal suppression by a reporter gene assay

To quantitate opal suppression, the SECIS variants and control sequences were cloned as N-terminal fusions to the α fragment of *lacZ* in pUC18. The resulting clones were sequenced, and Miller assays were performed in the α-complementing *E.coli* strain ER2738 (without pSelABC) to measure opal suppression. The results, summarized in Figure 5, show that all of the SECIS variants, as well as the TGACwt control, permitted UGA readthrough in a selenium-dependent manner, i.e. the error bars for experiments with and without media selenium do not overlap. In contrast, the selenium independence of readthrough of the TGGCwt control is consistent with tryptophan insertion. Readthrough of the TGAGwt control demonstrates selenium dependence, but suppression levels both in the presence and absence of selenium are elevated relative to the TGACwt. This is consistent with our earlier observation (17) that a purine immediately downstream from the UGA codon permits selenium-independent tryptophan insertion.

### Investigation of SECIS variants with a non-native lower stem

To determine whether the SECIS variants identified in this work were only functional in the context of the wt fdh SECIS

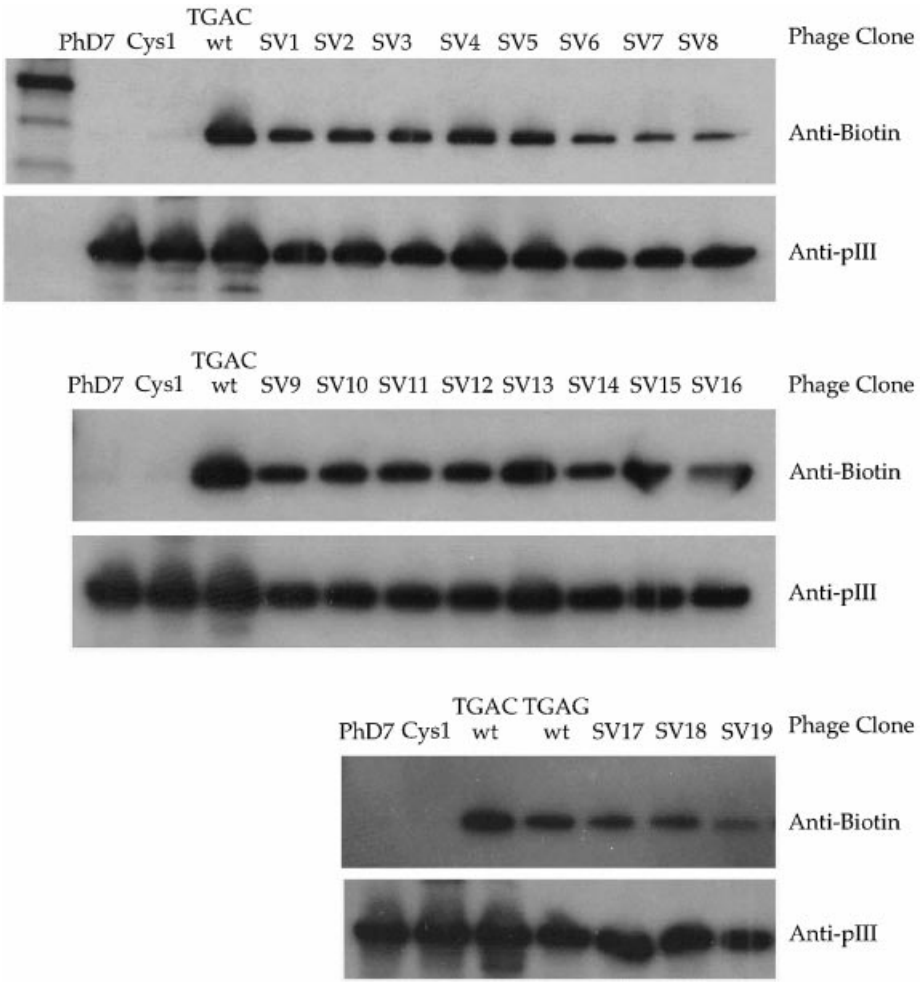


**Figure 4.** Immunoblots showing I-Bt modification of Sec residues on phage coat protein III. Reactions were carried out at pH 2.5. As controls, phage displaying a single cysteine (Cys-1; displayed peptide is SARVLCNH) and an aliquot of a 7mer peptide display library (PhD-7) were treated with I-Bt. The samples were divided in half and run on two gels to probe with anti-biotin and anti-pIII antibodies to normalize for variations in phage titering.

