Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 329 (2005) 147–151
BBRC
www.elsevier.com/locate/ybbrc

Selenocysteine tRNA identification in the model organisms *Dictyostelium discoideum* and *Tetrahymena thermophila*

Rajeev K. Shrimali[a,1], Alexey V. Lobanov[b,1], Xue-Ming Xu[a], Mahadev Rao[a,2], Bradley A. Carlson[a], Dana C. Mahadeo[c], Carole A. Parent[c], Vadim N. Gladyshev[b], Dolph L. Hatfield[a,*]

[a] *Molecular Biology of Selenium Section, Laboratory of Cancer Prevention, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA*
[b] *Department of Biochemistry, University of Nebraska, Lincoln, NE 68588, USA*
[c] *Laboratory of Cellular and Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA*

Received 22 December 2004

Abstract

Characterizing Sec tRNAs that decode UGA provides one of the most direct and easiest means of determining whether an organism possesses the ability to insert selenocysteine (Sec) into protein. Herein, we used a combination of two techniques, computational to identify Sec tRNA genes and RT-PCR to sequence the gene products, to unequivocally demonstrate that two widely studied, model protozoans, *Dictyostelium discoideum* and *Tetrahymena thermophila*, encode Sec tRNA in their genomes. The advantage of using both procedures is that computationally we could easily detect potential Sec tRNA genes and then confirm by sequencing that the Sec tRNA was present in the tRNA population, and thus the identified gene was not a pseudogene. Sec tRNAs from both organisms decode UGA. *T. thermophila* Sec tRNA, like all other sequenced Sec tRNAs, is 90 nucleotides in length, while that from *D. discoideum* is 91 nucleotides long making it the longest eukaryotic sequenced to date. Evolutionary analyses of known Sec tRNAs reveal the two forms identified herein are the most divergent eukaryotic Sec tRNAs thus far sequenced.

Keywords: Computational analysis; *Dictyostelium*; Protozoans; Selenium; Selenocysteine; Selenocysteine tRNA; *Tetrahymena*

The codeword UGA is generally used to dictate the termination of protein synthesis, but it also codes for selenocysteine (Sec) in many life forms within the eubacteria, archaea, and eukaryotic life kingdoms [1]. Thus, Sec is now acknowledged as the 21st amino acid in the genetic code. The tRNA for Sec is initially aminoacylated with serine by seryl-tRNA synthetase and the biosynthesis of Sec then occurs on its tRNA which is designated $tRNA^{[Ser]Sec}$. Transfer $RNA^{[Ser]Sec}$ is the only known tRNA that controls the expression of an entire class of selenium-containing proteins designated selenoproteins.

The number of selenoprotein genes varies considerably in different organisms. For example, in animals there is only one known selenoprotein gene encoded in the genome of *Caenorhabditis elegans*, while there are 24 in the genomes of rodents and 25 in humans [2]. Outside animals, selenoproteins and the selenoprotein synthetic machinery have been reported in another

* Corresponding author. Fax: +1 301 435 4957.
E-mail address: hatfield@dc37a.nci.nih.gov (D.L. Hatfield).
[1] These authors contributed equally to this project.
[2] Present address: Department of Oncology, Lombardi Cancer Center, Georgetown University Medical Center, New Research Building, Rm E320 3970 Reservoir Rd NW, Washington, DC 20057-1468, USA.

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.01.120

eukaryote, *Chlamydomonas reinhardtii*, a lower plant that encodes at least 10 selenoprotein genes [3]. The prokaryotes that utilize Sec for insertion into protein generally have few selenoprotein genes, and these are typically different from those found in eukaryotes [4].

Many life forms, however, do not have the apparatus for incorporating Sec into protein. For example, yeasts and higher plants do not utilize Sec in this manner. Of the 15 genomes sequenced in archaea, only three decode Sec, and among eubacteria, approximately one-third of the sequenced genomes possess Sec insertion machinery genes.

