ISSN 0026-8933, Molecular Biology, 2016, Vol. 50, No. 1, pp. 37–42. © Pleiades Publishing, Inc., 2016.
Original Russian Text © E.G. Varlamova, S.V. Novoselov, 2016, published in Molekulyarnaya Biologiya, 2016, Vol. 50, No. 1, pp. 44–50.

REVIEWS

UDC 577.321

Methods to Biosynthesize Mammalian Selenocysteine-Containing Proteins in vitro

E. G. Varlamova and S. V. Novoselov

Institute of Cell Biophysics, Russian Academy of Sciences, Pushchino, Moscow oblast, 142290 Russia;
e-mail: 1928lv@mail.ru

Received May 12, 2015; in final form, June 3, 2015

Abstract—The main problem in studying mammalian selenocysteine-containing proteins is that the proteins are difficult to obtain in a recombinant form because the amino acid selenocysteine (Sec), which is their component, is encoded by TGA, which is one of the stop codons. When only the open reading frame of a target protein is cloned in a plasmid, translation is prematurely terminated at the TGA codon. An intricate natural mechanism allows the codon to be recognized as a selenocysteine codon and involves various *cis*- and *trans*-acting factors, such as the selenocysteine insertion sequence (SECIS), mRNA secondary structure, selenocysteine tRNA Sec-tRNA$^{[Ser]Sec}$, SECIS-binding protein 2 (SBP2), selenocysteine-specific elongation factor EFsec, and others. Generation of recombinant selenoproteins in preparative amounts directly depends on the expression levels of the *cis*- and *trans*-acting transcription and translation factors to further complicate the problem, and cysteine homologs of selenoproteins are consequently used in many studies. Several methods designed to express mammalian selenoproteins in vitro are considered in the review.

Keywords: selenocysteine, mammalian selenocysteine-containing proteins, selenoprotein synthesis
DOI: 10.1134/S0026893316010210

GENERATION OF RECOMBINANT SELENOCYSTEINE-CONTAINING PROTEINS

Biosynthesis of Mammalian Selenocysteine-Containing Proteins in a Bacterial System

Arnér et al. [1] were the first to develop a system to express mammalian selenoprotein genes in *Escherichia coli*, using thioredoxin reductase (TrxR) as an example. In mammalian TrxR, selenocysteine (Sec) occurs in the penultimate position and is involved in the redox motif Gly-Cys-Sec-Gly [2]. Rather than being random, the choice of the enzyme was determined by the fact that the mechanism of Sec incorporation in selenocysteine-containing proteins somewhat differs between mammals and bacteria. One of the essential differences pertains to the spatial position of the Sec insertion sequence (SECIS) relative to the Sec UGA codon in mRNAs for selenocystein-containing proteins; i.e., this *cis*-acting element occurs in the mRNA 3'-untranslated region in mammals and within the open reading frame (ORF) in the vicinity of the UGA codon in bacteria. Because Sec is one amino acid away of the stop codon in mammalian TrxR, it was assumed that cloning the TrxR ORF in fusion with the SECIS element from the *M. melanoformicum* formate dehydrogenase gene (*fdh*F) would ensure a spatial proximity of the UGA codon and SECIS and that biosynthesis of recombinant mammalian TrxR would consequently follow the mechanism characteristic of bacteria. Overexpression of the SelA, SelB, and SelC genes and sodium selenite added to the culture medium further increase the production of recombinant TrxR.

Su et al. [3] developed another method to express mammalian selenoprotein genes in a bacterial system. Like in the above case, the gist is bringing the UGA codon and bacterial SECIS in spatial proximity in a genetic construct to ensure Sec incorporation in the protein product via the bacterial mechanism. A substantial difference of the method is that the UGA selenocysteine codon is several tens of base pairs away from the stop codon, rather than in its vicinity at the end of ORF as in mammalian TrxR. Mouse glutathione peroxidase 4 (Gpx4) was chosen as a protein to be expressed; its SECIS element corresponds to that of bacterial formate dehydrogenase (*fdh*F), like in the above case. To bring the two *cis*-acting elements close together, a 36-bp sequence complementary to the region immediately downstream of the selenocysteine UGA codon in the Gpx4 ORF was cloned upstream of the SECIS element in the genetic construct (Fig. 1). However, Sec incorporation in the recombinant pro-

