896–902 *Nucleic Acids Research, 1998, Vol. 26, No. 4*

The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in *Escherichia coli*

Zesheng Liu, Myriam Reches, Irina Groisman and Hanna Engelberg-Kulka*

Department of Molecular Biology, Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel

Received November 28, 1997; Revised and Accepted January 2, 1998

ABSTRACT

The UGA codon, usually a stop codon, can also direct the incorporation into a protein of the modified amino acid selenocysteine. This UGA decoding process requires a *cis*-acting mRNA element called 'selenocysteine insertion sequence' (SECIS) that can form a stem–loop structure. In *Escherichia coli* the SECIS of the selenoprotein formate dehydrogenase (FdhH) mRNA has been previously described to consist of at least 40 nucleotides following the UGA codon. Here we determined the nature of the minimal SECIS required for the *in vivo* UGA-directed selenocysteine incorporation in *E.coli*. Our study is based on extensive mutational analysis of the *fdhF* SECIS DNA located in a *lac′Z* fusion. We found that the whole stem–loop RNA structure of the *E.coli fdhF* SECIS previously described is not required for the UGA-directed selenocysteine incorporation *in vivo*. Rather, only its upper stem–loop structure of 17 nucleotides is necessary on the condition that it is located in a proper distance (11 nucleotides) from the UGA codon. Based on these observations, we present a new model for the minimal *E.coli* SECIS.

INTRODUCTION

The results of recent research in both prokaryotic and eukaryotic systems has permitted the establishment of an extended genetic code in which UGA, usually a stop codon, can also direct the incorporation of the modified amino acid selenocysteine (for reviews see 1–4). Proteins containing selenocysteine residues are called selenoproteins and these are mostly oxido-reductases (for reviews see 5,6). That selenocysteine residues have been shown to play an essential role in the activity of selenoproteins indicates the importance of the selenocysteine incorporation pathway. For UGA to be decoded ('recoded' in ref. 7) as selenocysteine, several additional elements must be present. These include several *trans* elements and a stem–loop structure forming a *cis*-acting mRNA element called a 'selenocysteine insertion sequence' (SECIS) (3,8,9; and for reviews see 10,11). Although prokaryotic and eukaryotic selenocysteine incorporation systems have some analogous elements, they nevertheless seem to operate through different mechanisms.

In *Escherichia coli*, a number of genes have been identified in which UGA directs the incorporation of selenocysteine instead of acting as a stop codon. These include genes *fdhF* (12) and *fdnG* (13), encoding the selenocysteine-containing enzymes formate dehydrogenase H and N, respectively. Immediately downstream from the selenocysteine specifying UGA in the mRNA of each of these polypeptides is found a SECIS that has been described as consisting of at least 40 nucleotides (nt) capable of forming a stem–loop RNA structure. Insertion of a DNA segment directing this specific SECIS into a heterologous gene such as *lacZ* results in the incorporation of selenocysteine into the corresponding polypeptide (9,12,14–17). A number of *trans* elements are also required, including $tRNA^{Sec}$ (SelC), which is a specialized tRNA containing a UCA anticodon (18). $tRNA^{Sec}$ is recognized by a special elongation factor, called SelB which has been shown *in vitro* to bind specifically to both $tRNA^{Sec}$ and to the mRNA stem–loop structure (formed by the SECIS) adjacent to the 3′ side of the UGA codon (19). Recent *in vitro* work by Kromayer and colleagues (20) has shown that selenocysteinyl-$tRNA^{Sec}$ can be bound by the N-terminal part of *E.coli* SelB (homologous to EF-Tu), and that the C-terminal subdomain of SelB specifically binds to the SECIS stem–loop RNA structure of the *E.coli* selenoprotein formate dehydrogenase (FdhH). They also found that the mRNA motif recognized by SelB can be reduced to a 17 nt minihelix without loss of binding capacity. This minihelix consists of the upper part of the stem–loop RNA structure of the *fdhF* SECIS, a loop of 6 nt and five adjacent base-paired nucleotides together with a bulged nucleotide.

Here we have determined the nature of the mRNA SECIS motif required for UGA-directed selenocysteine incorporation in *E.coli in vivo*. We based our study on an extensive mutational analysis of the *fdhF* SECIS DNA which we inserted into a *lac′Z* fusion. Previously it has been suggested that the presence of the whole stem–loop structure is required for UGA directed selenocysteine incorporation (9,12,21). On the basis of our results presented here, we suggest that the presence of only 17 nt consisting of the upper stem–loop structure of *fdhF*, are sufficient to permit this 'recoding' choice, on the further condition that this sequence is located 11 nt downstream from the UGA codon.

MATERIALS AND METHODS

Materials and media

[^{75}Se]selenite (350 mCi/mmol) was obtained from Amersham (Amersham, UK). Bacteria were grown in YT or in M9 minimal medium (pH 7.0) (22) supplemented with a mixture of amino acids, each at a final concentration of 20 μg/ml. Ampicillin (100 μg/ml)

*To whom correspondence should be addressed. Tel: +972 2 6758250; Fax: +972 2 6784010; Email: hanita@cc.huji.ac.il

was added to the media in which the plasmid-carrying strains were grown.

Bacterial strains and plasmid derivatives

The *E.coli* strains used in this study were MC4100, its Δ*selC* derivative RM1 (15), and its Δ*selB* derivative WL81300 (23).

Plasmid pRM4 was constructed by us previously (15). It carries the fused genes λ*cI′–lac′I′′Z*, into which we inserted the TGA region of the SECIS DNA (from –9 to +47, where the U of the UGA codon is designated as residue number 1) of *E.coli fdhF* gene in the junction of λ*cI′* and *lac′I′′Z*. Here we used the PCR technique for the generation of a series of plasmids carrying site-directed mutations in the SECIS of *fdhF* by using the required primers and with plasmid pRM4 as the template. The sequence of their mutated SECIS RNA is described in Figures 1–5.