Sec tRNAs and their corresponding genes have been sequenced from a variety of organisms. All animal Sec tRNAs sequenced to date are 90 nucleotides in length and their genes are 87 nucleotides as the CCA terminus is added post-transcriptionally [1]. Sec tRNAs from vertebrates, insects, and lower animals have little sequence homology, demonstrating that this molecule has undergone substantial evolutionary change [5]. *C. reinhardtii* Sec tRNA has also been sequenced [5], which marked the first non-animal eukaryotic Sec tRNA sequenced. This tRNA, like its animal counterparts, is 90 nucleotides in length and it has little sequence homology to other Sec tRNAs.

It is of interest to determine the sequences of Sec tRNAs in a variety of organisms, not only because this molecule is the key molecule or central component (see [1] and references therein) in selenoprotein biosynthesis, but also because its identity can be used to assess how widespread the Sec insertion machinery is in nature. In addition, determination of many Sec tRNA structures will provide insights into how this unique molecule has evolved and will characterize its relationship to the evolution of the Sec insertion machinery as a whole. Most importantly, identification of organisms that do and do not encode the Sec insertion machinery will help us understand some of the major unresolved questions in selenium biology. For example, why and how have such a sophisticated means of inserting selenium into protein evolved and then once it evolved, why has this elaborate machinery only been differentially maintained?

In the present study, we identified Sec tRNA genes in two widely studied, model organisms, *Tetrahymena thermophila*, a freshwater ciliate, and *Dictyostelium discoideum*, a soil living amoeba, and sequenced the Sec tRNAs within the tRNA populations of these two organisms. Both tRNAs contain a UCA anticodon that decodes UGA. *T. thermophila* Sec tRNA is 90 nucleotides long, while *D. discoideum* Sec tRNA is 91 nucleotides in length.

Materials and methods

Reagents. Tetrahymena thermophila, strain B/IA107, was obtained from ATCC (Manassas, Virginia) and *D. discoideum*, AX3, was obtained from the *Dictyostelium* Stock Center (Columbia University, NY). ^{75}Selenious acid (190 Ci/mmol) was purchased from the Research Reactor facility (University of Missouri, Columbia, MO) and [γ-^{32}P]ATP, [^{3}H]serine (30 Ci/mmol), and *Ex taq* DNA polymerase PCR beads were from Amersham Biosciences. *E. coli* polymerase A, yeast polymerase A, reverse transcriptase (Superscript II), DH5α-competent cells, *Eco*RI, and the pCR-2.1-TOPO cloning kit were from Gibco/Invitrogen, and the PCR purification kit, QIAquick spin columns, and miniprep kits were from Qiagen. T4 DNA quick ligase was from New England BioLabs and T4 RNA ligase was from Promega. All other reagents and chemicals were of the highest grade available.

Growth and labeling of the cells. T. thermophila was grown at 25 °C in ATCC 357 Tetrahymena media (Difco Protease peptone, 5 g; tryptone, 5 g; K_2HPO_4, 0.2 g; Fe EDTA [0.003%], 0.03 g, and distilled water, 1 L) and *D. discoideum* was grown at 23 °C in HL-5 media as described [6]. Both of the above media were supplemented with 5×10^{-7} M sodium selenite. After harvesting, cells (~50 g of each species) were stored at −80 °C until ready for use. ^{75}Se labeling of 5 g of cells of each species cells was carried out as given [7,8] and labeled cells were stored at −80 °C until used.