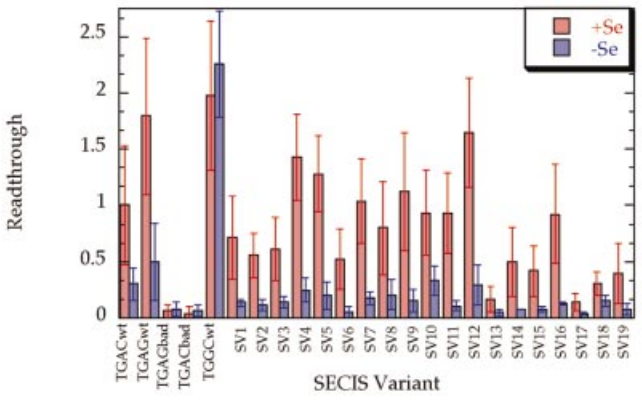


**Figure 5.** Quantitation of selenium-dependent opal codon readthrough for selected SECIS variants and control sequences by lacZ assay. For each clone, the TGA-SECIS was cloned upstream of lacZα in pUC18, and β-galactosidase assays were performed with or without 2 μM supplemental sodium selenite. Each data point represents six independent determinations. Error bars represent ± 1 standard deviation. All readthrough values are normalized relative to TGACwt (the wild-type fdh SECIS), which was set as 1.0.

lower stem, two clones, SV8 and SV15, were studied in the context of a different lower stem. In the two clones prepared, SV8* and SV15*, the 9 nt downstream of the TGA codon were TGGA**GT**C**A**T, with only the two nucleotides in bold conserved from the wt sequence, so as to abolish most of the base pairing interactions in the lower stem. Figure 6 shows that both sequences permitted selenium-dependent readthrough of the opal codon in β-galactosidase assays, although to a lesser extent (<50%) than the corresponding clones with the wt lower stem.

## Molecular modeling of SECIS variants

The SECIS variants were modeled in the mfold RNA folding program, and free energies determined from the predicted structures. The ΔG values for the variants ranged from –21.3 to –12.8 kcal/mol, compared with –17.8 kcal/mol for the wt SECIS. There was no correlation between predicted free energy and UGA readthrough (data not shown), possibly because the modeling did not take into account differences in structure and/or stability caused by binding of SelB to the SECIS.

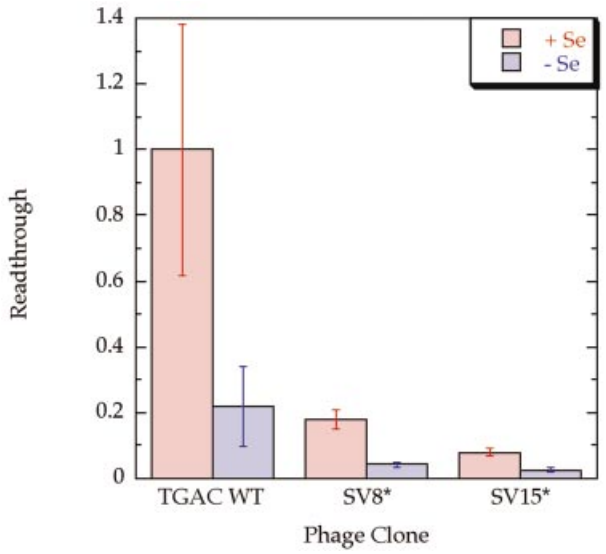


**Figure 6.** Quantitation of selenium-dependent opal codon readthrough for SV8* and SV15*, SECIS variants with a non-native lower stem, compared with the TGACwt SECIS. The TGA-SECIS was cloned upstream of lacZα in pUC18, and β-galactosidase assays were performed with or without 2 μM supplemental sodium selenite. Each data point represents three independent determinations. Error bars represent ± 1 standard deviation. Readthrough values are normalized relative to TGACwt (the wild-type fdh SECIS), which was set as 1.0.