Abbreviations: EFsec, elongation factor of selenocysteine; Sec, selenocysteine; SECIS, selenocysteine insertion sequence; Sec-tRNA$^{[Ser]Sec}$, selenocysteine tRNA; SBP2, SECIS-binding protein 2; SelA, prokaryotic selenocysteine synthase; SelB, selenocysteine-specific elongation factor; SelC, selenocysteine-specific tRNA for thioredoxin reductase (TrxR). ORF, open reading frame.

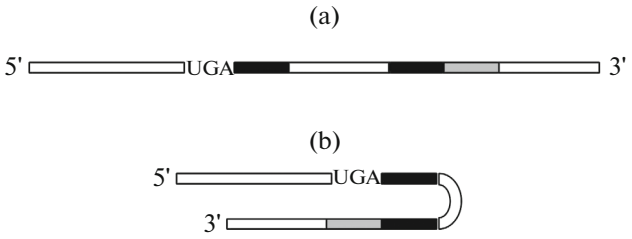

Fig. 1. Part of the genetic construct designed to produce recombinant selenocysteine-containing mGpx4 in a bacterial system. (a) Linear sequence of the *mgpx4* ORF with an internal UGA codon. Black bars, complementary sequences; gray bar, SECIS element from the bacterial formate dehydrogenase gene. (b) Loop formation brings two *cis*-acting elements, the selenocysteine UGA codon and SECIS, in spatial proximity of one another.

Fig. 2. Vector designed to efficiently express selenoprotein genes in mammalian cells (cited from [5]). The vector contains the *T. gondii selT* SECIS element wherein GGGA is replaced with ATGA. The SECIS element is immediately downstream of the cloning site (MCS) and the CMV promoter. In addition, the vector includes a human elongation factor 1α gene sequence ($P_{EF-1\alpha}$) followed by a sequence coding for the C-terminal region of rat SBP2 (SBP2).

tein expressed from the resulting construct was low efficient; i.e., approximately 3% of protein molecules had Sec.

Biosynthesis of Mammalian Selenocysteine-Containing Proteins in an Eukaryotic System

Heterologous expression of selenoproteins in mammalian cultured cells is also low efficient, as was demonstrated with a genetic construct containing the ORF of *selp*, which codes a human selenoprotein (SelP) [4]. The *selp* ORF codes for 10–12 Sec residues (depending on the species); in addition, SelP is Cys and His reach. For instance, the rat and human SelP proteins have two His-rich stretches each. One harbors eight (rat) or nine (human) out of 14 His residues occurring in the region; the other, seven and four residues, respectively. The rat SelP gene was expressed in HEK 293 human epithelial kidney cells via transient transfection; the product yield increased, but was still far from sufficient. Only trace amounts of recombinant SelP were obtained with other cell lines, such as CHO and HepG2. The *selp* expression level strongly depended on the selenium concentration in the medium. The presence of 10, 30, or 100 nM selenite in the culture medium increased the product yield by a factor of 1.2, 3.2, or 7.2, respectively. The recombinant SelP was detectable only by immune methods in any case.

Novoselov et al. [5] designed a more efficient system to express selenoprotein genes in mammalian cells with the use of the *Toxoplasma gondii* SECIS element. The SECIS sequences of *T. gondii* selenoprotein genes were identified by a similarity analysis with the SECIS elements of known selenocysteine-containing proteins. The *T. gondii* SECIS showed homology to four SECIS elements from genes for the mammalian selenoproteins SelT, SelS, SelK, and SelW. Two genetic constructs were obtained to contain *mselh*, which coded for mouse SelH fused with GFP. The method is different in that the *mselh* SECIS was replaced with the SECIS from the *T. gondii* SelT gene in one construct and the SECIS from the *T. gondii* SelS gene in the other. The genetic constructs were expressed in mammalian cells of two lines, HEK 293 and NIH 3T3. The results showed, first, that the amount of the full-size mSelH produced with either *T. gondii* SECIS was comparable with that obtained with the *mselh* SECIS. Second, the full-size mSelH protein content in HEK 293 cells was one order of magnitude higher than in NIH 3T3 cells, where premature translation termination at the selenocysteine UGA codon occurred more often upon expression of the two genetic constructs.