Isolation and fractionation of tRNA. Total tRNA was isolated from 50 g *T. thermophila* cells (yield was 1579 total A_{260} units) and from 50 g *D. discoideum* cells (yield was 982 total A_{260} units) as described [7,8] and the tRNA was stored in water at −20 °C until ready for use. Similarly, total tRNA was isolated from 5 g of cells labeled with ^{75}Se (yields were 87 and 150 total A_{260} units from *T. thermophila* and from *D. discoideum*, respectively) as described [7,8]. ^{75}Se-labeled tRNAs were mixed with the respective unlabeled tRNA to a total of 250 A_{260} units, the resulting mixture chromatographed over a RPC-5 column [9] in a 0.425–0.725 M NaCl buffer containing 10 mM Mg^{2+}, the eluted fractions of [^{75}Se]Sec-tRNA monitored by counting each fraction in a Wallac gamma counter (Model 1470), and the A_{260} of each sample measured in a Hewlett-Packard UV/VIS Spectrophotometer (Model 8453), and the individual ^{75}Se-containing fractions were pooled and collected as described [7,8,10]. The collected fractions were deacylated [8] and then electrophoresed in a 17.5% polyacrylamide TBE–urea gel (20 cm × 20 cm × 1.5 mm in 2.0 cm wide wells with 0.5 A260/well). The band that migrated at the same position on developed gels as a purified sample of bovine liver Sec tRNA [11] was eluted from gels, the sample was prepared for sequencing [11,12], and sequencing has carried out as described below.

Computational identity, sequencing of Sec tRNAs, and Sec tRNA genes, and phylogenetic analysis. Sec tRNA genes were identified in the genomic sequences of *Tetrahymena* and *Dictylostelium* using an in-house Sec tRNA search program (A.V. Lobanov and V.N. Gladyshev, unpublished). *T. thermophila* and *D. discoideum* Sec tRNAs were sequenced from the samples purified above by 3′-end polyadenylation [13] followed by reverse transcription, 5′-end anchor ligation, PCR [14,15], and cloning and sequencing [5], but with modification as follows. PCR amplification was carried out separately for sequencing 5′- and 3′-ends of the Sec tRNAs. The 5′-ends of Sec tRNAs were obtained by using primers 5′-CGCTCCGGAGGGAGAATCGA-3′ and 5′-CGCGC GTCAGGTTGAATCGA-3′ designed from the gene sequences of *T. thermophila* and *D. discoideum* Sec tRNAs, respectively, and the adaptor primer, 5′-GGCAATTAACCCTCACTAAAG-3′ [5], to amplify the cDNA yielded from reverse transcription followed by 5′-end anchor ligation. The 3′-ends of Sec tRNAs were obtained by using primers 5′-GCTTTGGTGAGTTGGACTGG-3′ and 5′-GCGTGTTTGATCG CTATCTG-3′ designed from the gene sequences of *T. thermophila* and *D. discoideum* Sec tRNAs, respectively, and the primer, 5′-TTGAA TTCGCATTGAGCACCTGC-3′ [5], to amplify the cDNA yielded from 3′-end polyadenylation followed by reverse transcription. The PCR products were directly sequenced after 2% agarose gel purification using Qiagen gel purification kits.

Nucleotide sequence alignments used in making the phylogenetic tree of Sec tRNAs (Fig. 3) were generated using the PileUP program.

Distances between sequences were calculated from these alignments by the DISTANCE program with the Kimura distance measuring methods. Finally, unrooted phylogenic trees were visualized by the GrowTree program.

Results

Tetrahymena thermophila and *D. discoideum* cells were labeled with ^{75}Se and the resulting labeled tRNA populations were chromatographed over a RPC-5 column (Fig. 1). *T. thermophila* contained a single, large late-eluting [^{75}Se]Sec-tRNA peak (Fig. 1A). Labeled fractions were collected from the column as shown in the figure and after deacylation the resulting sample was prepared for sequencing as described below.

Dictyostelium discoideum cells contained a late-eluting [^{75}Se]Sec-tRNA peak and a front-running shoulder (Fig. 1B). To determine the sequence of Sec tRNA in *D. discoideum*, we pooled the labeled fractions as shown in Fig. 1B and prepared them for sequencing (see below). It should be noted that the sequencing technique