Molecular cloning

All recombinant DNA manipulations were carried out by standard procedures (24). Restriction enzymes and other enzymes used in the recombinant DNA experiments were obtained from New England Biolabs (Beverly, MA). DNA sequencing was done with the Sequenase kit of United States Biochemicals (Cleveland, OH).

Bacterial growth, transformations, and measurements of β-galactosidase activity

Escherichia coli cells were transformed (24) by the plasmid of choice. Single colonies of freshly transformed cells were grown at 37°C on YT plates for 6–8 h and then in M9 liquid medium at 37°C in a rolling drum for 3–4 h to the OD_{600} = 0.4–0.6. The cells were examined for β-galactosidase activity as described by Miller (22).

[^{75}Se]selenite labeling and identification of its incorporation *in vivo* into the fused gene product

Freshly transformed cells were grown in M9 liquid medium with cysteine and methionine (final concentration 200 μg/ml each) at 37°C overnight in the presence of 1.0 μCi [^{75}Se]selenite/ml. During labeling, cold selenite was added to a final concentration of 1.5 μM. Cells were lysed, and the proteins were separated on 5% SDS–polyacryamide gel by electrophoresis and detected by autoradiography as we described previously (15).

RESULTS

A mini upper-stem–loop structure of *fdhF* SECIS mRNA is required for selenocysteine incorporation into a polypeptide

For this work we used our plasmid pRM4 (15), which we constructed previously to carry the TGA codon context of the *E.coli fdhF* gene fused to the *lac′Z* gene (*lac′Z* lacking the first eight codons). pRM4 carries nt –9 to +47 of *fdhF* gene (where the nucleotide U of the UGA codon was designated as nt 1). The mRNA of this region of the *fdhF* gene has been identified as a stem–loop structure forming SECIS (Fig. 1Aa), the *cis* acting mRNA element described previously to be required for the *in vivo* UGA-directed selenocysteine incorporation into the *lac′Z* fusion product of *E.coli* (12). Here we have identified the minimal sequence required for UGA-directed selenocysteine incorporation in *E.coli in vivo*. We carried out extensive mutational analysis on the *fdhF* SECIS DNA fused to the *lac′Z* gene on plasmid pRM4. We

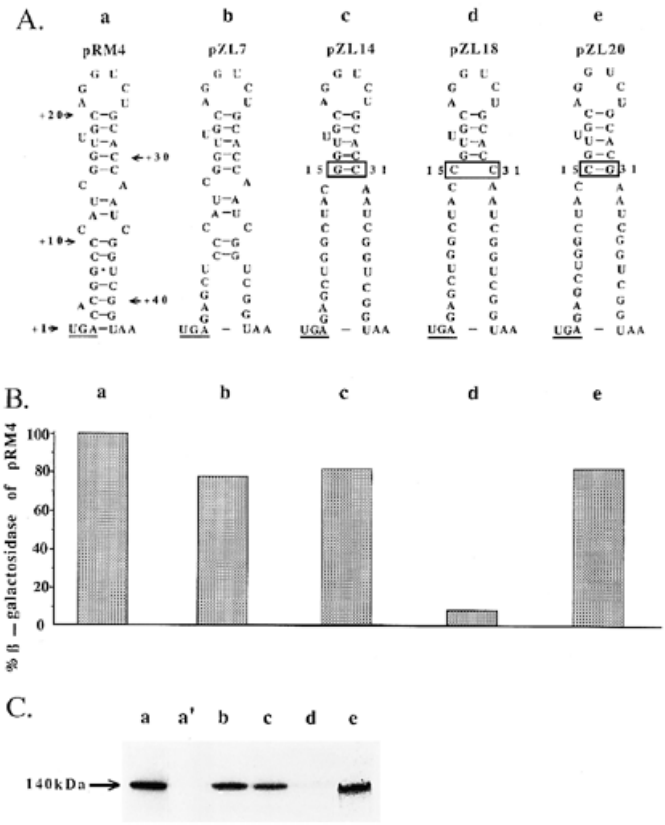

Figure 1. Mutations affecting the secondary structure of the *fdhF* SECIS mRNA downstream from the UGA codon of a *fdhF–lac′Z* gene fusion. (**A**) The positions of the mutations in the SECIS region of each plasmid. (**B**) Effect on the UGA-directed β-galactosidase activity. (**C**) Incorporation of [^{75}Se]selenite into the fused genes product. *Escherichia coli* MC4100 and its Δ*selC* derivative RM1 were transformed by plasmids pRM4 (a), pZL7 (b), pZL14 (c), pZL18 (d) and pZL20 (e). Freshly transformed cells were grown at 37°C and the level of β-galactosidase activity (B), and the incorporation of [^{75}Se]selenite into the polypeptide product of the fused genes (C) were determined as described in Materials and Methods. The percentage of β-galactosidase activity directed by the gene fusions in plasmids pZL7, pZL14, pZL18 and pZL20 is relative to that of pRM4. The numbers represent the average of the results of at least five experiments and were derived after subtracting the level of β-galactosidase activity in the Δ*selC* derivative from that in MC4100. After this background level was subtracted, the level of β-galactosidase activity was ~1000 Miller units. Labeling with [^{75}Se]selenite (C) was carried out in strain MC4100, except in sample Ca′ in which the labeling was carried out in a Δ*selC* derivative of MC4100 carrying pRM4. The protein product of the fused genes is found at the position equivalent to 140 kDa, indicated by the arrow. Note that in *fdhF* SECIS mRNA the nucleotide at position 18 is bulged as suggested by Zinoni *et al.* (12) and Heider *et al.* (9).

always studied selenocysteine incorporation into the gene product of the *lac′Z* fusion in two ways: (i) by measuring UGA-directed *(selC*-dependent) β-galactosidase activity; and (ii) by measuring the direct incorporation of [^{75}Se]selenite into the polypeptide of the fused genes. Since measuring β-galactosidase activity is the more accurate of the two methods, we used this method for our quantitative determinations of selenocysteine incorporation.