It was found additionally that the *T. gondii selT* SECIS element has GGGA in place of the conserved sequence AUGA in its SBP2-binding site. To obtain a genetic construct allowing high-level expression of mammalian selenoprotein genes in a eukaryotic system, experiments were performed to check which of the two SBP2-binding site variants is more efficient for expression.

A substitution of conserved AUGA for GGGA in the *selT* SECIS was found to provide for more efficient SECIS binding with SBP2 and, consequently, more efficient Sec incorporation in selenocysteine-containing proteins. A genetic construct optimal for expressing selenocysteine-containing protein genes in mammalian cells was thus obtained (Fig. 2) [5].

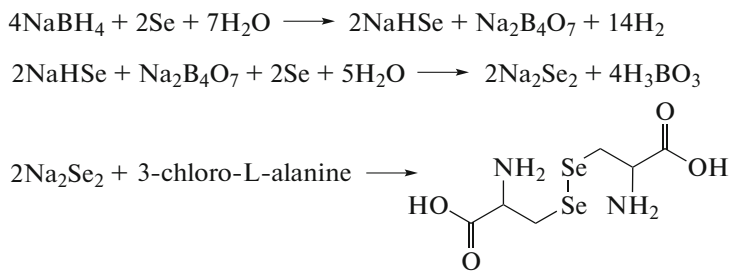

Fig. 3. Synthesis of the amino acid selenocystine from elemental selenium.

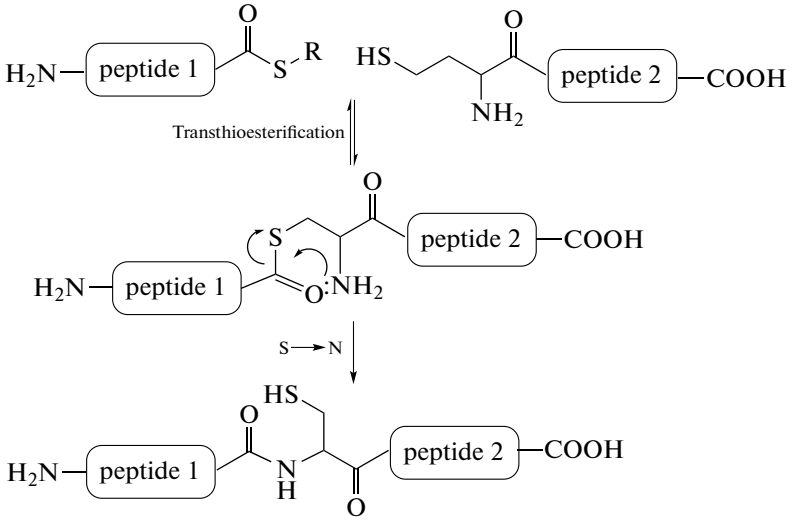

Fig. 4. Mechanism of chemical native peptide ligation. The thiol group of the N-terminal cysteine of peptide 2 attacks the C-terminal thioester of peptide 1 in a buffer with pH 7. The reaction is reversible and yields a thioester intermediate. Then an intramolecular substitution of nitrogen for sulfur generates a native peptide bond in the ligation site.

CHEMICAL METHODS TO SYNTHESIZE SELENOCYSTEINE-CONTAINING PROTEINS

Wu and Hilvert [6] pioneered chemical synthesis of selenocysteine-containing proteins. Their approach was generating the synthetic selenoprotein selenosubtilisin by chemical conversion of Ser to Sec in the sublilisin active site. The enzyme is treated first with phenylmethanesulfonyl fluoride to produce a sulfonylated Ser residue, which is then converted to selenocysteine via treatment with hydrogen selenide. The resulting selenosubtilizin acquired new enzymatic properties due to Sec, displaying glutathione peroxidase and acyltransferase activities [6, 7].