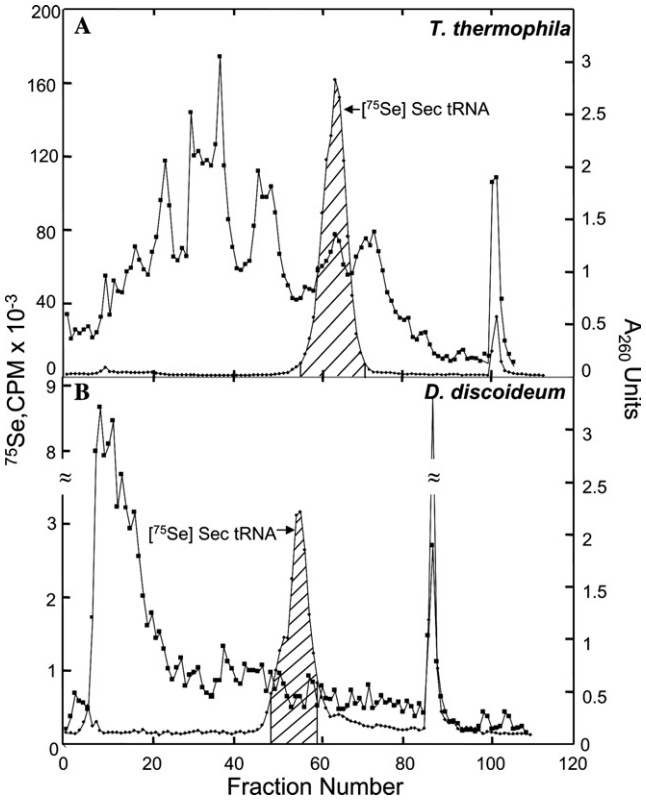

Fig. 1. Fractionation of *T. thermophila* and *D. discoideum* tRNA. *T. thermophila* tRNA (A) and *D. discoideum* tRNA (B) labeled with ^{75}Se were pooled with the respective unlabeled tRNA and chromatographed on a RPC-5 column as given in Materials and methods. Fractions containing [^{75}Se]Sec tRNA were pooled as shown by the hatched area, collected, and prepared for electrophoresis as described in Materials and methods.

used in this study does not distinguish between isoform variation in nucleoside modification unless the modification difference involved 1-methyladenosine (m^1A) (see below and [5]).

No Sec tRNA genes have been previously identified in *Tetrahymena* and *Dictyostelium*, and homology analyses using any known Sec tRNA gene failed to identify a corresponding gene in these organisms. We were able to identify candidate Sec tRNA genes, however, in the genomic sequences of *T. thermophila* and *D. discoideum* using a computational strategy that searched for Sec tRNA (Figs. 2A and B). The identification of both Sec tRNA sequences afforded us an opportunity to confirm their presence within the total tRNA population of each organism, provided the gene sequences were not pseudogenes.

Since we had probable Sec tRNA sequences (Fig. 2), we used a modified procedure for sequencing tRNAs than that used for sequencing Sec tRNA from *C. reinhardtii* [5]. The computational candidate sequences from the genomes of *T. thermophila* and *D. discoideum* were used as a guide to design primers for detection of the corresponding gene products within the tRNA populations. We used primers that partially overlapped either the 5′- or 3′-ends of each Sec tRNA during the PCR amplification step in order that tRNAs could be sequenced separately from both ends (see Materials and methods). In this manner, we could completely sequence both tRNAs and determine if their sequences matched the corresponding ones found by computational analysis. Indeed, the experimental sequences of both Sec tRNAs matched the predicted ones with the exception of U in place of A at position 81 in *T. thermophila* Sec tRNA and at position 82 in *D. discoideum* Sec tRNA (compare Figs. 2A and B gene sequence and tRNA sequence, respectively). Since we know from earlier studies that most Sec tRNAs contain 1-methyladenine (m^1A) at this position and m^1A produces an artifact of a U substituted for an A in the sequencing procedure (see [5] and references therein), we propose that this position contains an A.

The cloverleaf models of *T. thermophila* and *D. discoideum* Sec tRNAs are shown in a 9/4 arrangement (i.e., 9 paired bases in the acceptor stem and 4 paired bases in the TΨC stem; see Figs. 2C and D, respectively). A is shown at position 58 (in the tRNA numbering system) instead of the U observed by sequencing (see above). Interestingly, the extra base found in *D. discoideum* Sec tRNA compared to *T. thermophila* and animal Sec tRNAs occurs at position 13 as a bulged U base in the D-stem. Otherwise, the secondary structures of these tRNAs are similar to those of other eukaryotic Sec tRNAs [16].

