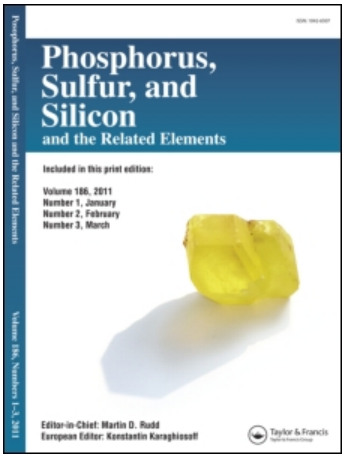

Phosphorus, Sulfur, and Silicon and the Related Elements
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713618290

Eukaryotic Selenocysteine Incorporation: Mechanistic Insights

Glover W. Martin Iii; Marla J. Berry

To cite this Article Martin Iii, Glover W. and Berry, Marla J.(1998) 'Eukaryotic Selenocysteine Incorporation: Mechanistic Insights', Phosphorus, Sulfur, and Silicon and the Related Elements, 136: 1, 308 — 320
To link to this Article: DOI: 10.1080/10426509808545957
URL: http://dx.doi.org/10.1080/10426509808545957

Biochemistry and Natural Products: Invited Lectures

Phosphorus Sulphur and Silicon
Vols. 136, 137 & 138 pp. 309–320
Reprints available directly from the publisher
Photocopying permitted by license only

EUKARYOTIC SELENOCYSTEINE INCORPORATION: MECHANISTIC INSIGHTS

GLOVER W. MARTIN, III[a] and MARLA J. BERRY[a,b]

[a]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts, 02115 USA; [b]Thyroid Division, Brigham and Women's Hospital, Boston, Massachusetts, 02115 USA

Cotranslational insertion of selenocysteine occurs in both prokaryotic and eukaryotic organisms. For the most part, what is known about this process in eukaryotes has been derived from extensive analysis of the RNA requirements for selenocysteine incorporation. Studies which have defined the selenocysteine insertion sequence (SECIS) element, identified its critical sequence and structural motifs, and established constraints on its spatial relationship with the UGA selenocysteine codon, have given mechanistic insights into this critical process.

Keywords: selenocysteine; translation; RNA structure; SELB

INTRODUCTION

Recent years have witnessed the growth of a small, yet very important, class of proteins. The defining characteristic of proteins of this class is the presence of one or more selenocysteine residues, the incorporation of which is specified by UGA codons. Table 1 lists the eubacterial and eukaryotic selenoproteins identified to date, as well as their functions, if known. Sequences encoding hydrogenases and dehydrogenases, like those found in

eubacteria, have also recently been identified in the archaeobacterium *Methanococcus jannaschi.*[1] Overwhelmingly, selenoenzymes catalyze redox-like reactions, in which the selenocysteine residue(s) plays a critical role. The ability to synthesize these selenoproteins, while not needed for viability in *E. coli*, is predicted to be is essential for proper development and homeostasis in vertebrates.

TABLE 1: Selenoproteins and known or proposed functions

EUKARYOTES	
Glutathione peroxidase	ROOH→ ROH + H_2O; protection from peroxide damage
Cytoplasmic	
Plasma	
Phospholipid hydroperoxide	
Gastrointestinal	
Iodothyronine deiodinases	
Types 1 and 2	T4→ T3; thyroid hormone activation
Type 3	T3→ T2, T1, T0; thyroid hormone deactivation
Thioredoxin reductase	provides reducing equivalents for ribonucleotide reductase catalyzes proper folding of NFκB
Selenophosphate synthase	Se + ATP→Se~P + ADP Provides active selenium donor for selenocysteine biosynthesis
Selenoprotein P	plasma selenoprotein of unknown function contains 10-12 UGA codons
Selenoprotein W	muscle selenoprotein of unknown function
PROKARYOTES	
Formate dehydrogenase	HCOOH→ CO_2 + H_2O
Ni-Fe-Se hydrogenase	H_2 → 2H + 2e
Glycine reductase	glycine→ acetate + NH_4^+

