Recognition of UGA as a selenocysteine codon in eukaryotes: a review of recent progress
Marla J. Berry* and P. Reed Larsen
Thyroid Division, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School,
75 Francis Street, Boston MA 02115, U.S.A.

Introduction
UGA codons function alternatively either as selenocysteine codons or as termination signals both in prokaryotes and in eukaryotes. This cotranslational incorporation of selenocysteine in UGA codons has earned it the designation of the 21st amino acid [1]. Considerable progress has been made recently in unravelling the process of selenocysteine incorporation, particularly in prokaryotes. Specific tRNAs with anticodon complementarity to the UGA codon have been identified in eubacteria [2], in protists [3], in fungi and in plants [4] and throughout the animal kingdom [5]. Designated tRNA^{ser}_{sec}, these tRNAs are charged with serine, which is then converted to selenocysteine by specific cellular enzymes. In addition, a selenocysteyl-tRNA specific elongation factor, SELB, has been identified and cloned in Escherichia coli [6]. Selenocysteine incorporation into E. coli formate dehydrogenase H (fDH) requires a stem-loop structure in the mRNA immediately adjacent to the UGA codon [7], and recent studies have demonstrated specific interactions between the SELB protein, the mRNA stem–loop and selenocysteyl-tRNA^{ser}_{sec} [8]. Furthermore, mutagenesis has shown that the position of the stem–loop and the sequence of the loop are critical to the function [9].

In eukaryotes, much less is known about UGA codon recognition. The existence of an eukaryotic SELB homologue has not been demonstrated, but is postulated on the basis of similar tRNA^{ser}_{sec} species and a requirement for distinguishing selenocysteine codon function from enzymes. No stem–loops analogous to the one in E. coli fDH are found near to the UGA codons in eukaryotic selenoprotein mRNAs. Our studies of the selenoenzyme type 1 iodothyronine 5' deiodinase (5'DI) have provided new information about selenocysteine incorporation in eukaryotes. 5'DI contains a selenocysteine in the active site, which confers specific kinetic properties and inhibitor sensitivities on the enzyme [10, 11]. Conversion of the UGA codon in the rat 5'DI to a UGU cysteine codon results in the production of a cysteine mutant enzyme that has altered biochemical properties. Expression studies of the wild-type selenoenzyme and of the cysteine mutant revealed that sequences in the 3' untranslated region (3' ut) of the mRNA are absolutely required for the translation of the former, but not for that of the latter [12]. Since the only difference between these two mRNAs is the UGA to UGU mutation, this result demonstrated that the 3' sequences are required for selenocysteine incorporation at the UGA codon. Further studies showed that the level of expression of the selenoenzyme in transient transfections was about two orders of magnitude lower than that of the cysteine mutant, indicating that UGA codon read-through is an inefficient process [13].

Deletion analyses in the 3' ut delineated the critical region as a stretch of ~ 250 nucleotides, and computer folding predicts the formation of a stable stem–loop structure in this region of the mRNA [12]. Further evidence for the existence of this secondary structure is our demonstration that a small deletion of nine nucleotides, removing only the putative loop, results in loss of function. Cloning of the human 5'DI revealed a similar structure in this 3' ut, and functional analysis also confirmed its role.

Having identified apparently critical 3' ut sequences in the 5'DI genes, we were interested in determining if these elements are part of a general mechanism for recognizing UGA as a selenocysteine codon. The only other cloned eukaryotic selenoprotein was glutathione peroxidase (GPX). We therefore examined the 3' ut region of the rat GPX cDNA for structures similar to those in the 5'DI genes. A stable stem–loop was indeed predicted in this sequence and in the other GPX genes, and certain nucleotides in the loop and unpaired bulge regions were conserved between the 5'DI and GPX 3' ut sequences [12]. Substitution of the putative GPX stem–loop for that of the rat 5'DI resulted in UGA codon read-through, as assessed by the production of 5'DI activity (Figure 1) [12], and a loop deletion of only eight nucleotides abolished the function of this element. These results not only
Chimeric constructs between the coding region of 5'DI and SECIS elements from GPX or from Sel P

Constructs contained nucleotides 1–906 from the rat 5'DI cDNA linked to the SECIS elements indicated. Sel P1, Sel P2, and Sel P1,2 designate constructs containing the first, second or both first and second SECIS elements from rat Sel P. UGA codon read-through was assessed by measuring 5'DI activity, and is expressed relative to a value of 1.0 for the wild-type 5'DI construct. (Adapted from [12, 22].)