The phylogeny of known Sec tRNA sequences was assessed by placing them in a phylogenetic tree (Fig. 3). *T. thermophila* Sec tRNA exists on a branch, albeit

A *Tetrahymena thermophila*

```
Gene sequence: GCTTTGGTGA GTTGGACTGG TGTCCAGGAC GGGCTTCAAA CCCGTGGGTG TCGAAAGGCA CAGGAGTTCG ATTCTCCCTC CGGAGCG
tRNA sequence: GCUUUGGUGA GUUGGACUGG UGUCCAGGAC GGGCUUCAAA CCCGUGGGUG UCGAAAGGCA CAGGAGUUCG UUUCUCCCUC CGGAGCGCCA
```

B *Dictyostelium discoideum*

```
Gene sequence: GCGTGTTTGA TCGCTATCTG GGGATGCGTG TGGACTTCAA ATCCATAGCT CGATAACTCG AGAGTTGTTC GATTCAACCT GACGCGCG
tRNA sequence: GCGUGUUUGA UCGCUAUCUG GGGAUGCGUG UGGACUUCAA AUCCAUAGCU CGAUAACUCG AGAGUUGUUC GUUUCAACCU GACGCGCGCCA
```

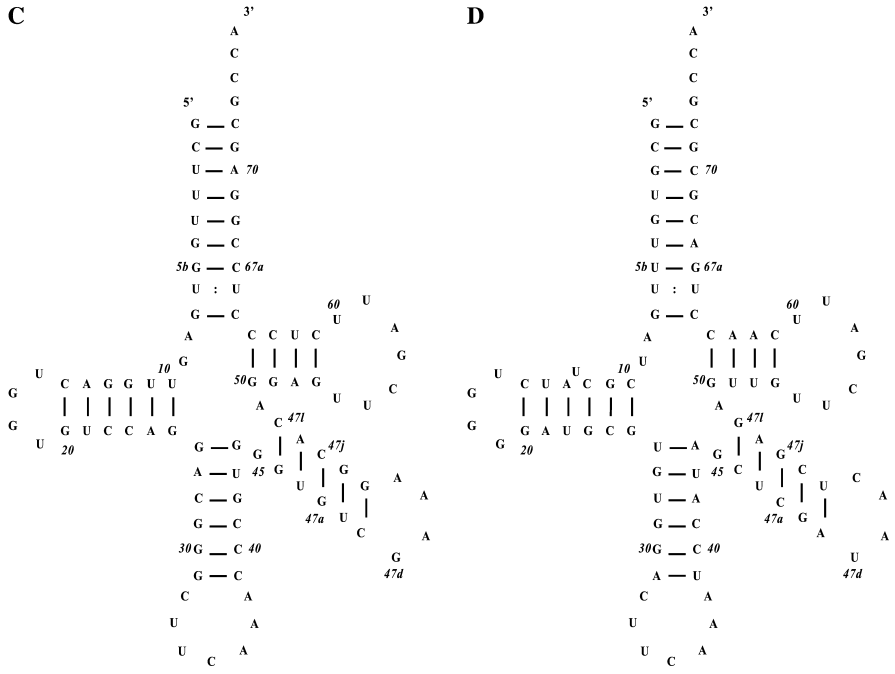

Fig. 2. Sequences of *T. thermophila* and *D. discoideum* Sec tRNA genes and Sec tRNAs. (A) The computational sequence of the *T. thermophila* Sec tRNA gene is shown, and in (B) the *D. discoideum* Sec tRNA gene sequence is shown. The respective sequences derived by direct sequencing of each Sec tRNA are shown under the gene sequence. The anticodon is shown in bold. The base at positions 81 in *T. thermophila* and 82 in *D. discoideum* is most likely 1-methyladenine in both Sec tRNAs as this modified base produces an artifact in the sequencing technique yielding T instead of the expected A (see text and [5]). (C) *T. thermophila* and (D) *D. discoideum*, Sec tRNAs are shown in cloverleaf model as a 9/4 representation. See text for details.

distantly, with those of other eukaryotes. *D. discoideum*, however, appears to be the most divergent eukaryotic Sec tRNA sequenced to date.

