Chapter 1
Selenocysteine tRNA$^{[Ser]Sec}$: From Nonsense Suppressor tRNA to the Quintessential Constituent in Selenoprotein Biosynthesis

Bradley A. Carlson, Byeong Jae Lee, Petra A. Tsuji, Ryuta Tobe, Jin Mo Park, Ulrich Schweizer, Vadim N. Gladyshev, and Dolph L. Hatfield

Abstract When selenocysteine (Sec) tRNA$^{[Ser]Sec}$ was originally discovered, it was proposed to be the first nonsense suppressor tRNA found in mammalian and avian tissues, since it exclusively decoded the nonsense codon, UGA, which normally dictates the cessation of protein synthesis. This tRNA was subsequently shown to be Sec tRNA, which inserted Sec into protein as the 21st proteinogenic amino acid. Once it was established that this tRNA was aminoacylated with serine by seryl-tRNA synthetase and served as the scaffold for Sec synthesis, Sec tRNA was appropriately named Sec tRNA$^{[Ser]Sec}$. The mammalian Sec-tRNA$^{[Ser]Sec}$ population consists

B.A. Carlson (✉) • D.L. Hatfield (✉)
Molecular Biology of Selenium Section, Mouse Cancer Genetics Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
e-mail: carlsonb@mail.nih.gov; hatfield@mail.nih.gov

B.J. Lee
School of Biological Sciences, Seoul National University, Seoul, South Korea

P.A. Tsuji
Department of Biological Sciences, Towson University, 8000 York Road, Towson, MD 21252, USA

R. Tobe
Department of Biotechnology, College of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

J.M. Park
Cutaneous Biology Research Center, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA, USA

U. Schweizer
Institut für Biochemie und Molekularbiologie, Rheinische Friedrich-Wilhelms-Universität Bonn, Nussallee 11, 53115 Bonn, Germany

V.N. Gladyshev
Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

D.L. Hatfield et al. (eds.), *Selenium*, DOI 10.1007/978-3-319-41283-2_1

of two isoforms that differ from each other by a single 2′-*O*-methyl moiety on the uridine at position 34, designated Um34. The non-Um34 isoform is involved in the synthesis of a subclass of selenoproteins, called housekeeping selenoproteins, while the Um34 isoform supports synthesis of stress-related selenoproteins. These novel functions and other unique features of Sec tRNA are the subjects of this chapter, supporting the idea that this tRNA is the quintessential constituent in selenoprotein biosynthesis.

Keywords Housekeeping selenoproteins • Selenocysteine • Selenocysteine tRNA • Selenoprotein biosynthesis • Selenoproteins • Stress-related selenoproteins • UGA codon

1.1 Introduction

Selenocysteine (Sec) tRNA was discovered in 1970 when a seryl-tRNA was found to form phosphoseryl-tRNA [1] and decode exclusively the UGA codon [2] in mammalian and avian livers. Subsequently, phosphoseryl-tRNA and the UGA-decoding seryl-tRNA were found to be the same tRNA [3]. However, since the UGA-decoding seryl-tRNA suppressed the UGA termination codon in rabbit β-globin mRNA [3], its earlier proposal as a nonsense suppressor tRNA was retained until the biosynthesis of Sec was shown to occur on this tRNA in prokaryotes and eukaryotes identifying it as Sec-tRNA$^{[Ser]Sec}$ [4, 5]. From its discovery in 1970 until the finding that it is an indispensable component in selenoprotein biosynthesis, many unusual features of this tRNA were reported. Two essential cellular functions of Sec tRNA are to synthesize Sec, the 21st amino acid in the genetic code, using tRNA$^{[Ser]Sec}$ as the scaffold upon which the synthesis occurs (discussed in Chap. 4) and donate Sec to protein in response to UGA Sec codons in selenoprotein mRNAs (discussed in Chaps. 2 and 5). Since the synthesis of selenoproteins is dependent on Sec tRNA$^{[Ser]Sec}$, manipulation of the expression of Sec tRNA has been used to modulate selenoprotein expression in assessing the role of selenoproteins in health and development (discussed in Chap. 46).