For this set of experiments, we generated site-directed mutations that cause changes in the secondary structure of the *fdhF* SECIS mRNA downstream from the UGA codon. First, we altered five

of the nucleotides (pZL7, Fig. 1Ab) and then 11 of the nucleotides (pZL14, Fig. 1Ac) in the stem region immediately after the UGA codon. We did this such that the nucleotides on the left side of the stem are identical rather than complementary to the corresponding nucleotide on the right side of the stem. As shown, the UGA-directed selenocysteine incorporation is still permitted even when the 11 nt following UGA are not base paired to their corresponding partners in the stem (Fig. 1Bc and Cc). Quantitatively, the level of the UGA-directed β-galactosidase activity is reduced only by ~25% (Fig. 1Bb and Bc). We obtained similar results by changing 11 nt at the right side of the stem to be identical rather than complementary to the unchanged left side of the stem (data not shown). These results suggest that neither the primary nor the secondary structure of this region in *fdhF* mRNA, designated here as the 'intermediary domain' (nt 4–14 and 32–42 from the UGA), are required for the incorporation of selenocysteine into the *lac′Z* fused polypeptide. However, the mutated construct pZL14 (Fig. 1Ac) still carries a 'mini upper-stem–loop' (a stem of 5 bp, one bulged nucleotide, and a loop of 6 nt) that may be responsible for the UGA-directed selenocysteine incorporation. Indeed, when we constructed and then studied plasmid pZL18 (Fig. 1Ad), we found that selenocysteine incorporation was almost completely abolished (Fig. 1Bd and Cd) when only one additional base-pairing is also prevented, thereby reducing the 'stem' part of the mini upper-stem–loop structure from 5 to only 4 bp (Fig. 1Ad). Furthermore, selenocysteine incorporation was again permitted (Fig. 1Be and Ce) when we introduced a compensatory mutation (pZL20, Fig. 1Ae). This mutation permits 5 bp in the mini upper-stem–loop due to a C–G bp (Fig. 1Ae) instead of the G–C bp in the SECIS of the pZL14 (Fig. 1Ac). Altogether, the results described so far suggest that the mini upper-stem–loop structure downstream of the UGA on *fdhF* SECIS mRNA, is required for the process of selenocysteine incorporation into a polypeptide *in vivo*, while the intermediary domain (nt 4–11 and 32–42) is not.

It was previously suggested that the presence of a C residue, after the UGA of the *E.coli* SECIS is crucial because it prevents the recognition of UGA by RF2 that actually recognizes 4 nt, the UGA and the first nucleotide at its 3′ side; the poorest 4 base terminator is UGAC (25). The results described herein show that the C residue following UGA in *fdhF* mRNA can be changed to a G residue without affecting the UGA-directed selenocysteine incorporation (Fig. 1). It seems therefore, that at least in the case of *fdhF* SECIS, the presence of the C residue at this position is not crucial for the prevention of the recognition of UGA by RF2.

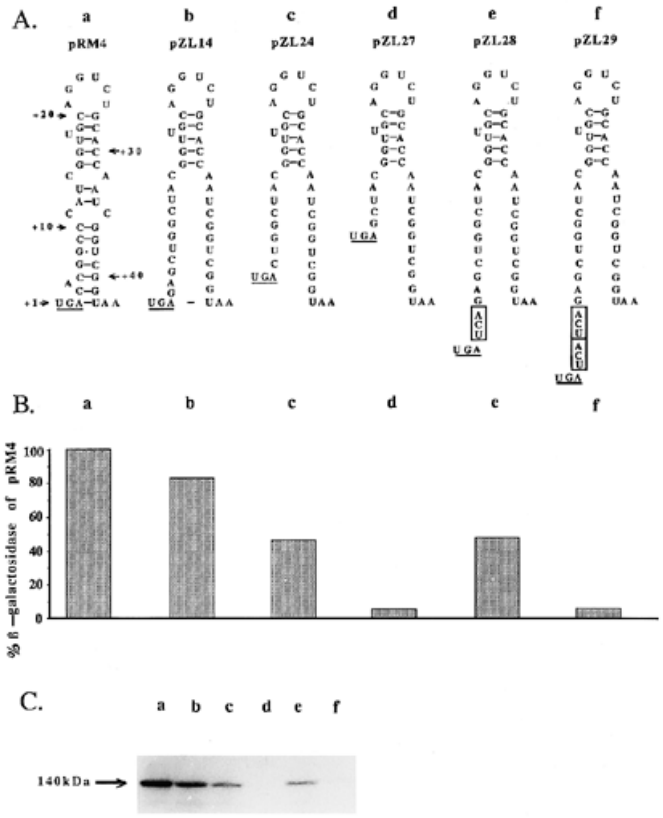

Figure 2. Mutations affecting the distance between the UGA codon and the mini upper-stem–loop in the *fdhF* SECIS mRNA of a *fdhF–lac′Z* gene fusion. (**A**) The positions of the mutations in the SECIS region of each plasmid. (**B**) Effect on the UGA-directed β-galactosidase activity. (**C**) Incorporation of [^{75}Se]selenite into the fused genes product. *Escherichia coli* MC4100 and its Δ*selC* derivative RM1 were transformed by plasmids pRM4 (a), pZL14 (b), pZL24 (c), pZL27 (d), pZL28 (e) and pZL29 (f). The level of β-galactosidase activity (B), and the incorporation of [^{75}Se]selenite into a fused protein (C) were determined as described in Figure 1 and in Materials and Methods. In pZL24 (c) the 3 nt $G_4A_5G_6$ and in pZL27 (d) the 6 nt $G_4A_5G_6C_7U_8G_9$ downstream from the UGA were deleted from pZL14 (b), and in pZL28 (e) the 3 nt (UCA) (boxed), and in pZL29 (f) the 6 nt (UCAUCA) (boxed) were inserted downstream from the UGA codon in pZL14 (b). Note that in *fdhF* SECIS mRNA the nucleotide at position 18 is bulged as suggested by Zinoni *et al.* (12) and Heider *et al.* (9).