## DISCUSSION

Since the discovery that prokaryotic Sec insertion is dependent on a downstream mRNA sequence (14) that is recognized by SelB (24), several groups have attempted to define the recognition elements within the SECIS, to further elucidate the mechanism of Sec insertion and to minimize the sequence constraints. In the first work to substantially minimize the SECIS requirements relative to the wt sequence, Engelberg-Kulka and coworkers (11,13) investigated a variety of SECIS variants cloned upstream of β-galactosidase and measured readthrough and radioactive selenium incorporation. A critical determination was that the lower stem (positions +4 to +14 and +32 to +41) could be completely unpaired with no deleterious effect on opal suppression. A series of upper stem variants was prepared, in the context of a fully unpaired lower stem, and it was concluded that the upper stem must be fully paired.

Using the same selenopeptide–pIII fusion vector as in the present work, our earlier phage display studies (16,17) confirmed that the lower stem nucleotides in positions +4 to +12 could be randomized, in the context of the wt upper stem and loop, with efficient Sec insertion. The randomization scheme in both studies was (NNK) for each codon downstream of UGA, where N is any nucleotide and K is G or T. Fortuitously, none of the positions randomized as K were G or T in the wt paired lower stem, so our SECIS variants were inherently biased against a paired lower stem. Moreover, in the present study we randomized the lower stem nucleotides +13, +14, +32 and +33 and obtained functional SECIS variants with almost every possible combination of nucleotides in these positions (data not shown). Our data from screening combinatorial libraries of lower stem sequences demonstrates that the lower stem can indeed be fully unpaired, and that this region is tolerant to highly randomized sequences.

The goal of the present work was to employ combinatorial methods to determine the requirements of the upper stem and loop regions of the SECIS. Previous studies of the upper stem and loop (11,12) investigated individual SECIS variants, rather than performing a thorough sampling of sequence space. In one study (25), a combinatorial aptamer approach was employed to identify upper stem–loop sequences that bound SelB *in vitro*. Although that work did underscore the importance of the upper stem bulged uridine in SelB recognition, the aptamers were screened *in vitro* for SelB binding, and many SelB binding sequences were not functional *in vivo* for Sec insertion. Our goal was to perform library screening *in vivo* (i.e. during phage morphogenesis) in order to identify functional SECIS variants.

Our findings revealed an unexpected degree of flexibility in the SECIS requirements (Fig. 3). The upper stem was considerably more tolerant to mutations than predicted. The positions closest to the loop were somewhat fixed, with only two functional variants (SV6 and SV13) arising in the pair closest to the loop (+19 and +28) and three (SV9, SV13 and SV19) in the next pair down (+18 and +29). In contrast the lower two pairs in the upper stem (+15 and +31, and +16 and +30) had a wide range of permissible mutations. SV2 was especially relevant, because its G15C mutation had been shown previously to abolish UGA readthrough in the context of a completely unpaired lower stem (11). To further investigate the interaction between a non-native upper stem and the pairing of the lower stem, SECIS variants SV8 and SV15 were prepared with a mostly unpaired lower stem. The resulting clones, SV8* and SV15*, permitted selenium-dependent UGA readthrough, but less than half as much as their parent clones (Fig. 6). Similar results were observed using a SECIS variant selected from RNA aptamers for its SelB-binding ability (25): a clone with a non-native upper stem revealed no readthrough with a non-native lower stem, but suppression rose to 17% when the upper stem variant was fused to the wt lower stem. These results together suggest that the SECIS requirements are not limited to the upper stem. Perhaps some minimal level of stem pairing is required in order to present the required elements for SelB recognition, and this pairing can be distributed over the entire length of the stem, rather than just the upper region.