1.2 Primary and Secondary Structures of Sec tRNA

Since its discovery as a seryl-tRNA [1, 2], Sec tRNA was shown to occur in very small amounts compared to the corresponding total seryl-tRNA population in all tissues and cells where it was detected (reviewed in [6]). This tRNA existed in two isoforms [7], which differed by a single 2′-*O*-methylribosyl at position 34, designated Um34. The highly modified base at position 34 is 5-methoxycarbonylmethyluridine

(mcm^{5}U), while the nucleoside at this position is 5-methoxycarbonylmethyl-2′-*O*-methyluridine (mcm^{5}Um). Interestingly, the mammalian isoforms are 90 nucleotides long making them the longest tRNAs sequenced in higher vertebrates. Another unique feature about these tRNAs is that they have relatively few modified bases, and thus, are highly undermodified compared to all other known tRNAs that normally contain 15–17 amended bases.

The secondary structures of the two isoforms are shown in Fig. 1.1 in a clover leaf model along with the four modified bases, mcm^{5}U, *N*6-isopentenyladenosine (i^{6}A), pseudouridine (ψ), and 1-methyladenosine (m^{1}A) at positions 34, 37, 55, and 58, respectively [7]. The nucleoside, mcm^{5}Um, is also shown. Unlike any other known tRNA, Sec tRNA exists in a novel clover leaf form having 13 bases in the acceptor and TψC stems compared to 12 in other tRNAs and the acceptor stem contains nine bases, while the TψC stem contains four. Thus, the tRNA resides in a 9/4 cloverleaf form compared to a 7/5 form in other tRNAs [8, 9]. The D-stem of Sec tRNA has more base pairs, five to six, than all other tRNAs, which have three to four. Furthermore, Sec tRNA does not have the dihydrouracil base found in the D-loop in other tRNAs. The long variable arm and the extra base in the acceptor/TψC stems account for the bases that make this tRNA much longer than canonical tRNAs. When comparing Sec tRNA to all other tRNAs, it is indeed the most unique adaptor RNA described to date. These features account for the inability of elongation factors TU or 1alpha to bind tRNA[Ser]Sec and the requirement instead for dedicated elongation factors SelB and EF-SEC in bacteria and eukaryotes, respectively.

Sec tRNA is rightfully named Sec tRNA[Ser]Sec, since it is initially aminoacylated with serine (Ser) by seryl-tRNA synthetase (SERS); but as a result of synthesizing Sec from Ser on the tRNA, it inserts Sec into protein. Historically, tRNAs were named by the amino acid attached to them by their corresponding aminoacyl-tRNA synthetase, and unlike any other known tRNA in eukaryotes, Sec is synthesized directly on its tRNA. The novelty of these events are highlighted by the uniqueness of its name, tRNA[Ser]Sec.

1.3 Um34 Addition to Sec tRNA[Ser]Sec, a Most Highly Specialized Modification

In the maturation of Sec tRNA[Ser]Sec, the final modification is the addition of Um34, which is indeed a highly specialized event. For example, its synthesis depends on the secondary and tertiary structure of Sec tRNA [10], and its inclusion in the tRNA has a dramatic impact on tertiary structure of the molecule [7]. The synthesis is dependent on selenium status [7, 11, 12] and its presence governs stress-related selenoprotein synthesis [13, 14]. The extent of Um34 modification is dependent on selenium status that in turn governs the magnitude of stress-related