The phenomenon of cotranslational incorporation of selenocysteine into polypeptides follows two paradigms, of which one is much better understood than the other. In eubacteria, the process requires the action of four gene products[2,3]. Two of these are necessary for the synthesis of selenocysteine, and a third is the selenocysteine-specific tRNA ($tRNA^{[ser]sec}$). The fourth gene product, SELB, is a specialized and specific translation elongation factor which mediates the incorporation of the amino-acid during translation. On the RNA side, selenocysteine incorporation in *E. coli* requires a hairpin structure within the protein-coding region, which is constrained to lie immediately downstream of the requisite UGA codon.[4,5]

The availability of genetic methods in *E. coli* greatly accelerated the pace of discovery soon after the identification of selenoproteins in this

The availability of genetic methods in *E. coli* greatly accelerated the pace of discovery soon after the identification of selenoproteins in this organism. Unfortunately, no such system has been available in higher eukaryotes, significantly lengthening the data-gathering process. To date, the eukaryotic counterpart of only one of the enzymes in the selenocysteine biosynthetic pathway has been identified, a serendipitous discovery as an artifact of a two-hybrid screen for proteins associated with splicing factors[6]. Due to anticodon identity and structural similarity with their *E. coli* counterpart, many eukaryotic $tRNA^{[ser]sec}$ homologs have also been identified.[7]

However, it is the discovery of a eukaryotic selenocysteyl-tRNA-specific elongation factor, or SELB homolog, which is anxiously awaited. In *E. coli*, this specialized EF-Tu analog binds both to $tRNA^{[ser]sec}$ and the mRNA hairpin structure immediately adjacent to the UGA codon. The prevailing model holds that it is through this proximity of ribosome, elongation factor and tRNA that incorporation of selenocysteine is achieved at the appropriate point in the peptide elongation cycle (see below).[7] In contrast, the secondary structural motifs required in eukaryotes, the selenocysteine insertion sequence (SECIS) elements, lie at significant and variable distances from the selenocysteine UGA codon.[8,9] This difference between the eubacterial and eukaryotic systems is highly suggestive of fundamentally distinct mechanisms of selenocysteine incorporation. Concurrent with the search for a SELB homolog, a number of studies have been carried out in our lab and in others to better characterize the SECIS element, obtain detailed information about its structure and identify parameters which affect its function. This review will summarize these studies and the insights they have provided about the molecular mechanism of selenocysteine incorporation in eukaryotes.

RNA REQUIREMENTS

The Selenocysteine Insertion Sequence Element

Due to the lack of a genetic system, elucidation of the mechanism of selenocysteine incorporation in eukaryotes has proceeded from a different origin. Much work has been done to understand the mRNA requirements for eukaryotic selenocysteine incorporation. After the discovery that a UGA codon could, under precise circumstances, specify selenocysteine during translation, further studies demonstrated the requirement for *cis*-acting sequences in the 3'UTR. Our initial studies identified SECIS elements in the type 1 iodothyronine deiodinase (D1) and cellular glutathione peroxidase (GPX) mRNAs,[8] and subsequently in the selenoprotein P (sel P) mRNA.[9] We showed that the 3'UTRs from these three mammalian selenoprotein mRNAs are functionally interchangeable in directing selenocysteine incorporation[8,9]. Folding and alignment programs revealed conserved primary sequence and secondary structural features among SECIS elements, and mutagenesis studies have shown that the conserved nucleotides are critical for function (Table 2)[9,10]. These studies provided the first insights into the requirements for selenocysteine incorporation in eukaryotes.