The distance of the mini upper-stem–loop structure from the UGA of *fdhF* mRNA is crucial for selenocysteine incorporation into a polypeptide

Having shown that the base pairing of the intermediary domain of the *fdhF* SECIS is not required for UGA-directed selenocysteine incorporation into the polypeptide product of the *fdhF–lac′Z* gene fusion, the following question arose. Does the length of this intermediary domain affect the efficiency of selenocysteine incorporation? Using pZL14 (Fig. 2Ab) as the basis for these studies, we either effected deletions of 3 nt $\Delta G_4A_5G_6$ (pZL24, Fig. 2Ac) or of 6 nt $\Delta G_4A_5G_6C_7U_8G_9$ (pZL27, Fig. 2Ad), or additions of 3 nt UCA (pZL28, Fig. 2Ae) or 6 nt, UCAUCA (pZL29, Fig. 2Af). Reducing or increasing the length of the intermediary region by 3 nt decreases the level of the UGA-directed selenocysteine incorporation by ~50% (Fig. 2Bc,e and Cc,e). Moreover, either reducing or increasing the length of the intermediary region by 6 nt almost completely abolishes the UGA-directed selenocysteine incorporation (Fig. 2Bd,f and Cd,f).

U17 is a specific and bulged nucleotide of the mini upper-stem–loop structure downstream of the UGA of *fdhF* mRNA that is required for selenocysteine incorporation into a polypeptide

The solution structure of RNA hairpins within *fdhF* mRNA has recently been determined by Hüttenhofer and colleagues (21) who showed that there are three bulged nucleotides in the stem of the upper-stem–loop. What is the role *in vivo*, if any, of the bulged nucleotides in the mini upper-stem–loop RNA structure? As shown in Figure 3Aa, in plasmid pRM4, these are nucleotides U_{17}, U_{18} and A_{29}. Does replacing these specific bulged nucleotides affect the efficiency of *in vivo* selenocysteine

incorporation? We found that the incorporation of selenocysteine in the *lac′Z* fusion polypeptide is completely prevented (Fig. 3Bb,d and Cb,d) by replacing each of the following nucleotides: U_{17}→A_{17} (pZL42; Fig. 3Ad), U_{18}→A_{18} (pZL38; Fig. 3Ab) and A_{29}→U_{29} (data not shown). We could, however, cause the return of selenocysteine incorporation (Fig. 3Bc and Cc) by introducing the compensatory mutation A_{29}→U_{29} into pZL38, creating pZL43 in which A_{18} and U_{29} are base-paired (Fig. 3Ac). Thus, selenocysteine incorporation depends not on the presence of a U residue at position 18 (which is replaced by A in the mutant) but rather on its ability to base pair with the nucleotide at position 29. Note, however, that selenocysteine incorporation was completely abolished by either the single change U_{17}→A_{17} (pZL42, Fig. 3Ad, Bd and Cd) or the double change U_{17}→A_{17} and the compensatory mutation A_{29}→U_{29} (pZL44; Fig. 3Ae, Be and Ce). Furthermore, selenocysteine incorporation was also completely abolished by the single changes U_{17}→C_{17} or U_{17}→G_{17} (data not shown). Thus, selenocysteine incorporation appears to be dependent on both (i) the specificity of the nucleotide at position 17 as a U residue, and (ii) its presence as a bulged nucleotide. The importance of the bulged U_{17} has been recently shown also by a SELEX approach (26).