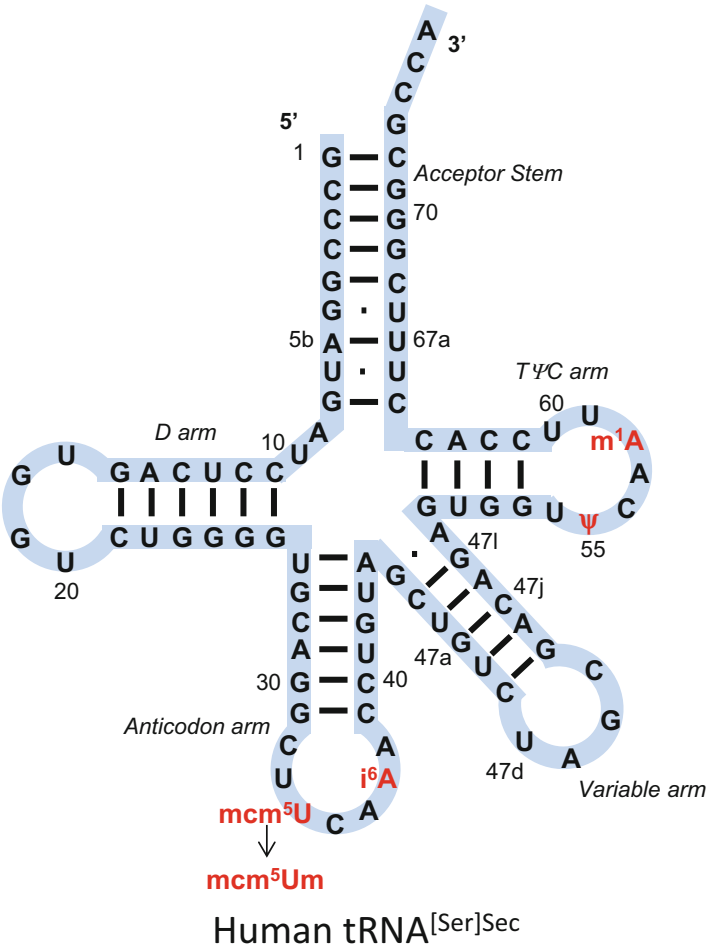

Fig. 1.1 The primary structure of human tRNA[Ser]Sec is shown in a cloverleaf model. There are 90 bases in mammalian tRNA[Ser]Sec and the bases are numbered as shown in the figure (see also [9]). The acceptor stem constitutes the paired 5′ and 3′ terminal bases, the D stem and loop constitute the six paired and four unpaired bases of the left portion of the tRNA, the anticodon stem and loop, the six paired and seven unpaired bases of the lower portion of the tRNA, the variable stem and loop, the five paired and four unpaired bases, and the TψC stem and loop, the four paired and seven unpaired bases of the right portion of the tRNA. Mammalian tRNA[Ser]Sec contains base modifications at positions 34 (mcm^5U), 37 (i^6A), 55 (ψ) and 58 (m^1A) as described in the text. The two isoforms differ from each other by a single methyl group on the 2′-*O*-ribosyl moiety at position 34

selenoprotein expression; i.e., the higher the selenium level, the greater the degree of mcm^5Um occurrence and thus the enhanced expression of stress-related selenoproteins [7, 11, 12].

An unresolved question regarding Um34 addition to the mcm^5U isoform is when exactly does this addition occur during Sec tRNA[Ser]Sec maturation. Synthesis of the

modified bases, Ψ and m^{1}A at positions 55 and 58, respectively, occurs in the nucleus, while the synthesis of mcm^{5}U occurs in the cytoplasm [15, 16]. However, base analysis following injection of the tRNA[Ser]Sec transcript into *Xenopus* oocytes revealed that the i^{6}A modification at position 37 was present in tRNA[Ser]Sec both with and without Um34 [15]. Since only the mcm^{5}Um isoforms, mcm^{5}U and mcm^{5}Um, were found in cells and tissues, we proposed that i^{6}A has to be present on the tRNA for Um34 synthesis to take place and i^{6}A addition was reversible [17]. The fact that tRNA[Ser]Sec apparently must have an amino acid attached, presumably Sec, prior to Um34 addition suggested that the corresponding methylase is highly specific for Sec-tRNA[Ser]Sec and that the i^{6}A modification must be present for Um34 addition. The enzyme that adds i^{6}A, isopentenyl transferase 1 (TRIT1), has been characterized in mammalian cells, and although the reversibility of this enzyme has not been characterized *per se*, it seems highly unlikely that the reaction is sufficiently reversible [18] to accommodate the earlier proposal that addition of the i^{6}A base is reversible (see above and [17]). Sec-tRNA[Ser]Sec must have both Um34 and i^{6}A modifications for stress-related selenoprotein synthesis to occur as the isoform lacking i^{6}A cannot express this subclass of selenoproteins [13, 14]. Thus, it appears that i^{6}A must be present for Um34 synthesis in vivo. These observations raise questions as to where in the cell and at what stage during tRNA[Ser]Sec maturation do Um34 synthesis take place in vivo, if the TRIT1 reaction is poorly reversible and Um34 cannot be synthesized on mcm^{5}U lacking i^{6}A?