Discussion

We used a combination of a computational approach to identify potential Sec tRNA genes and a RT-PCR sequencing technique to determine the primary structures of the gene products as a novel approach to demonstrate that two widely studied organisms, *T. thermophila* and *D. discoideum*, contain Sec tRNA for insertion of Sec into protein. These organisms, therefore, utilize the trace element selenium for insertion into protein through the highly sophisticated Sec incorporation apparatus.

The computational tRNA genomic analysis program is proving to be a powerful tool to identify Sec tRNA genes in organisms with fully or partially sequenced genomes. As modeled in the present study, Sec tRNA genes can be readily identified by computational analysis and then their products confirmed within the tRNA population by the direct sequencing technique. Thus, this approach will provide a simple means of determining how widespread the use of UGA as a Sec codon is in nature. The distribution of Sec insertion into protein in nature, however, will likely be assessed only on an organism by organism basis.

A [^{75}Se]Sec-tRNA that decoded UGA was previously observed in *T. pseudonana* [7]. This was a surprising finding in light of the fact that *Tetrahymena* was known to use UAA and UAG as Gln codons [17,18], and that this organism must then share the function of its sole termination codon with Sec. The sequence of *T. thermophila* Sec tRNA reported in this study unequivocally confirmed the existence of Sec tRNA in this phylum and the use of UGA as a codon for Sec.

Sec-containing tRNAs that decode UGA have also been identified in a higher plant, *Beta vulgaris*, and in a filamentous fungus, *Gliocladium virens* [8]. In addition, *Saccharomyces cerevisiae* contain a tRNA that is acylated with Sec, but its coding properties have not been identified [7]. The possibility that the yeast Sec-containing tRNA might be Cys-tRNA in which selenium has replaced sulfur seems unlikely since the cells were grown in