Although Heider *et al.* (12) observed up to 22% readthrough with certain loop mutations, we did not observe any permissible loop variants using our randomization scheme. The loop results might have been different in the context of a completely wt stem; the combined effects of upper stem and loop mutations presumably diminished the chances of identifying functional loop variants. In two cases (SV15 and SV17), the 'required bulged' U17 was mutated to a G; in both of these cases, however, G16 was mutated to a U, and U18 was unmutated, so that there were two uridines, either of which could possibly function as the recognition element for SelB. Further experiments are necessary to fully delineate the importance of the bulged U in the upper stem. For many of the SECIS variants studied, molecular modeling of the RNA sequences did not portray the bulged uridine (data not shown), but it is likely that the protein–RNA contacts afforded by SelB binding would stabilize the appropriate binding conformation.

Although much of the Sec insertion mechanism has been defined, the nature of the SelB–SECIS recognition is not fully elucidated. It has been shown by binding and toeprinting studies that the upper stem and loop region of the SECIS is contacted by SelB (24,26,27). In the absence of structural data showing the SelB–SECIS complex, it is difficult to predict the exact recognition elements in the SECIS. Given the permissiveness of the entire stem, it seems likely that SelB principally recognizes the loop and bulged U, but probably does not recognize sequence or structural elements within the stem. The stem probably does, however, need to be stable enough to present the appropriate bulged U and loop for SelB recognition.

Based on our earlier results (17), we expected the TGAGwt SECIS sequence to allow Trp-inserting opal suppression in addition to Sec insertion. Although the β-galactosidase assay did reveal a somewhat elevated level of readthrough in the absence of supplemental selenite, and overall readthrough higher than TGACwt SECIS, this clone still had substantial selenium enhancement of both lacZα readthrough and phage plaque diameter. This result is consistent with the report that the Arg and Val codons immediately upstream of Sec in the native fdh sequence, together with a TGAC sequence, help to prevent readthrough when selenium concentration is low (13,28). In our previous phage display study we randomized the four codons upstream of Sec, and found the resulting TGAG clones to have selenium-independent UGA readthrough as measured by phage production and plaque diameter. In the present work the upstream codons were fixed as the native sequence (Arg-Val), which partially counteracted the tendency of the downstream G to direct Trp-inserting opal suppression in the absence of selenium.

Overexpression of the SelABC genes has been shown to enhance selenoprotein expression in the context of a native SECIS (5,29). Our early attempts at screening for functional SECIS variants were performed in *E.coli* ER2738, without the pSelABC plasmid. These efforts were hampered by the instability of many Sec-inserting clones; phage carrying mutations in the UGA codon had a significant growth advantage over UGA-containing clones. The overexpression of the SelABC genes stabilized the clones during amplification so that we could identify SECIS variants. To ascertain that these variants were functional in the normal *E.coli* background, the β-galactosidase assays shown in Figures 5 and 6 were performed in ER2738 without pSelABC. It is expected that overexpression of the SelABC genes would enhance selenoprotein expression in the context of any of our SECIS variants. This would be especially useful if protein sequence requirements dictated the use of a minimally efficient SECIS variant.

In addition to the growing number of approaches for obtaining selenoproteins [e.g. expression in cultured *Chlamydomonas reinhardtii* plant cells (30,31)], expression in *E.coli* may be a viable option for more selenoprotein sequences than previously predicted. Although our initial attempts revealed no permissible variations in the SECIS loop sequence, the upper stem requirements appear to be substantially more flexible than had been reported. It seems plausible that the 'rules' of SECIS structure are not as easily defined as thought previously: while the full-length SECIS requires some minimum level of hairpin stability, this stability can evidently be distributed along the length of the stem structure. It may be possible to express selenoproteins in *E.coli* by simply overexpressing the SelABC genes and designing a SECIS sequence that allows expression of the native downstream amino acid sequence of the protein, but which preserves the base-pairing and length of the stem region and as much of the sequence of the loop region as possible.