Sec-tRNA[Ser]Sec$_{mcm5Um}$ is essential for stress-related selenoprotein synthesis and the presence of Um34 [19] and i^{6}A modifications (see [20] and references therein) enhance the accuracy of codon:anticodon recognition. It should also be noted that mcm^{5}U synthesis at position 34 on tRNA was reported in *Arabidopsis* [21], while the mammalian tRNA methyltransferase, ALKBH8, along with an accessory protein, TRM112, have been found to perform the final methylation step in formation of mcm^{5}U in tRNA[Ser]Sec [22]. ALKBH8 knockout mice lacked the Um34 modification and had reduced GPx1 expression, indicating that mcm^{5}U formation is necessary for Um34 synthesis [22].

1.4 *Trsp*, the Sec tRNA[Ser]Sec Gene

The Sec tRNA[Ser]Sec gene is designated *Trsp*. It is a single copy gene in all organisms examined except zebrafish, which has two gene copies [23]. An RT-PCR technique was developed for identifying and sequencing tRNAs[Ser]Sec from lower eukaryotes [24] and the corresponding tRNAs from *Chlamydomonas reinhardtii* [24] and *Dictyostelium discoideum* and *Tetrahymena thermophila* [25] were sequenced. The longest eukaryotic Sec tRNA[Ser]Sec gene sequenced to date was from *Plasmodium falciparum*, which is 93 nucleotides in length [26, 27]. *Trsp* is present on chromosome 19 in humans [28] and chromosome 7 in mice [29].

1.5 Transcription of *Trsp*

Trsp is transcribed by Pol III as are all other eukaryotic tRNAs (reviewed in [17]), although *Trsp* in *Trypanosoma brucei* has been reported to be transcribed by Pol II [30]. The transcription of *Trsp*, however, is governed largely by three upstream regulatory elements, a *TATA*-like box that resides between positions about −20 and −35, a proximal sequence element (*PSE*) that resides between positions about −46 and −66, and a distal sequence element (*DSE*) that resides between positions about −195 and −210 [31, 32] (Fig. 1.2). On the other hand, transcription of canonical tRNAs is primarily dictated by two intragenic promoter elements, the A and B boxes. *Trsp* also encodes an intragenic A-like box and a B box [33], but the role of these two internal promoters in transcription of the Sec tRNA$^{[Ser]Sec}$ gene is largely not understood [34] and, for the most part, is still controversial [16, 17]. Transcription of *Trsp* begins at the first nucleotide within the gene, and thus, the transcript lacks a leader sequence found in all other tRNAs; however, it has a 3′-trailer sequence characteristic of canonical tRNAs that is removed leaving a transcript consisting of 87 nucleotides in all animals examined, wherein the CCA terminus is then added to complete the tRNA sequence [35]. The 5′-triphosphate at the first nucleotide in Sec tRNA$^{[Ser]Sec}$ remains intact through maturation of the tRNA, but its possible role in Sec biosynthesis and incorporation into protein has not been resolved.

The *TATA*-like box is an essential element for *Trsp* transcription efficiency as observed in vivo in *Xenopus* oocytes [34, 36] and in vitro in HeLa cell extracts [36]. The TATA box and *PSE* encode the elements also found in other Pol III transcribed RNAs that serve as basal promoters for attachment of the regulatory factors governing their transcription [31], albeit variability in sequences exists in these regions and particularly in the *Trsp PSE*, even within different species (reviewed in [17]).