Gross anatomical studies of the D1, GPX and sel P SECIS elements utilized ~175 - 230 nucleotide segments of the 3'UTRs.[8,9] However, if the mechanism of selenocysteine incorporation in eukaryotes involves recognition of the RNA structures by a specific protein, this protein is likely to recognize a much smaller region of these structures, based on current understanding of RNA-protein interactions. Therefore, we recently sought to define the minimal functional D1 SECIS element. We examined the effects of progressive deletions from the 5' side of the SECIS element on ability to direct selenocysteine incorporation[11]

TABLE 2: Alignment of SECIS elements from eukaryotic selenoproteins. Abbreviations: D1, type 1 deiodinase; D3, type 3 deiodinase; cGPX, cytoplasmic glutathione peroxidase; PLGPX, plasma GPX; PHGPX, phospholipid hydroperoxide GPX; GPX-GI, gastrointestine-specific GPX; S. mans., *Schistosoma mansoni*; M. cont., *Molluscum contagiosum*; SELP, selenoprotein P; SELW, selenoprotein W; Thiored. red., thioredoxin reductase. Invariant nucleotides are underlined.

D1	AUGA	Y	GGYCACAGYNU	AAA	GYNCNCACRGCUGUGACU	U	GA	U
D3	AUGA	C	GANCCGCCUCU	AAC	UGGGCUUGACCACGGGCNGGNUC	Y	GA	N
GPX	AUGA	N	GGUGUUYCCUCU	AAA	YYURCRNGGAGRAAYRCC	U	GA	U
PLGPX	AUGA	R	GGAGGGGCCCCR	AAG	CCCUUGUGGGCGGRCCUCCCC	U	GA	G
PHGPX	AUGA	C	RGUCUGCCURA	AAA	CCAGCCCNNUGGUGGGGCNAGNCY	C	GA	G
GPX-GI	AUGA	U	GCACCUUCCU	AAA	CCCUCAUGGGUGGUGUC	U	GA	G
S.mans. GPX	AUGA	C	GAUGGCAGUCUC	AAA	UGUUCAUUGGUUGCCAUU	U	GA	U
M.cont. GPX	AUGA	C	GGCGUCUCUCG	AAC	ACCGACAAGGAGGGAGAGCUGCC	C	GA	G
SELP1	AUGA	G	AAYAGAAACRU	AAA	CUAUGACCUAGGGGUUUNCUGU	G	GA	U
SELP2	AUGA	Y	GGUUUAAUAGRN	AAA	CYRARYCCUAURAACC	U	GA	N
SELW	AUGA	Y	AGGAAGGACUG	AAA	GUCUYNNRGACNYNUGGUCYUUCYY	Y	GA	U
Thiored. red.	AUGA	A	GUCACCAGUCUC	AAG	CCCAUGUGGUAGGCGGUGAU	G	GA	A

Strikingly, a distinct 5' boundary between full activity and loss of function was found to correspond precisely with the first base of the invariant AUGA sequence in the 5' arm of the stem.[11] The 3' boundary was mapped to the point directly across from the 5' boundary in the base-paired structure. Interestingly, although sequences below this point are not conserved in primary sequence, and therefore not required for function, the propensity of these nucleotides to base pair is inversely related to resultant function of the SECIS element. This data suggests that the ability of a SECIS element to adopt a higher order structure is crucial to its function, and is potentially required for the formation of a ligand for a specific RNA-binding protein.

Conserved sequence and structural features of the SECIS element

More detailed comparative analysis identified conserved sequence and structural features in the D1, GPX, and sel P SECIS elements. These include the sequence AUGA in the 5' arms of the stems, three consecutive A's in the loops, and a conserved GA dinucleotide in the 3' arms of the

stems (Table 2 and Figure 1A)[8-10,12]. These conserved nucleotides (shown in bold in Figure 1, underlined in Table 2) are present in all eukaryotic SECIS elements identified to date. The 3'UTRs of D1 and GPX each contain one SECIS element. Two functional elements are present in the sel P 3'UTR. The first of these directs selenocysteine incorporation at ~3-fold higher levels than either the second sel P or the D1 SECIS element[9]. Several recently identified selenoprotein sequences contain SECIS elements that are predicted to form alternative structures to those initially identified. These predicted structures (Figure 1B) contain the same conserved nucleotides as those originally identified, but the conserved A's are predicted to be present in an internal bulge rather than a terminal loop. Functional differences between these two SECIS element forms are yet to be revealed.