The amino acid selenocysteine is possible to synthesize from sodium diselenide, which is obtained in situ via selenium reduction with sodium borohydride ($NaBH_4$) (Fig. 3). β-Chloroalanine added to the reaction mixture produces Sec in the form of yellow crystals. Then Sec is treated with *p*-methoxybenzyl chloride to obtain a pMob derivative [8]. The α-nitrogen atom can then be protected with a fluorenylmethoxycarbonyl (Fmoc) group by treating Se-*p*-methoxybenzyl selenocysteine with Fmoc-OSu, thus producing N-9-fluorenylmethoxycarbonyl-Se-4-methoxybenzyl selenocysteine (Fmoc-Se(pMob)-OH). The derivative can be recrystallized; the overal product yield is approximately 60% [9].

Early studies to synthesize Sec-containing peptides were performed by Walter and colleagues [10–12], who obtained a Se-benzyl derivative of selenocysteine and synthesized selenium-containing analogs of glutathione, oxytocin, and somatostatin. The benzyl group is then usually removed by sodium in liquid ammonia. However, the method has several substantial drawbacks. Fmoc-Se(pMob)-OH is prone to racemization in solid-phase peptide synthesis; dehydroalanine can form at the elongation step of protein synthesis with the use of a tertiary base, such as diisopropylethylamine; etc.

Another chemical procedure generating synthetic selenoproteins is based on two methods, chemical native peptide ligation, which was developed by Kent and colleagues [13], and a semisynthetic method utilizing inteins [14, 15]. The methods take advantage of

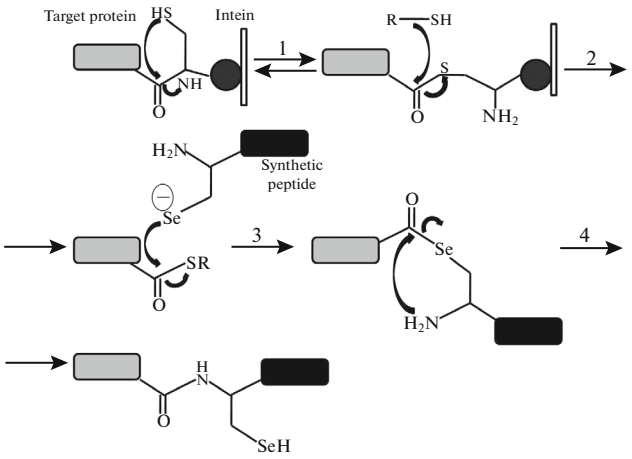

Fig. 5. Mechanism of intein-mediated peptide ligation. An intein immobilized in a chitin column interacts with a target protein to produce a bond, which exists in equilibrium between the peptide and thioester forms (reaction 1). When a small thiol molecule is added into the reaction mixture to a high concentration, a thioester-labeled protein forms (reaction 2). The addition of a synthetic peptide containing N-terminal Sec is accompanied by an attack on the thioester in the target protein, producing a selenoester (reaction 3). The selenoester is then rapidly converted to a peptide form.

the fact that selenium is similar in physicochemical properties to sulfur; in particular, selenium attacks the α-thioester bond as readily. The ligation method consists in chemical synthesis of medium-sized proteins by chemical coupling of two or more peptide fragments. One of the peptides has an α-thioester at the C end, and the other has a cysteine residue at the N end, so that a tight peptide bond forms between the thioester and cysteine. The gist of the method is shown in Fig. 4. The method was modified by two research teams; i.e., Sec was used in place of Cys to synthesize selenocysteine-containing proteins [16, 17].

Hondal and colleagues [17, 18] used intein-mediated peptide ligation to obtain semisynthetic selenoproteins, an intein being added to the C end of a protein of interest. The peptide bond at the junction site exists in equilibrium between the peptide and thioester forms. A synthetic peptide containing Sec or Cys at the N end can be used to attack the thioester and ligates the two peptides together. The method makes it possible to obtain large amounts (several milligrams per milliliter of a bacterial culture) of a thioester-bound recombinant protein. A scheme of intein-mediated peptide ligation is shown in Fig. 5.