Materials and methods

SECSIS and UGA mutant constructs

The rat Sel P cDNA was a generous gift of Drs. Kristina Hill and Raymond Burk [14]. Indicated regions or mutations of the rat 5'DI, GPX or Sel P cDNAs were cloned into a CDM-8 vector containing the first 906 nucleotides of the rat 5'DI cDNA (coding region nt 7–777). PCR or oligonucleotide-directed mutagenesis was performed as described previously [12]. All mutated regions were sequenced in their entirety to confirm that no other mutations were present. All manipulations and sequencing were performed at the DNA level. However, to avoid confusion, selenocysteine codons are referred to as UGA, rather than TGA codons, throughout this paper.

Deiodinase assays

Transient transfections in COS-7 cells were performed as described previously [17]. Cell sonicates were incubated with [125I] reverse T3 (600 nM) for 30 min at 37°C, and release of 125I- quantitated as described previously [11]. Activity is normalized to human growth hormone in the media produced from a cotransfected, constitutively expressed, human-growth-hormone-encoding plasmid. All transfections and assays were in duplicate and all experiments were performed at least twice.

BrAc[125I]T3, labelling

Synthesis of BrAc[125I]T3 has been described in detail elsewhere [13]. Affinity labelling was performed with 100 μg of transfected COS cell sonicate protein in a reaction volume of 50 μl containing 0.05 μCi of BrAc[125I]T3, 0.1 M potassium phosphate, pH 6.9, 1 mM EDTA and 20 mM dithiothreitol. Reactions were incubated for 10 min at 37°C and analysed on denaturing 12.5% polyacrylamide gels and then autoradiographed.
**Results and discussion**

**Sel P SECIS elements**

The presence of ten in-frame UGA codons in the Sel P mRNAs, and of \( \sim 8 \) mol of Se per mol of Sel P protein [18] indicate that the process of selenocysteine incorporation functions quite efficiently in this protein. Computer folding of the Sel P 3' ut predicts two putative SECIS elements with conserved nucleotides and secondary structures in common with those of the rat 5'DI and GPX. To assess the function of the Sel P SECIS elements, the two stem-loops were linked, in combination and individually, downstream of the rat 5'DI coding region and the resulting chimeric constructs were assayed for the production of deiodinase activity. We have shown previously that the translation of the rat 5'DI UGA as selenocysteine, as measured by the expression of deiodinase activity, required a functional SECIS element linked downstream [12]. Surprisingly, a segment of Sel P 3' ut consisting of both stem-loops produced nearly fourfold higher activity than the rat 5'DI SECIS element (Figure 1). The first stem-loop alone was about threefold more active than the 5'DI SECIS element, while the activity of the second Sel P stem-loop was comparable with that of the 5'DI. Interestingly, the relative activities of the Sel P, 5'DI, and GPX SECIS elements corresponds to the order in which these proteins reappear on selenium supplementation in selenium-deficient animals [14, 19, 20]. This suggests that the hierarchy of selenoprotein synthesis may reflect the relative efficiencies of the SECIS elements, in addition to the regulation of selenoprotein mRNA levels by selenium.

**Critical features in SECIS elements**

In our initial report describing the 5'DI and GPX SECIS elements, we identified a number of conserved nucleotides, including UAAA in the loops and UGAU in the 3' arms of the stems. A recent report [21] demonstrates that deletion of either of these four-nucleotide sequences in the GPX element abolishes function, and the authors concluded that this result specifically shows the necessity of these sequences. However, both these deletions and those used in our earlier studies would be predicted to alter the stem-loop structure. With the identification of two new functional SECIS elements, several apparently conserved nucleotides can be eliminated from the consensus sequence. The remaining consensus includes three As in the loop region of each element, an AUG in an unpaired bulge in the 5' arm of each stem and a UGR in a bulge in the 3' arms. Furthermore, the spacing from these unpaired bases to the loops is 10-12 nucleotides in all four elements. The conserved positions are shown in the rat 5'DI SECIS element by the circled or boxed nucleotides in Figure 2. Site-directed mutagenesis was used to generate nucleotide substitutions in the various SECIS elements, and the mutant elements were tested for their ability to direct the translation of the 5'DI UGA codon. The results with the 5'DI element, which are summarized in Figure 2, showed that each of the As contributed to SECIS element function, as did the UGR on the right side of the stem. Similar analyses in the GPX and Sel P2 elements confirmed the UGR requirement [22]. Further mutations in Sel P2 delineated the critical
role of the AUG bulge nucleotides on the left side of the stem.