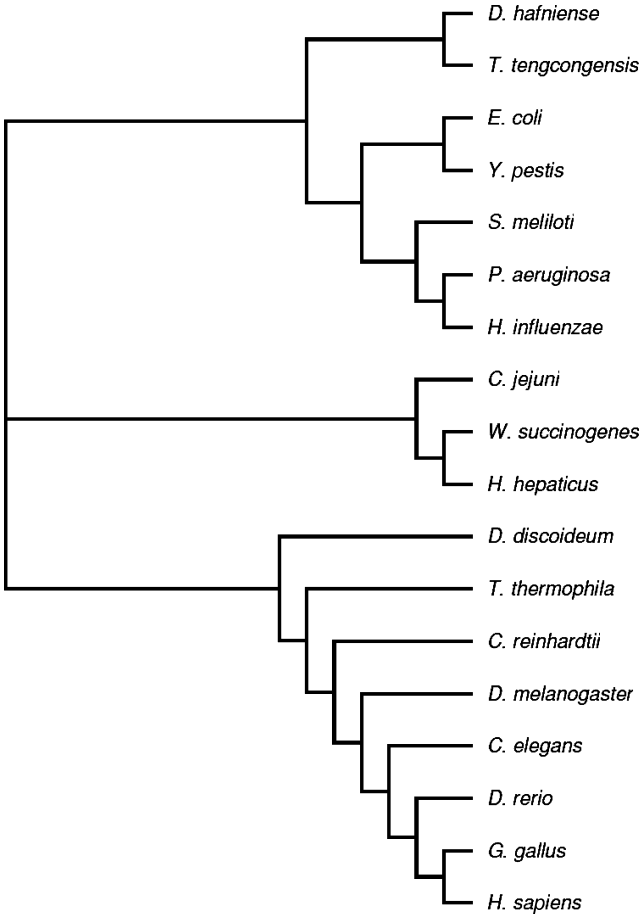

Fig. 3. Phylogenetic tree of eukaryotic Sec tRNAs. Accession numbers for Sec tRNA sequences were as follows: *C. reinhardtii* (AY268554), *Drosophila melanogaster* (M34509), *Caenorhabditis elegans* (M34508), *Danio rerio* (AF135237), *Gallus gallus* (K01941), *Homo sapiens* (K02923), *Wolinella succinogenes* (BX571659), *Helicobacter hepaticus* (AE017145), *E. coli* (Y00299), *Yersinia pestis* (NC_003143), *Sinorhizonium meliloti* (AE007196), *Pseudomonas aeruginosa* (AE004893), *Haemophilus influenzae* (U32753), *Desulfitobacterium hafniense* (NZ_AAAW00000000), *Thermoanaerobacter tengcongensis* (AE013138), and *Campylobacter jejuni* (AF486635).

the presence of high concentrations of cysteine and methionine to reduce any background that might arise by replacement of sulfur with selenium. The means of now being able to isolate partially purified samples of these tRNAs with [^{75}Se]Sec attached to them, deacylate, and then determine their primary sequences should permit us to assess where and how these novel tRNAs may fit into the overall evolutionary scheme of selenium metabolism.

Acknowledgment

This work was supported in part by NIH Grant GM061603 to V.N.G.

References

[1] D.L. Hatfield, V.N. Gladyshev, Mol. Cell. Biol. 22 (2002) 3565–3576.
[2] G.V. Kryukov, S. Castellano, S.V. Novoselov, A.V. Lobanov, O. Zehtab, R. Guigo, V.N. Gladyshev, Science 300 (2003) 1439–1443.
[3] S.V. Novoselov, M. Rao, N.V. Onoshko, H. Zhi, G.V. Kryukov, Y. Xiang, D.P. Weeks, D.L. Hatfield, V.N. Gladyshev, EMBO J. 21 (2002) 3681–3693.
[4] G.V. Kryukov, V.N. Gladyshev, EMBO Rep. 5 (2004) 538–543.
[5] M. Rao, B.A. Carlson, S.V. Novoselov, D.P. Weeks, V.N. Gladyshev, D.L. Hatfield, RNA 9 (2003) 923–930.
[6] M. Sussman, Methods Cell Biol. 28 (1987) 9–29.
[7] D.L. Hatfield, B.J. Lee, N.M. Price, T.C. Stadtman, Mol. Microbiol. 5 (1991) 1183–1186.
[8] D.L. Hatfield, I.S. Choi, S. Mischke, L.D. Owens, Biochem. Biophys. Res. Commn. 184 (1992) 254–259.
[9] A.D. Kelmers, D.E. Heatherly, Anal. Biochem. 44 (1971) 486–495.
[10] D. Hatfield, C.R. Matthews, M. Rice, Biochim. Biophys. Acta 564 (1979) 414–423.
[11] A. Diamond, B. Dudock, D. Hatfield, Cell 25 (1981) 497–506.
[12] A.M. Diamond, I.S. Choi, P.F. Crain, T. Hashizume, S.C. Pomerantz, R. Cruz, C.J. Steer, K.E. Hill, R.F. Burk, J.A. McCloskey, J. Biol. Chem. 268 (1993) 14215–14223.
[13] J.R. Sampson, M.E. Saks, Methods Enzymol. 267 (1996) 384–410.
[14] J.B. Edwards, J. Delort, J. Mallet, Nucleic Acids Res. 19 (1991) 5227–5232.
[15] D.P. Morse, B.L. Bass, Biochemistry 36 (1997) 8429–8434.
[16] N. Hubert, C. Sturchler, E. Westhof, P. Carbon, A. Krol, RNA 4 (1998) 1029–1033.
[17] S. Horowitz, M.A. Gorovsky, Proc. Natl. Acad. Sci. USA 82 (1985) 2452–2555.
[18] Y. Kuchino, N. Hanyu, F. Tashiro, S. Nishimura, Proc. Natl. Acad. Sci. USA 82 (1985) 4758–4762.