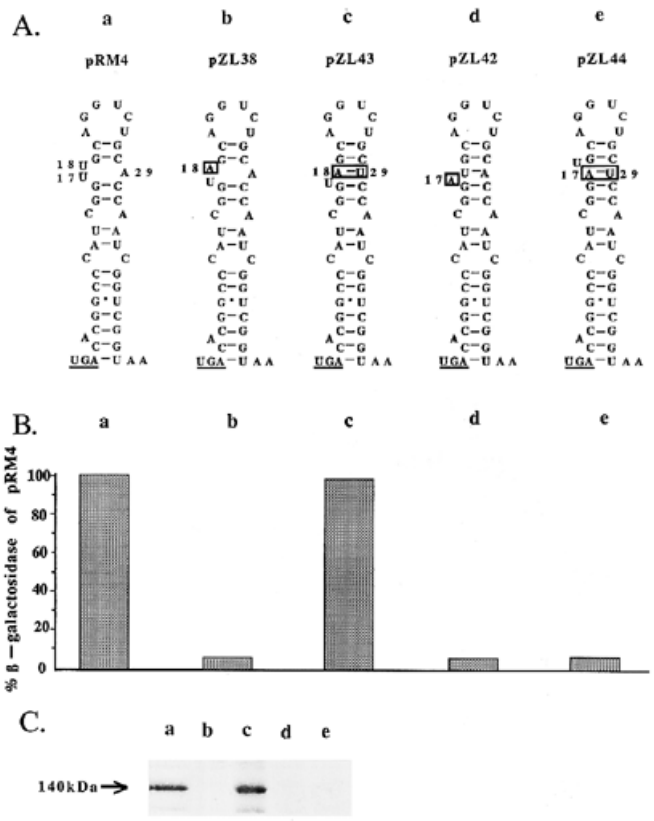

Figure 3. Mutations affecting the bulged nucleotides in the mini upper-stem–loop of the *fdhF* SECIS mRNA of a *fdhF–lac′Z* gene fusion. (**A**) The positions of the mutations in the SECIS region of each plasmid. (**B**) The effect on the UGA-directed β-galactosidase activity. (**C**) Incorporation of [^{75}Se]selenite into the fused genes product. *Escherichia coli* MC4100 and its Δ*selC* derivative RM1 were transformed by plasmids pRM4 (a), pZL38 (b), pZL43 (c), pZL42 (d) and pZL44 (e). The level of β-galactosidase activity (B), and the incorporation of [^{75}Se]-selenite into the fused genes product (C) were determined as described in Figure 1 and in Materials and Methods. The mutated nucleotides in positions 17, 18 and 29 are boxed. Here these nucleotides at these positions in pRM4 are shown as bulged as suggested by Hüttenhoffer *et al.* (21).

The nature of the mini upper-stem–loop structure downstream of the UGA of an *fdhF* mRNA that is required for selenocysteine incorporation into a polypeptide

Having determined that selenocysteine incorporation is possible even when the base pairing of the intermediary domain of the SECIS is opened, we asked how the specific structure of the mini upper-stem–loop affects selenocysteine incorporation. We particularly wished to distinguish between the role of single- and double-stranded regions in this structure. We did this by asking how the inversion of potentially base paired nucleotides might affect selenocysteine incorporation. It is assumed that 2 nt in the stem are in a base-paired state if selenocysteine incorporation is permitted even when they are inverted. Based on this assumption, we found that the following nucleotides must be paired in the mini upper-stem–loop structure: 15 with 31 (Fig. 4Ab, Bb and Cb), 16 with 30 (Fig. 4Ac, Bc and Cc), 18 with 29 (Fig. 4Ad, Bd and Cd), and 19 with 28 (Fig. 4Ae, Be and Ce). On the other hand, since the inversion of the C_{20}–G_{27} to G_{20}–C_{27} abolishes selenocysteine incorporation not only when the intermediary region is opened as in pZL66 (Fig. 4Af, Bf and Cf), but also when it is closed as in a pRM4 derivative (data not shown), it is possible that the pairing between nt 20 and 27 is not required and/or that the nature of the nucleotides in this position is important for selenocysteine incorporation (see Discussion). In addition, since selenocysteine incorporation is not detected when A_{17} is paired with U_{29} (Fig. 3Ae, Be and Ce), we conclude that U_{17} is required as a bulged nucleotide. Our experiments also reveal that, as previously shown by Heider *et al.* (9), for the ‘closed’ stem, selenocysteine incorporation is also completely prevented by single mutations in the loop region (A_{21}→U_{21}, G_{22}→C_{22}) when the intermediary region is ‘open’ (data not shown). Based on all our results, we suggest that the mini upper-stem–loop structure required for selenocysteine incorporation consists of 17 nt including (Fig. 6): (i) a region of 4 bp of non-specific nucleotides 15–31, 16–30, 18–29 and 19–28; (ii) a bulged U at position 17; and (iii) a sequence of eight specific nucleotides from C_{20} to G_{27}. Among these, C_{20} and G_{27} may be weakly paired and the rest are located in a loop (see Discussion).

Only the mini upper-stem–loop structure downstream of the UGA of *fdhF* mRNA is required for selenocysteine incorporation into a polypeptide

Based on the chemical and enzymatic probing data of experiments designed to elucidate the solution structure of mRNA hairpins within the *fdhF* SECIS, Hüttenhofer and colleagues (21) have suggested that upstream from the UGA codon there is an extended helix of 7 bp in which the U and G residues of the UGA codon are included, but the A residue is bulged (Fig. 5Aa). Does this extended helix upstream from the UGA codon have a role in the incorporation of selenocysteine *in vivo*? To answer this question, we mutated the extended helix upstream from the UGA such that the secondary structure 5′ to the UGA could not be formed. In pRM4, we changed both nucleotides at the 5′ side and the 3′ side of the UGA; at the 5′ side 5 nt were changed to A residues and nt 42–50 at the 3′ side of the UGA were changed to AUUUAAAAA, respectively (pZL40; Fig. 5Ab). As shown, these changes did not significantly affect selenocysteine