The requirement for secondary structure in the region at the top of the 5'DI SECIS stem was also examined. When five nucleotides predicted to be involved in base-pairing (CAGCU, indicated by solid circles in Figure 2) were mutated to eliminate complementarity, SECIS element function was lost. Compensating mutations to the opposite side of the putative stem (AGUG) restoring the ability to base-pair also restored function. Thus, the secondary structure in this region is critical, but the specific sequence is not. The requirement for specific nucleotides in the loop and bulge regions, separated by base-paired regions of 10–12 nucleotides, suggests multiple contacts between these RNAs and their putative binding protein(s). Similar requirements for primary sequence and for secondary structure have been described in numerous other RNA–protein complexes [23–26]. The specific nature of these interactions in SECIS elements must await the characterization of the proteins involved.

**UGA codon context**

As discussed above, the spacing between the UGA selenocysteine codon and the adjacent stem–loop in *E. coli* fDH is critical for selenocysteine incorporation into this protein. However, the role of UGA codon context in eukaryotic selenocysteine insertion has not been defined previously. To examine this requirement, we used site-directed mutagenesis to introduce UGA codons in to the rat 5'DI coding sequence. Four mutant constructs each contained a single UGA codon in an ‘out-of-context’ position (codons 95, 105, 124 or 194), with a cysteine codon in place of the wild-type UGA (codon 126). Two additional constructs each contained an ‘out-of-context’ UGA (codons 194 or 248) in addition to the wild-type UGA, necessitating translation through two UGA codons. Translation of these mutants was assessed by BrAc[125]T3 affinity labelling of COS-cell homogenates after transient expression. BrAcT3 is a substrate analogue of T3 that covalently labels both the wild-type and the cysteine-126 mutant 5'DI proteins (Figures 3a and

**Figure 3**

Affinity labelling of 5'DI proteins produced by translation through mutant UGA codons

(a) Rat 5'DI constructs contain UGA codons at the positions indicated [10] and either the wild-type UGA or a cysteine codon in place of the wild-type UGA. Labelling with BrAc[125]T3 was done as described in the Materials and methods section. Lane 1, CD8 vector control; lane 2, 5'DI cysteine mutant at position 126; lane 3, wild-type 5'DI; lanes 4–9, 5'DI constructs, SECIS elements and positions of UGAs are indicated above the lanes. The far left lane contains 14C-labelled protein standards of 200, 94, 68, 45, 30, 18, and 14 kDa. (b) The 3' ut region consists either of that of the 5'DI, both 3' ut stem-loops of Sel P, or of the 3' ut of 5'DI with the SECIS region deleted. Lanes 1–3 are as in (a) above, lanes 4–12, 5'DI constructs, SECIS elements and positions of UGAs are indicated above the lanes. (Reproduced from [22]).
The New Biology of Protein Synthesis

3b, lanes 2 and 3) [13]. BrAcT3-labelled proteins of the same size as the wild-type enzyme were produced by all constructs containing a single 'out-of-context' UGA codon (Figures 3a and 3b). These proteins were not detectable in extracts from cells transfected with the vector alone (Figure 3a and 3b, lanes 1) or with constructs from which the 5'DI SECIS element had been deleted (Figure 3a, lanes 5, 7, 9; Figure 3b, lanes 6, 9, 12). In addition to the full-length protein, the UGA-248 mutant also produced a prominent BrAcT3-labelled protein of the size expected from termination at the second UGA (Figure 3a), indicating that the truncated protein also binds to the substrate analogue. In the light of our demonstration that the Sel P SECIS region was ~fourfold more active than that of the 5'DI, this sequence (nt 1209–2550, containing both Sel P loops) was introduced in the 3' ut region of three of the UGA mutants. The construct containing the wild-type UGA and a UGA-194 mutation expressed only a small amount of full length protein when the 5'DI SECIS element was present (Figure 3b), but the amount increased significantly with the Sel P SECIS element linked downstream. Three to four-fold increases were also observed with the two single UGA mutants tested (Figure 3b, lanes 7 compared with 8, 10 compared with 11), consistent with the results in Figure 1. As shown in Figures 3(a) and 3(b), the efficiency of translation of the wild-type 5'DI UGA codon in transfected cells is low compared with the cysteine mutant. In accordance with this result, the lighter bands from cells transfected with the UGA-126, 194 constructs reflects the difficulty in translating two UGAs as compared to one in the wild-type protein.