Computer folding programs predict the conserved nucleotides in the 5' and 3' arms, the AUGA and GA, to be unpaired[8,9]. However, chemical and enzymatic probing of the rat and human D1 and rat GPX SECIS elements recently demonstrated that this region is likely to be base-paired[12]. Based on these findings, the authors proposed a novel structural motif in this region, containing four non-Watson-Crick base pairs. Recently, through extensive mutational analysis, we obtained *in vivo* data that strongly corroborate the existence of such a structure[13]. Definitive evidence for the existence of this motif is likely to require the solution or crystal structure of SECIS elements.

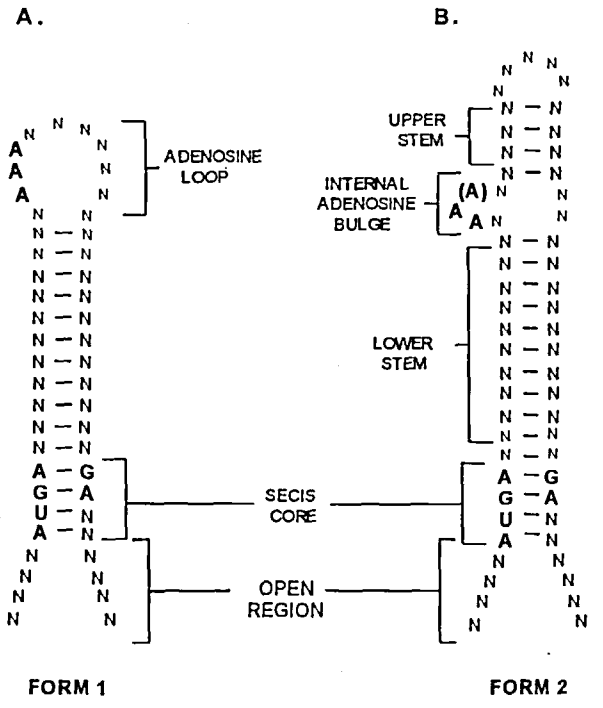

FIGURE 1: Proposed form 1 and form 2 SECIS element structures. A, Form 1 consensus (D1, GPX, selP loop 2, see figure 3 legend for abbreviations). B, Form 2 consensus (all other elements identified to date). Invariant nucleotides are shown in bold.

Effect of Relative Positions of SECIS Element and UGA Codon

In our initial studies of SECIS element function, we examined the effects of altering the spacing between the D1 SECIS element and UGA selenocysteine codon. We found that increasing this spacing from the native 1.2 kb to ~2.7 kb produced wild-type levels of selenocysteine incorporation[9], whereas decreasing the spacing to ~0.6 kb increased incorporation by 20 to 40%[8]. Although the spacing between selenocysteine codons and SECIS elements varies in nature and has been experimentally manipulated over a considerable range, the minimal spacing

allowed for function had not been examined until recently. Nearly all eukaryotic selenoproteins described thus far utilize UAA or UAG as their termination codons. Two exceptions are a small selenoprotein found in skeletal muscle, selenoprotein W (sel W)[14], and *Schistosoma mansoni* GPX[15]. These mRNAs each contain two in-frame UGA codons, the first of which encodes selenocysteine, the second acting as a stop codon. The UGA termination codons are located 55 or 38 nucleotides, respectively, from the SECIS elements. Although this arrangement of a UGA codon in proximity to the SECIS element is reminiscent of the situation in prokaryotes, it appears to be non-functional in eukaryotes. In fact, the close proximity apparently precludes SECIS function for the second UGA codons.

To determine if spacing, *per se*, is the reason for lack of selenocysteine incorporation at the SECIS proximal UGA codons, we set out to define the minimal UGA-SECIS spacing that would allow selenocysteine incorporation. DNA constructs were generated with progressive reduction in the spacing between the D1 TGA and SECIS element, and selenocysteine incorporation assessed following transient transfection. In these experiments, we found that a spacing of 111 nucleotides between UGA and SECIS element was sufficient to promote selenocysteine incorporation, but 51 nucleotides was not[11], consistent with the non-permissive spacing between the second UGA codons and SECIS elements in sel W and *S. mansoni* GPX.