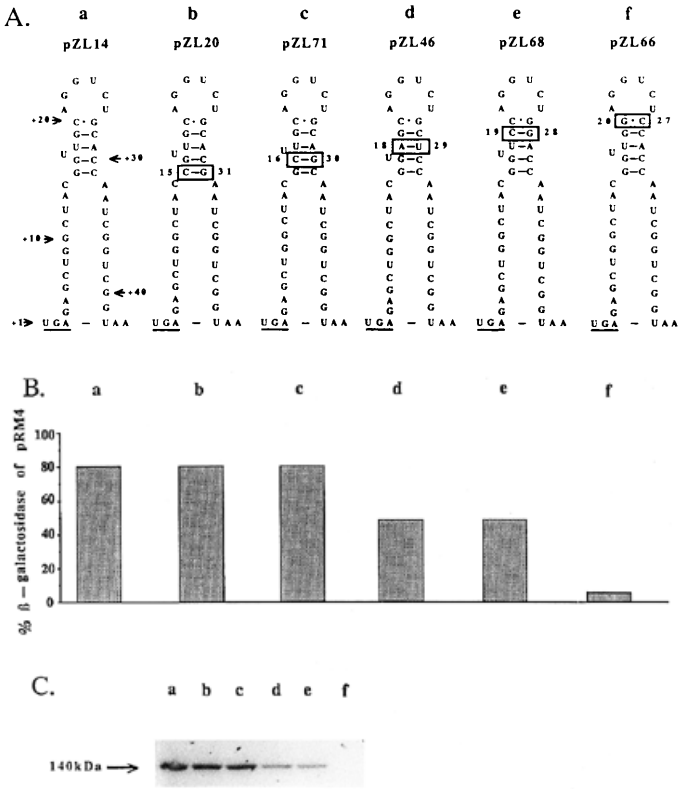

Figure 4. Mutations affecting the mini upper-stem–loop structure of the *fdhF* SECIS mRNA of a *fdhF–lac′Z* gene fusion, when the 'intermediary domain' of the SECIS is opened. (**A**) The positions of the mutations in the SECIS region of each plasmid. (**B**) The effect on the UGA-directed β-galactosidase activity. (**C**) Incorporation of [^{75}Se]selenite into the fused genes product. *Escherichia coli* MC4100 and its Δ*selC* derivative RM1 were transformed by plasmids pZL14 (a), pZL20 (b), pZL71 (c), pZL46 (d), pZL68 (e) and pZL66 (f). The level of β-galactosidase activity (B), and the incorporation of [^{75}Se]selenite into the fused genes product (C) were determined as described in Figure 1 and in Materials and Methods. The mutated nucleotides are numbered and boxed. Here the nucleotide in position 17 is bulged, according to results as presented in Figure 3.

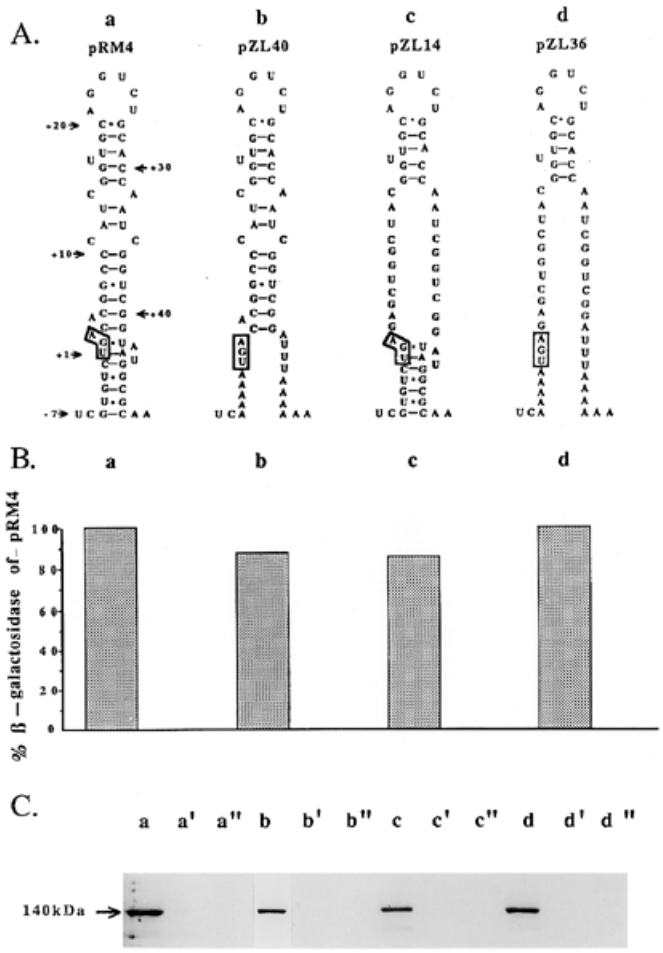

Figure 5. Mutations affecting the secondary structure of the *fdhF* SECIS mRNA both upstream and downstream from the UGA codon of a *fdhF–lac′Z* gene fusion. (**A**) The positions of the mutations in the SECIS region of each plasmid. (**B**) The effect on the UGA-directed β-galactosidase activity. (**C**) Incorporation of [^{75}Se]selenite into the fused genes product. *Escherichia coli* MC4100, its Δ*selC* derivative RM1 and its Δ*selB* derivative WL81300 were transformed by plasmids pRM4 (a), pZL40 (b), pZL14 (c) and pZL36 (d). The level of β-galactosidase activity (B), and the incorporation of [^{75}Se]selenite into the fused genes product (C) were determined as described in Figure 1 and in Materials and Methods. Labeling with [^{75}Se]selenite (C) was carried out in MC4100 (a–d), in Δ*selC* RM1(a′–d′), and in the Δ*selB* derivative WL81300 (a″–d″). The UGA codon is boxed. Here the nucleotide in position 17 is bulged, according to the results presented in Figure 3.

incorporation (Fig. 5Bb and Cb). Moreover, even when the same changes were made in pZL14, so that both the upstream UGA stem and the intermediary domain were 'opened' (pZL36; Fig. 5Ad), selenocysteine incorporation was unaffected (Fig. 5Bd and Cd). Compare particularly the results found for plasmids pRM4 (5Ba and Ca) and pZL36 (Fig. 5Bd and Cd): pRM4 carries the whole extended helix of *fdhF* (that is the region upstream from UGA), the intermediary domain, and the mini upper-stem–loop structure (Fig. 5Aa). Though in pZL36 only the mini upper-stem–loop structure is unchanged, however, while at a proper distance of 11 nt from the UGA (Fig. 5Ad), it still supports the same level of selenocysteine incorporation as does pRM4. These results further support our conclusion that *in vivo* selenocysteine incorporation into the polypeptide product of the *fdhF–lac′Z* fused genes requires only the mini upper-stem–loop structure placed at the proper position.