## NOTE ADDED IN PROOF

Following submission of our manuscript, a report appeared in which the authors used a combination of rational mutagenesis and NMR spectroscopy to investigate the SelB–SECIS interaction (32). Using techniques that perfectly complemented our combinatorial approach, the authors arrived at nearly identical conclusions regarding the invariance of the loop region and sequence flexibility of the upper stem in the vicinity of the bulged U.

## ACKNOWLEDGEMENTS

The authors thank Laurie Mazzola and Jennifer Ware of the NEB sequencing lab for extensive technical assistance and Donald Comb for support of this research. We thank Corinna Tuckey, Paul Riggs and Jamie Williamson for helpful discussions, and Karen Noren, Corinna Tuckey and Kamini Hingorani for critical reading of the manuscript.

## REFERENCES

1. Zinoni,F., Birkmann,A., Leinfelder,W. and Böck,A. (1987) Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proc. Natl Acad. Sci. USA*, **84**, 3156–3160.
2. Böck,A., Forchhammer,K., Heider,J., Leinfelder,W., Sawers,G., Veprek,B. and Zinoni,F. (1991) Selenocysteine: the 21st amino acid. *Mol. Microbiol.*, **5**, 515–520.
3. Stadtman,T.C. (1996) Selenocysteine. *Annu. Rev. Biochem.*, **65**, 83–100.
4. Berry,M.J., Martin,G.W., Tujebajeva,R., Grundner-Culemann,E., Mansell,J.B., Morozova,N. and Harney,J.W. (2002) Selenocysteine insertion sequence element characterization and selenoprotein expression. *Methods Enzymol.*, **347**, 17–24.
5. Arner,E.S., Sarioglu,H., Lottspeich,F., Holmgren,A. and Böck,A. (1999) 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.*, **292**, 1003–1016.
6. Bar-Noy,S., Gorlatov,S.N. and Stadtman,T.C. (2001) Overexpression of wild type and SeCys/Cys mutant of human thioredoxin reductase in *E. coli*: the role of selenocysteine in the catalytic activity. *Free Radic. Biol. Med.*, **30**, 51–61.
7. Gieselman,M.D., Xie,L. and van der Donk,W.A. (2001) Synthesis of a selenocysteine-containing peptide by native chemical ligation. *Org. Lett.*, **3**, 1331–1334.
8. Hondal,R.J., Nilsson,B.L. and Raines,R.T. (2001) Selenocysteine in native chemical ligation and expressed protein ligation. *J. Am. Chem. Soc.*, **123**, 5140–5141.
9. Hondal,R.J. and Raines,R.T. (2002) Semisynthesis of proteins containing selenocysteine. *Methods Enzymol.*, **347**, 70–83.
10. Berry,S.M., Gieselman,M.D., Nilges,M.J., van Der Donk,W.A. and Lu,Y. (2002) An engineered azurin variant containing a selenocysteine copper ligand. *J. Am. Chem. Soc.*, **124**, 2084–2085.
11. Liu,Z., Reches,M., Groisman,I. and Engelberg-Kulka,H. (1998) The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in *Escherichia coli*. *Nucleic Acids Res.*, **26**, 896–902.
12. Heider,J., Baron,C. and Böck,A. (1992) Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. *EMBO J.*, **11**, 3759–3766.