Of the three upstream regulatory regions governing *Trsp* transcription, the *DSE* has sustained more interest. It contains a SPH motif and an octamer sequence that constitute the *Trsp activator element* (*AE*) to which the Sec tRNA transcription activating factor (STAF) binds ([37] and see [38] and references therein). The *AE* is essential for optimal transcription in *Xenopus* oocytes [31], but this regulatory element does not function in *Xenopus* oocyte extracts [39]. A large fragment of DNA was inserted between the *PSE* and *DSE* in mice that resulted in embryonic lethality due to reduced expression of *Trsp* transcription and generation of selenoprotein transcripts [40]. STAF has been characterized in frogs and mice and shown to have

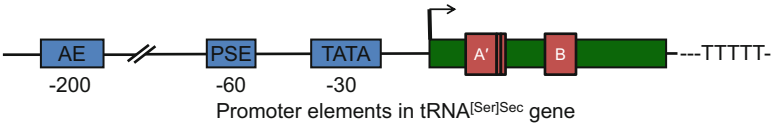

Fig. 1.2 Transcription elements governing *Trsp* transcription. The upstream and intragenic regulatory sites governing transcription of *Trsp* are discussed in the text. The downstream, –TTTTT- sequence, designates the termination signal that dictates the end of transcription leaving a 3′-trailer sequence that must be removed to yield the gene-only transcript

roles in numerous genes transcribed by RNA Pol II and III ([41, 42] and references therein). Mice carrying a deleted *AE* transgene (designated Δ*AE*^{r}) and also lacking *Trsp* (designated Δ*Trsp*) were dependent on the Δ*AE* transgenic mice for expression of tRNA[Ser]Sec [43]. Interestingly, these mice generated tRNA[Ser]Sec populations in dramatically varying amounts in different organs and tissues than their wild type counterparts. Some organs and tissues expressed more total tRNA[Ser]Sec, while others expressed less. One consistent observation in all of the organs and tissues examined was the amount of the mcm^{5}Um isoform was always less in Δ*AE*^{r}/Δ*Trsp* mice than wild type mice [43]. The finding of how the defective *AE* region influenced the levels of the tRNA[Ser]Sec population and in particular the level of mcm^{5}Um is intriguing and must require further work.

1.6 The Genetic Codeword for Sec tRNA[Ser]Sec is UGA

In 1986, two groups reported that TGA occurred in the open reading frames of the genes for mammalian glutathione peroxidase 1 (*Gpx1*) [44] and bacterial formate dehydrogenase (*fdhF*) [45] and their location coincided with Sec in the corresponding proteins. This was a surprising finding as UGA is used as a termination codon in protein synthesis in mammals and *Escherichia coli*. These findings provided the first evidence that the genetic code could contain a 21st amino acid.

Mutation of the *TGA* codeword in *fdhF* to any of a number of other codons resulted in the incorporation of the synonymous amino acid into formate dehydrogenase that in turn was not modified to Sec [46]. This observation further suggested that Sec was most likely the 21st proteinogenic amino acid. Since the minor phosphoseryl-tRNA in eukaryotes was found to specifically decode UGA [3], the possibility that phosphoserine was initially incorporated into selenoproteins and then post-translationally modified to Sec needed to be ruled out. When tRNA[Ser]Sec encoded in *selC* in *E. coli* and mammalian tRNA[Ser]Sec were shown to synthesize Sec on the corresponding tRNAs [4, 5], Sec was unequivocally demonstrated as the 21st amino acid in the genetic code.

1.7 Crystallization of tRNA[Ser]Sec

Numerous elegant studies have been carried out on three-dimensional structures of tRNA[Ser]Sec transcripts from a variety of sources including human [47–49], mice [50], archaea [51, 52] and bacteria [53, 54], either solely as a transcript [47, 50] or in a complex with a protein [48, 49, 51–54]. Although the transcript had to be mutated at several locations to achieve crystallization or to enhance tRNA[Ser]Sec: protein complex formation, these studies revealed important findings about the function of tRNA[Ser]Sec and its interaction with proteins involved in selenoprotein biosynthesis. A major limitation in these studies, however, was that naturally

occurring $tRNA^{[Ser]Sec}$ with its base modifications could not be used. The lack of i^6A in non-modified mouse $tRNA^{[Ser]Sec}$ leads to a disordered anticodon loop involving an illicit A_{37}-U_{33} base pair [50]. The secondary and tertiary structures of $tRNA^{[Ser]Sec}_{mcm5U}$ and $tRNA^{[Ser]Sec}_{mcm5Um}$ are known to be quite different [7] and crystal analyses of the two naturally occurring isoforms would most certainly provide novel insights into how they have such different roles in selenoprotein synthesis.