DISCUSSION

The UGA stop codon in the mRNA of the *E.coli* formate dehydrogenase gene *fdhF* has been described to be decoded as selenocysteine when it is in the context of a SECIS of at least 40 nt following the UGA (12). This process is dependent on the ability of this SECIS to form a stem–loop RNA structure (9). These conclusions were based on the results of mutational analyses of this stem–loop structure directed by the fusion of the *fdhF* SECIS DNA and the *lac′Z* gene in an *E.coli* strain MC4100 (*rpsL*$^+$ *rpsE*). It has been shown convincingly that the specificity of each of the nucleotides in the loop and the base pairing of the adjacent two nucleotides (mostly the second) located in the stem are crucial for the process of the UGA-directed selenocysteine incorporation into the fused polypeptide (12). These analyses, however, have not dealt fully with the rest of this mRNA region, and thus do not distinguish among the following possibilities: is UGA-directed selenocysteine incorporation dependent on (i) the nucleotide sequence of the SECIS; and/or (ii) the secondary structure of the

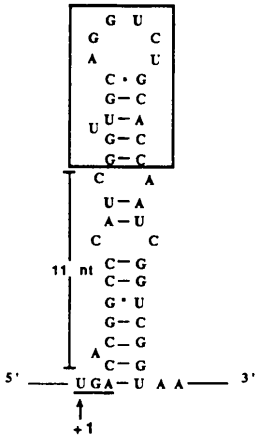

Figure 6. The specific characteristics, based on this study, of the minimal SECIS of *E.coli fdhF* mRNA required for *in vivo* UGA-directed selenocysteine incorporation. The SECIS includes a mini upper-stem–loop structure (boxed) that probably provides the SelB binding domain *in vivo*. For selenocysteine incorporation to take place, the locus of this mini upper-stem–loop structure must be at a distance of 11 nt (unpaired or paired) from the UGA codon (underlined). The nucleotide residue at position 17 must be U and is bulged. The pairing of bases C_{20} and G_{27} is questionable (see Discussion) and is therefore designated by a dot instead of by a dash.

SECIS; and/or (iii) the distance of the loop from the UGA codon? Also, does the extended helix upstream from the UGA codon (21) play a role in the *in vivo* process of selenocysteine incorporation?

By a process of progressive elimination, we arrived at a minimal SECIS which can still facilitate *in vivo* UGA-directed selenocysteine incorporation into the polypeptide product of a plasmid borne *fdhF–lac′Z* gene fusion. We determined the level of the UGA-directed selenocysteine incorporation as reflected by: (i) the level of the *selC*-dependent β-galactosidase activity; and (ii) [^{75}Se]selenite incorporation into the fused polypeptide. These studies were carried out in *E.coli* strain MC4100 (*rpsL150*) (15). We found that in order to permit selenocysteine incorporation *in vivo* the previously described *fdhF* SECIS (9,12,21) can still function properly even when it is reduced to only 17 nt forming a mini upper-stem–loop RNA structure (Figs 1, 4 and 5). We found, however, that two major regions of the SECIS were not crucial for this *in vivo* process, and these are what we call: (i) the intermediary domain, the 11 bp double-stranded RNA region, including nt 4–14 and 32–42 downstream from the UGA codon (Fig. 1); and (ii) the upstream UGA domain, the double-stranded RNA structure, including nt –5 to 1 upstream from the UGA and nt 41–46 downstream from the UGA (Fig. 5). Preventing the formation of the secondary structure of either the intermediary domain (Fig. 1) or the upstream UGA domain (Fig. 5) caused a reduction of only ~20–30% in the level of the UGA-directed selenocysteine incorporation. We have also shown that neither the primary nor the secondary structure of the intermediary domain are required for the process. Our experiments suggest that the intermediary domain assures the existence of the proper distance of 11 nt (from the UGA to the mini upper-stem–loop structure) requisite for optimal selenocysteine incorporation. Thus, these intermediary RNA nucleotides can be in either single or double-stranded form (Fig. 2). It is the distance of the mini upper-stem–loop structure from the UGA that is crucial for the process.

Having found the lower size limit for the SECIS to be 17 nt, we further analyzed the mini upper-stem–loop RNA structure necessary for selenocysteine incorporation. By examining the effects of inverting potentially base-paired nucleotides, we particularly distinguished between single- and double-stranded RNA regions in this structure required for the process (Figs 3 and 4). Our results show that the mini upper-stem–loop structure required for selenocysteine incorporation consists of 17 nt including (Fig. 6): (i) a region of 4 bp, 15–31, 16–30, 18–29 and 19–28 of non-specific nucleotides; (ii) a bulged U at position 17; and (iii) a sequence of eight specific nucleotides from C_{20} to G_{27}. In our experiments we were not able to show the involvement of the base pair C_{20}–G_{27}; inversion of the base pair to G_{20}–C_{27} does not permit selenocysteine incorporation (see Results and Fig. 4Af, Bf and Cf). Since Heider *et al.* (9) have previously shown that such an inversion partially affects selenocysteine incorporation (reduction to 30%), we assume that C_{20}–G_{27} base pairing is questionable. Therefore, in our model on the nature of the minimal *E.coli* SECIS (Fig. 6), the base pair C_{20}–G_{27} is represented by a dot.

Based on our results, we suggest a minimal *E.coli* SECIS that consists of a mini upper-stem–loop structure with a proper distance from the UGA as described in Figure 6. It is interesting that this minimal SECIS in genes *fdhF* and *fdnG* is very similar (21).