IMPLICATIONS FOR THE MECHANISM OF SECIS ELEMENT FUNCTION

Eukaryotic selenoproteins cannot be expressed in bacterial systems, nor *vice versa* (our unpublished observations)[16], highlighting the mechanistic differences between the two pathways. As has been previously discussed, the RNA requirements for selenocysteine incorporation in prokaryotes and eukaryotes differ in sequence, predicted structure and position within the mRNA. This last distinction is potentially the most critical, and is the source of the proposed mechanistic differences between the eubacterial and eukaryotic implementations of selenocysteine incorporation.

Most important, due to the location of the SECIS element within the coding region, translation of prokaryotic UGA codons would require subsequent melting of the stem-loop structures, presumably either by helicase activities associated with the ribosome or by other unidentified unwinding activities. While the proximity and effective tethering of the prokaryotic quaternary complex (tRNA-SELB-GTP-stem loop) to the UGA at the A site on the ribosome would be predicted to result in highly efficient selenocysteine incorporation in prokaryotes as compared to eukaryotes, subsequent melting of the stem with each ribosomal passage might have the opposite effect. Therefore, ribosomes occupying the region of the melted stem-loop would be expected (at least transiently) to prevent refolding, and consequently, selenocysteine incorporation by upstream ribosomes. Our initial experiments have shown that the activity of the D1 SECIS element is reduced by more than 90% when placed within the open reading frame, suggesting that selenoprotein synthesis in eukaryotes may be much more efficient than in *E. coli.*[11]

A related implication is that, although not known to exist, a eubacterial selenoprotein with multiple UGA selenocysteine codons would require a stem-loop for each one. Presumably this arrangement would multiply the inefficiency of selenocysteine incorporation at each UGA, resulting in an

extremely low yield of the full-length polypeptide. In contrast, mammalian selenoproteins containing multiple UGA codons occur naturally, the prototype for which is sel P. In this case, selenocysteine incorporation at up to twelve UGA codons is driven by two SECIS elements in the 3'UTR.[17] In addition, we showed that eukaryotic SECIS elements are able to direct selenocysteine incorporation at any upstream UGA, and that a single SECIS element can function for multiple UGAs, unlike the prokaryotic situation where UGA-stem loop context is critical. Finally, the necessity in *E. coli* for a secondary structural element containing conserved nucleotides to lie immediately downstream of the UGA selenocysteine codon in the open reading frame places some degree of constraint on the amino acids that may be coded for in this region. Moving the SECIS beyond the open reading frame has removed this constraint in eukaryotes.

CONCLUSION

Figure 2 illustrates the current model of selenocysteine incorporation in eukaryotes and summarizes the data that have been presented here. The SELB-tRNA$^{[ser]sec}$ complex binds the SECIS element, most likely interacting directly with the two regions of nucleotide invariance.

We hypothesize that the flexible region below the SECIS element facilitates the assumption of a functional orientation. We previously proposed that the mRNA between the UGA codon and the SECIS element is looped out to allow the direct interaction of the SELB-tRNA$^{[ser]sec}$-SECIS element with the ribosome approaching the UGA codon.
Our minimal spacing studies suggest a steric interaction of SELB and the ribosome that precludes function at close distances. Alternatively, close proximity may also prevent SELB from adopting an orientation which allows it to function. If any of the RNA requirements for selenocysteine incorporation are not met, release factor will bind to the ribosome when it

encounters the UGA codon and mediate termination of protein synthesis and release of the peptide. For comparison, the prokaryotic model of selenocysteine incorporation is also depicted.