1.8 Concluding Remarks

A key role of Sec in protein function is certainly evidenced by the appearance and preservation of *Trsp* during evolution, the synthesis of Sec on $tRNA^{[Ser]Sec}$, the selenium-dependent Um34 methylation of $tRNA^{[Ser]Sec}_{mcm5U}$ giving rise to a second isoform, $tRNA^{[Ser]Sec}_{mcm5Um}$, the selective use of these isoforms in synthesizing two subclasses of selenoproteins, housekeeping and stress-related selenoproteins, the additional machinery used specifically in the incorporation of Sec from both isoforms into protein in response to UGA Sec codons and the generation of numerous selenoproteins used in many aspects of cellular metabolism involving health and development.

The appearance of selenoproteins containing selenium in the form of Sec likely occurred early in evolution, prior to the appearance of the three domains of life [55]. The reasons for preservation of Sec utilization during evolution most certainly reside in the fact that Sec is used in catalytic sites of redox-active proteins, thus being an integral and essential part of redox homeostasis. This raises an important question which is why then only some organisms take advantage of such a system? The chemical basis for the selection of Sec in proteins is discussed in Chap. 7.

There are several outstanding questions regarding Sec $tRNA^{[Ser]Sec}$ and its use in cellular metabolism. For example, what are the turnover rates of both isoforms in selenium deficient and replete conditions? What is the identity of the Um34 methylase that converts mcm^5U to mcm^5Um and does methylation occur in the cytoplasm or the nucleus? What are the governing factors responsible for the selective use of mcm^5Um in synthesizing stress-related selenoproteins? Is the $tRNA^{[Ser]Sec}_{mcm5U}$ isoform used exclusively for housekeeping selenoprotein synthesis or is it also used in stress-related selenoprotein synthesis? How does a defective *AE* regulatory region at the −200 position in mice cause such dramatic changes in the levels of the two Sec isoforms in different organs and tissues, always resulting in greater reduction in amounts of mcm^5Um compared to mcm^5U? What are the roles of the intragenic A and B boxes in *Trsp* in transcription? How do the crystal structures of naturally occurring $tRNA^{[Ser]Sec}_{mcm5U}$ and $tRNA^{[Ser]Sec}_{mcm5Um}$ differ from each other? What would their structures divulge about the functions of these two isoforms in interactions with proteins and in donation of the respective Sec to the nascent polypeptide chain (i.e., in decoding UGA)? Are there additional factors in the machinery involved in transcription of $tRNA^{[Ser]Sec}$, including novel ones? Assessing these issues will provide greater insight into the roles of the two Sec $tRNA^{[Ser]Sec}$ isoforms,

elucidate the enigma of how a 2′-*O*-hydroxylmethyl group can selectively govern the synthesis of an entire subclass of selenoproteins and pinpoint how selenium enrichment plays a role in the upregulation of Sec $tRNA^{[Ser]Sec}_{mcm5Um}$ and stress-related selenoprotein expression.

Acknowledgements This work was supported by the Intramural Research Program of the National Institutes of Health, NCI, Center for Cancer Research to DLH, NIH grants CA080946, GM061603 and GM065204 to VNG, DFG Priority Program 1784 to US (Schw914/5-1) and Towson University's Jess and Mildred Fisher College of Science and Mathematics to PAT, who is a Jess and Mildred Fisher Endowed Chair of Biological Sciences.