The special elongation factor SelB was shown to be structurally divided into two parts: (i) an N-terminal half which shares extensive sequence homology with EF-Tu and binds to selenocysteinyl-tRNA^{Sec} and GTP; and (ii) a separate C-terminal domain, required for specifically binding to the *fdhF* mRNA stem–loop structure 3′ adjacent to the UGA codon *in vitro* (19). Moreover, the mRNA motif recognized by SelB can be reduced to a 17 nt minihelix without loss of binding capacity *in vitro* (20). This minihelix is identical to the mini upper-stem–loop of *fdhF* mRNA described by us here to be the minimal SECIS required for the *in vivo* incorporation of selenocysteine in *E.coli* (Fig. 6). The C-terminal domain of SelB probably exclusively binds also *in vivo* to this upper stem–loop RNA structure. Hüttenhofer *et al.* (27) have recently proposed a model where the upper part of the *fdhF* mRNA hairpin which binds to the C-terminus of SelB can be placed ‘outside’ the ribosomal mRNA track during UGA decoding. The nature of the minimal *E.coli* SECIS described here (Fig. 6) corroborates with such a model; only small variations are suggested. The model suggested by Hüttenhoffer *et al.* (27) was based on *in vitro* toeprint experiments with 30S ribosomal subunits. They contend that when the mRNA is unfolded by the 30S ribosomal subunit there are 16 unfolded nucleotide residues from the first nucleotide at the P site to the 3′ side of the UGA of *fdhF* mRNA. In our *in vivo* experiments described here, we found that selenocysteine is incorporated when there are not more than 17 unfolded nucleotides in this region; that is 11 nt from the UGA to the mini upper-stem–loop structure (Fig. 6) plus the 6 nt included in the P and A sites. Therefore, we propose that there may be a safety measure whereby the *E.coli* 70S elongating ribosome can even unfold 17 nt and still permit UGA-directed selenocysteine incorporation into the translated protein product. Such a model is supported by the fact that the *E.coli* ribosome covers 18 ± 1 nt downstream from the P-site codon (28).

In conclusion, our results show that just the mini upper-stem–loop domain of *fdhF* mRNA (located at the proper distance of 11 nt from the UGA) is required for selenocysteine incorporation *in vivo*. However, the UGA codon context of *fdhF* and *fdnG* mRNAs

generates a more extensive secondary structure (9,12,21), designated here as the 'intermediary domain' and the 'upstream UGA domain'. We are presently investigating the possibility that under specific physiological conditions the intermediary and upstream domains may have a role in selenocysteine incorporation into polypeptides.

ACKNOWLEDGEMENTS

We thank Dr August Böck (Munich, Germany) for kindly providing us with bacterial strains. We are deeply grateful to F. R. Warshaw-Dadon (Jerusalem, Israel) for her critical reading of the manuscript. This research was supported by grants from the United States–Israel Binational Science Foundation (BSF) and the Israel Science Foundation administrated by The Israel Academy of Science and Humanities.

REFERENCES

1 Chambers, I. and Harrison, P. R. (1987) *Trends Biochem. Sci.* **12**, 255–256.
2 Böck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B. and Zinoni, F. (1991) *Mol. Microbiol.* **5**, 515–520.
3 Berry, M. J. and Larson, P. R. (1993) *Biochem Soc. Trans.* **21**, 827–832.
4 Stadtman, T. C. (1996) *Annu. Rev. Biochem.* **65**, 83–100.
5 Böck, A., Forchhammer, K., Heider, J., Leinfelder, W. and Baron, C. (1991) *Trends Biochem. Sci.* **16**, 463–467.
6 Stadtman, T. C. (1991) *J. Biol. Chem.* **266**, 16257–16260.
7 Gesteland, R. F., Weiss, R. B. and Atkins, J. F. (1992) *Science* **257**, 1640–1648.
8 Lee, B. J., Worland, P. J., Davis, J. N., Stadtman, T. C. and Hatfield, D. L. (1989) *J. Biol. Chem.* **264**, 9724–9727.
9 Heider, J., Baron, C. and Böck, A. (1992) *EMBO J.* **11**, 3759–3766.
10 Baron, C. and Böck, A. (1995) In Söll, D. and RhajBhandary, U. (eds), *tRNA: Structure, Biosynthesis and Function*. American Society for Microbiology, ASM Press, Washington, DC. pp. 529–544.
11 Low, S. C. and Berry, M. J. (1996) *Trends Biochem. Sci.* **21**, 203–208.
12 Zinoni, F., Heider, J. and Böck, A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4660–4664.
13 Berg, B. L., Li, J., Heider, J. and Stewart, V. (1991) *J. Biol. Chem.* **266**, 22380–22385.
14 Chen, G. T., Fang, L. and Inouye, M. (1993) *J. Biol. Chem.* **268**, 23128–23131.
15 Reches, M., Zhao, C. and Engelberg-Kulka, H. (1994) *J. Appl. Env. Microbiol.* **60**, 45–50.
16 Zhao, C., Reches, M., Depalo, D., St Germain, D. L. and Engelberg-Kulka, H. (1994) *Gene* **148**, 351–356.
17 Liu, Z., Reches, M. and Engelberg-Kulka, H. (1997) *Anal. Biochem.* **244**, 40–44.
18 Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M. A. and Böck, A. (1988) *Nature* **331**, 723–725.
19 Baron, C., Heider, J. and Böck, A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4181–4185.
20 Kromayer, M., Wilting, R., Tormay, P. and Böck, A. (1996) *J. Mol. Biol.* **262**, 413–420.
21 Hüttenhofer, A., Westhof, E. and Böck, A. (1996) *RNA* **2**, 354–366.
22 Miller, J. H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laroratory Press, Cold Spring Harbor.
23 Tormay, P., Sawers, A. and Böck, A. (1996) *Mol. Microbiol.* **21**, 1253–1259.
24 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
25 Tate, W. P. and Mannering, S. A. (1996) *Mol. Microbiol.* **21**, 205–211.
26 Klug, S. J., Hüttenhofer, A., Kromayer, M. and Famulok, M. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 6676–6681.
27 Hüttenhofer, A., Heider, J. and Böck, A. (1996) *Nucleic Acids Res.* **24**, 3903–3910.
28 Beyer, D., Skripkin, E., Wadzack, J. and Nierhaus, K. (1994) *J. Biol. Chem.* **269**, 30713–30717.