13. Engelberg-Kulka,H., Liu,Z., Li,C. and Reches,M. (2001) An extended *Escherichia coli* 'selenocysteine insertion sequence' (SECIS) as a multifunctional RNA structure. *Biofactors*, **14**, 61–68.
14. Zinoni,F., Heider,J. and Böck,A. (1990) Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc. Natl Acad. Sci. USA*, **87**, 4660–4664.
15. Chen,G.-F.T., Fang,L. and Inouye,M. (1993) Effect of the relative position of the UGA codon to the unique secondary structure in the *fdhF* mRNA on its decoding by selenocysteinyl tRNA in *Escherichia coli*. *J. Biol. Chem.*, **268**, 23128–23131.
16. Sandman,K.E., Benner,J.S. and Noren,C.J. (2000) Phage display of selenopeptides. *J. Am. Chem. Soc.*, **122**, 960–961.
17. Sandman,K.E. and Noren,C.J. (2000) The efficiency of *Escherichia coli* selenocysteine insertion is influenced by the immediate downstream nucleotide. *Nucleic Acids Res.*, **28**, 755–761.
18. Miller,J.H. and Albertini,A.M. (1983) Effects of surrounding sequence on the suppression of nonsense codons. *J. Mol. Biol.*, **164**, 59–71.
19. Wilson,K. (1994) Preparation of genomic DNA from bacteria. In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology*. John Wiley & Sons, Vol. 1, pp. 2.4.1–2.4.5.
20. Noren,K.A. and Noren,C.J. (2001) Construction of high-complexity combinatorial phage display peptide libraries. *Methods*, **23**, 169–178.
21. Miller,J.F. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
23. Zuker,M., Mathews,D.H. and Turner,D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In Barciszewski,J. and Clark,B.F.C. (eds), *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 11–43.
24. Baron,C., Heider,J. and Böck,A. (1993) Interaction of translation factor SELB with the formate dehydrogenase H selenopolypeptide mRNA. *Proc. Natl Acad. Sci. USA*, **90**, 4181–4185.
25. Klug,S.J., Hüttenhofer,A., Kromayer,M. and Famulok,M. (1997) *In vitro* and *in vivo* characterization of novel mRNA motifs that bind special elongation factor SelB. *Proc. Natl Acad. Sci. USA*, **94**, 6676–6681.
26. Ringquist,S., Schneider,D., Gibson,T., Baron,C., Böck,A. and Gold,L. (1994) Recognition of the mRNA selenocysteine insertion sequence by the specialized translational elongation factor SELB. *Genes Dev.*, **8**, 376–385.
27. Kromayer,M., Wilting,R., Tormay,P. and Böck,A. (1996) Domain structure of the prokaryotic selenocysteine-specific elongation factor SelB. *J. Mol. Biol.*, **262**, 413–420.
28. Liu,Z., Reches,M. and Engelberg-Kulka,H. (1999) A sequence in the *Escherichia coli* fdhF 'Selenocysteine insertion Sequence' (SECIS) operates in the absence of selenium. *J. Mol. Biol.*, **294**, 1073–1086.
29. Arner,E.S. (2002) Recombinant expression of mammalian selenocysteine-containing thioredoxin reductase and other selenoproteins in *Escherichia coli*. *Methods Enzymol.*, **347**, 226–235.
30. Novoselov,S.V., Rao,M., Onoshko,N.V., Zhi,H., Kryukov,G.V., Xiang,Y., Weeks,D.P., Hatfield,D.L. and Gladyshev,V.N. (2002) Selenoproteins and selenocysteine insertion system in the model plant cell system, *Chlamydomonas reinhardtii*. *EMBO J.*, **21**, 3681–3693.
31. Fu,L.H., Wang,X.F., Eyal,Y., She,Y.M., Donald,L.J., Standing,K.G. and Ben-Hayyim,G. (2002) A selenoprotein in the plant kingdom. Mass spectrometry confirms that an opal codon (UGA) encodes selenocysteine in *Chlamydomonas reinhardtii* gluththione peroxidase. *J. Biol. Chem.*, **277**, 25983–25991.
32. Fourmy,D., Guittet,E. and Yoshizawa,S. (2002) Structure of prokaryotic SECIS mRNA hairpin and its interaction with elongation factor SelB. *J. Mol. Biol.*, **324**, 137–150.