When homogenates of cells transfected with these mutants were assayed for the expression of deiodinase activity, the activities were low compared with the wild-type and cysteine mutant proteins (results not shown). However, all of the constructs produced activity significantly above background, but only when a SECIS element was present. The activities were three to four-fold higher with the Sel P SECIS element, compared with the 5'DI SECIS element. The UGA mutants in the wild-type background were considerably more active than any of the UGA mutants in the cysteine background, despite the reduced efficiency of read-through with two UGAs, consistent with our previous studies showing that the catalytic efficiency of the wild-type enzyme is about two orders of magnitude greater than that of the cysteine mutant [13]. These studies confirm an important difference between prokaryotic and eukaryotic selenoprotein synthesis that was predicted in our earlier studies [12], that the SECIS element will allow recognition of UGA as a selenocysteine codon in a variety of contexts. Also in agreement with these results is a recent report demonstrating UGA codon read-through and selenium labelling in a UGA mutant non-selenoprotein when the rat GPX SECIS element was linked 3' of the coding region [21]. The mechanism by which SECIS elements function to direct this process is under investigation currently.

This work was supported by NIH grants DK36256 and DK45944. We are grateful for the technical assistance of Laila Banu and John Harney, and to Drs. Kristina Hill and Raymond Burk for the generous gift of the rat Sel P cDNA.

Selenium has been recognized as an essential trace element since the late 1950s [1]. One of its functions is protection against damage caused by oxidant species. The anti-oxidant protection afforded by selenium is associated with its incorporation into proteins. Selenium is specifically incorporated into proteins as the amino acid selenocysteine. The first mammalian selenoprotein to be identified and characterized was cellular glutathione peroxidase, which detoxifies hydroperoxides. Table 1 lists other selenocysteine-containing proteins that have been identified, purified and characterized to varying extents.

Selenoprotein P is thought to provide protection against oxidant damage. Selenium-deficient rats are exceedingly sensitive to oxidative damage caused by the herbicide diquat. Administration of selenium to selenium-deficient rats affords protection against diquat toxicity within hours [9]. This protection is not due to glutathione peroxidase because its activity does not increase by the time protection is restored. However, selenoprotein P concentrations are increased when protection returns [10]. Thus, selenoprotein P is thought to play a role in protection against oxidant species.

Selenoprotein P has been purified from rat plasma by immunoaffinity chromatography [11]. It is a glycosylated plasma protein with an apparent molecular mass of 57 kDa, as determined by SDS/PAGE [12]. Deglycosylation of the protein produced a polypeptide that migrated at 43 kDa on SDS/PAGE. The protein has a very high selenium content and the selenium is incorporated into the primary structure of the protein as selenocysteine. Removal of selenoprotein P from plasma by immunoaffinity chromatography resulted in the removal of 65% of the selenium from rat plasma. Thus, selenoprotein P is the major selenoprotein in rat plasma.

Cloning of selenoprotein P

Partial amino acid sequence data was obtained by sequencing the N-terminus of the purified protein and of peptides generated from the protein by cyanogen bromide treatment or by enzymic digestion with V8 protease [8]. At the same time, a rat liver cDNA library was screened for the expression of selenoprotein P with polyclonal antibodies raised against rat selenoprotein P. One clone was selected for sequencing. Comparison of the N-terminal amino acid sequence of one of the peptide fragments and the deduced amino acid sequence of a portion of the cDNA clone showed that the sequences were identical. The cDNA clone contained an in-frame TGA codon corresponding with a selenocysteine residue in the peptide. Since the cDNA clone was not full-length, a second rat liver cDNA library was screened with the partial clone and a full-length cDNA clone was obtained. The nucleotide sequence of the full-length cDNA clone contained ten in-frame TGA codons. Two of the ten TGA codons were confirmed to be selenocysteine.

*To whom correspondence should be addressed.