Several examples of using the semisynthetic approach to obtain selenocysteine-containing proteins are available in the literature. A synthetic form of the protein azurin was produced by chemical native peptide ligation by replacing Cys112 with Sec [19, 20]. The copper atom binds with Cys112 and two histidine residues in the active site, allowing the copper atom to change its valence (cycling between Cu^{1+} and Cu^{2+}) during biological electron transport [21]. A substitution of any other of the 19 amino acid residues for Cys substantially alters the properties of the enzyme. Although Cys and Sec are similar in chemical properties, the semisynthetic protein variant has unique spectroscopic properties [19, 20]. A similar method was used to obtain the bovine pancreatic trypsin inhibitor by replacing Cys38 with Sec [22].

Another example of using the semisynthetic approach is production of RNase A [9, 23], which harbors eight cysteine residues, which are oxidized to form four disulfide bonds. Theoretically, each of the eight Cys residues can be replaced with Sec. Three recombinant RNase A forms were obtained: one had Sec in place of Cys110, another had Sec in place of Cys95, and a third one had both Cys110 and Cys95 replaced with Sec. Cys110 is involved in disulfide bonding with Cys58, and Cys95 binds with Cys40. The two disulfide bonds were found to play a crucial role in enzyme activity. Protein refolding is distorted in their absence, which was used as a test to demonstrate the formation of Se–S bonds.

Hondal and colleagues [24] used the semisynthetic method to obtain thyoredoxin reductase TrxR. As mentioned above, Sec occurs in the penultimate position and belongs to the Gly-Cys-Sec-Gly redox motif in mammalian TrxR, preventing the use of chemical native peptide ligation in this case. A total of 487 TrxR amino acid residues were obtained as an intein-tagged polypeptide in a bacterial system. This intein-involving technique is known as an IMPACT system, being developed by New England Biolabs. The IMPACT system is designed to produce an intein-tagged protein in *Escherichia coli* cells and includes affinity purification of the protein on chitin agarose via the use of a chitin-binding domain attached to the intein. The purified protein is combined with the thiol N-methyl mercaptoacetamide, which either attacks the thioester in the TrxR–intein coplex to produce a thioester-tagged thioredoxin reductase or keeps the tripeptide in the reduced form. The tripeptide can then attack the TrxR–intein–thioester complex or the N-methyl mercaptoacetamide–TrxR–thioester complex. Once the thioester forms, the free amino group at the N end of the peptide can rapidly attack the carbonyl group to produce a stable peptide bond, and the enzyme does not undergo further modification after its formation. The semisynthetic TrxR produced by this method has high catalytic activity.

In addition to TrxR, the same research team produced semisynthetic methionine R-sulfoxide reductase (MsrB1), which is also known as the selenocysteine-containing protein R (SelR) [25]. There are two other proteins, MsrB2 and MsrB3, wherein Sec is replaced with Cys; the replacement shows that Sec is not chemically essential to catalyze the reduction of methionine sulfoxide to methionine. In spite of the

relatively small size of the protein (116 amino acid residues) and the favorable position of Sec (95), the protein was synthesized in the form of separate modules, which were then joined via chemical ligation. Region 1–94 was fused with an intein and produced in *E. coli* cells, allowing for the formation of a reactive thioester group (F94). The other region, 95–116, was then produced as a synthetic peptide with a Sec residue at the N end. A selenoester and a peptide bond were then formed. Chemical ligation of the two protein modules was verified by SDS-PAGE.

CONCLUSIONS

It is beyond doubt now that the microelement selenium and selenium-containing proteins are essential for the normal function of organisms. The unique character of selenocysteine and the complex mechanism of its incorporation in nascent polypeptides attract great interest, but many issues are still unclear in the field. The functions of selenoproteins are difficult to study primarily because the proteins are difficult to obtain in preparative amounts, and this problem arises because selenocysteine is encoded by a translation termination codon and because the selenoprotein biosynthesis mechanism is unique. Several recombinant selenocysteine-containing proteins have been obtained to date, including mammalian thioredoxin reductase, SelR, SelP, glutathione peroxidase 4, etc.