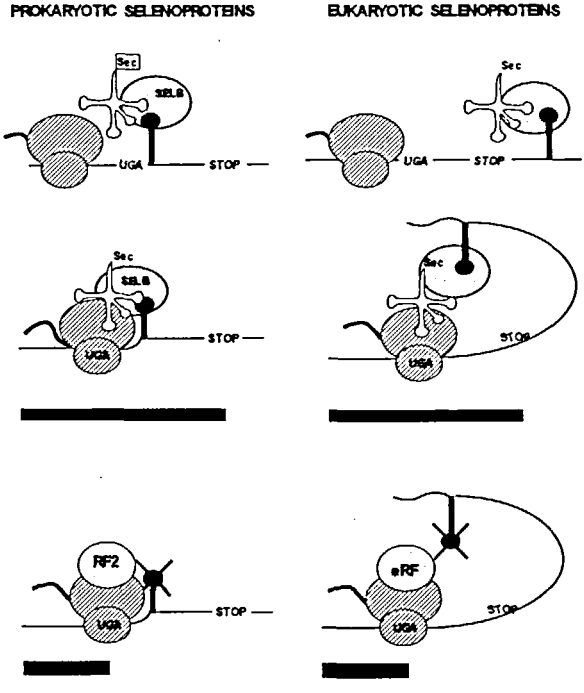

FIGURE 2: Proposed mechanisms of selenoprotein synthesis in prokaryotes (adapted from ref. 7), and eukaryotes (adapted from ref. 18). Hatched ovals - ribosomes; shaded ovals - selenocysteine specific elongation factors or release factors; solid bars - translated proteins; sec - selenocysteine charged to tRNA

The dissection of the RNA requirements for selenocysteine incorporation in eukaryotes has yielded enough details for the construction of this model. However, the critical unknown is the identity of SELB homolog. Once identified and cloned, its interaction with the SECIS element, $tRNA^{[ser]sec}$, and ribosome can be investigated in detail. Perhaps

most importantly, a much clearer understanding of the logistics of this process at the molecular level can then be achieved.

References

1. R. Wilting, S. Schorling, B. C. Persson, A. Böck, J Mol Biol **266**, 637 (1997).
2. W. Leinfelder, K Forchammer, F. Zinoni, G. Sawers, M.-A. Mandrand-Berthelot, A. Böck, J. Bacteriol. **170**, 540 (1988)
3. A. Böck, K. Forchammer, J. Heider, W. Leinfelder, G. Sawers, B. Verpek, F. Zinoni, Molec. Microbiology **5**, 515 (1991).
4. F. Zinoni, J. Heider, A. Böck, Proc. Natl. Acad. Sci. USA **87**, 4660 (1990).
5. G.-F. T. Chen, L. Fang, M. Inouye, J. Biol. Chem. **268**, 23128 (1993).
6. S. C. Low, J. W. Harney, M. J. Berry, J. Biol. Chem. **270**, 21659 (1995).
7. C. Baron, A. Böck, in tRNAs: Structure, Biosynthesis, and Function, D. Söll and U. RajBhandary, Eds. (American Society for Microbiology, Washington, DC, 1995), p. 529.
8. M. J. Berry, L. Banu, Y. Chen, S. J. Mandel, J. D. Kieffer, J. W. Harney, P. R. Larsen, Nature **353**, 273 (1991).
9. M. J. Berry, L. Banu, J. W. Harney, P. R. Larsen, EMBO J **12**, 3315 (1993).
10. Q. Shen, J. L. Leonard, P. E. Newburger, RNA **1**, 519 (1995).
11. G. W. Martin, J. W. Harney, M. J. Berry, RNA **2**, 171 (1996).
12. R. Walczak, E. Westhof, P. Carbon, A. Krol, RNA **2**, 367 (1996).
13. G. W. Martin, III, J. Harney, M. J. Berry, RNA (in press).
14. S. C. Vendeland, M. A. Beilstein, J-Y. Yeh, W. Ream, P. D. Whanger, Proc Natl Acad Sci USA **92**, 8749 (1995).
15. D. L. Williams, R. Pierce, E. A Cookson, A. Capron, Mol. Biochem. and Parasitology **52**, 127 (1992).
16. C. Rocher, C. Faucheu, F. Hervé, C. Bénicourt, J. L. Lalanne, Gene **98**, 193 (1991).
17. K. E. Hill, R. S. Lloyd, R. F. Burk, Proc Natl Acad Sci USA **90**, 537 (1993).
18. S. C. Low, M. J. Berry, Trends Biochem. Sci. **21**, 203 (1996).