References

1. PH Maenpaa, MR Bernfield 1970 *Proc Natl Acad Sci U S A* 67:688
2. D Hatfield, FH Portugal 1970 *Proc Natl Acad Sci U S A* 67:1200
3. D Hatfield et al 1982 *Proc Natl Acad Sci U S A* 79:6215
4. BJ Lee et al 1989 *J Biol Chem* 264:9724
5. W Leinfelder et al 1989 *J Biol Chem* 264:9720
6. DL Hatfield et al 2006 *Prog Nucleic Acid Res Mol Biol* 81:97
7. AM Diamond et al 1993 *J Biol Chem* 268:14215
8. A Bock et al 1991 *Trends Biochem Sci* 16:463
9. C Sturchler et al 1993 *Nucleic Acids Res* 21:1073
10. LK Kim et al 2000 *RNA* 6:1306
11. HS Chittum et al 1997 *Biochim Biophys Acta* 1359:25
12. D Hatfield et al 1991 *Nucleic Acids Res* 19:939
13. BA Carlson et al 2007 *J Biol Chem* 282:32591
14. BA Carlson et al 2005 *J Biol Chem* 280:5542
15. IS Choi et al 1994 *Biochemistry* 33:601
16. C Sturchler et al 1994 *Nucleic Acids Res* 22:1354
17. DL Hatfield et al 1999 in *Comprehensive Natural Products Chemistry,* JW Kelly Ed (Elsevier Sc Ltd, Oxford) Vol 4 p 353
18. N Fradejas et al 2013 *Biochem J* 450:427
19. R Tobe et al 2013 *J Biol Chem* 288:14709
20. EM Gustilo et al 2008 *Curr Opin Microbiol* 11:134
21. V Leihne et al 2011 *Nucleic Acids Res* 39:7688
22. L Songe-Moller et al 2010 *Mol Cell Biol* 30:1814
23. XM Xu et al 1999 *FEBS Lett* 454:16
24. M Rao et al 2003 *RNA* 9:923
25. RK Shrimali et al 2005 *Biochem Biophys Res Commun* 329:147
26. AV Lobanov et al 2006 *Nucleic Acids Res* 34:496
27. BA Carlson et al 2006 in *Selenium: Its Molecular Biology and Role in Human Health,* DL Hatfield et al Eds (Springer Science+Business Media, LLC, New York) p 29
28. OW McBride et al 1987 *J Biol Chem* 262:11163
29. T Ohama et al 1994 *Genomics* 19:595
30. E Aeby et al 2010 *Nucleic Acids Res* 38:5833
31. E Myslinski et al 1992 *Nucleic Acids Res* 20:203
32. JM Park et al 1995 *Gene* 162:13
33. DL Hatfield et al 1983 *Proc Natl Acad Sci U S A* 80:4940
34. P Carbon, A Krol 1991 *EMBO J* 10:599

35. BJ Lee et al 1987 *Proc Natl Acad Sci U S A* 84:6384
36. BJ Lee et al 1989 *J Biol Chem* 264:9696
37. C Schuster et al 1995 *EMBO J* 14:3777
38. K Adachi et al 2000 *Biochem J* 346 Pt 1:45
39. JM Park et al 1996 *Biochem Biophys Res Commun* 226:231
40. VP Kelly et al 2005 *Mol Cell Biol* 25:3658
41. OA Barski et al 2004 *Genomics* 83:119
42. M Schaub et al 1999 *J Biol Chem* 274:24241
43. BA Carlson et al 2009 *Biochem J* 418:61
44. I Chambers et al 1986 *EMBO J* 5:1221
45. F Zinoni et al 1986 *Proc Natl Acad Sci U S A* 83:4650
46. F Zinoni et al 1987 *Proc Natl Acad Sci U S A* 84:3156
47. Y Itoh et al 2009 *Nucleic Acids Res* 37:6259
48. S Palioura et al 2009 *Science* 325:321
49. C Wang et al 2015 *Nucleic Acids Res*
50. OM Ganichkin et al 2011 *PLoS One* 6:e20032
51. S Chiba et al 2010 *Mol Cell* 39:410
52. RL Sherrer et al 2011 *Nucleic Acids Res* 39:1034
53. Y Itoh et al 2013 *Science* 340:75
54. Y Itoh et al 2013 *Nucleic Acids Res* 41:6729
55. VM Labunskyy et al 2014 *Physiol Rev* 94:739