There are two main approaches to production of selenocysteine-containing proteins, gene engineering with expression of selenoprotein genes and chemical methods. Main examples of their employment in producing selenoproteins are considered in the review.

Although the methods to produce semisynthetic and gene-engineering selenoproteins seem relatively simple at first glance, certain problems are encountered when synthesizing selenoproteins. For instance, the use of genetic constructs to obtain selenoproteins still fails to ensure their high-level expression and synthesis in preparative amounts. Optimal conditions are to be selected and the techniques optimized for each particular protein; i.e., the approach is not universal enough to be broadly used in the field. Chemical methods to produce selenoproteins are similarly neither ideal nor productive, each having substantial drawbacks. In particular, certain components utilized in the production are not involved in the natural mechanism of selenoprotein biosynthesis, as is the case with inteins, thiols, etc., potentially allowing the structural and functional interactions that would be impossible with routine recombinant methods.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research ("My First Grant;" project nos. 14-04-31460 and 13-04-00576).

REFERENCES

1. Arnér E.S., Sarioglu H., Lottspeich F., Holmgren A., Böck A. 1999. High-level expression in Escherichia coli of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the selA, selB and selC genes. *J. Mol. Biol.* **292**, 1003–1016.
2. Gladyshev V.N., Jeang K.T., Stadtman T.C. 1996. Selenocysteine, identified as the penultimate C-terminal residue in human T-cell thioredoxin reductase, corresponds to TGA in the human placental gene. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6146–6151.
3. Su D., Li Y., Gladyshev V.N. 2005. Selenocysteine insertion directed by the 3'-UTR SECIS element in *Escherichia coli*. *Nucl. Acids Res.* **33**, 2486–2492.
4. Tujebajeva R.M., Harney J.W., Berry M.J. 2000. Selenoprotein P expression, purification, and immunochemical characterization. *J. Biol. Chem.* **275**, 6288–6294.
5. Novoselov S.V., Lobanov A.V., Hua D., Kasaikina M.V., Hatfield D.L., Gladyshev V.N. 2007. A highly efficient form of the selenocysteine insertion sequence element in protozoan parasites and its use in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7857–7862.
6. Wu Z.-P., Hilvert D. 1989. Conversion of a protease into an acyl transferase: selenosubtilisin. *J. Am. Chem. Soc.* **111**, 4513–4514.
7. Wu Z.-P., Hilvert D. 1990. Selenosubtilisin as a glutathione peroxidase mimic. *J. Am. Chem. Soc.* **112**, 5647–5648.
8. Koide T., Itoh H., Otaka A., Yasui H., Masataka K., Esaki N., Soda K., Fuji N. 1993. Synthetic study on selenocystine-containing peptides. *Chem. Pharm. Bull.* **41**, 502–506.
9. Hondal R.J., Raines R.T. 2002. Semisynthesis of proteins containing selenocysteine. *Meth. Enzymol.* **347**, 70–83.
10. Theodoroulos D., Shuwartz I.L., Walter R. 1967. Synthesis of selenium-containing peptids. *Biochemistry.* **6,** 3927–3932.
11. Hartrodt B., Neuber K., Bierwolf B., Blech W., Jakubke H.D. 1980. Synthese von [3,14-L-selenocystein, 8-D-tryptophan]somatostatin. *Tetrahedron Lett.* **21**, 2393–2398.
12. Walter R., Chan W.Y. 1967. Syntheses and pharmacological properties of selenium isologs of oxytocin and deamino-oxytocin. *J. Am. Chem. Soc.* **89**, 3892–3893.
13. Dawson P.E., Muir T.W., Clark-Lewis I., Kent S.B. 1994. Synthesis of proteins by native chemical ligation. *Science.* **266**, 776–9.
14. Hackeng T.M., Griffin J.H., Dawson P.E. 1999. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10068–10073.
15. Evans T.C., Jr., Benner J., Xu M.Q. 1998. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.* **7**, 2256–2264.
16. Quaderer R., Sewing A., Hilvert D. 2001 Selenocysteine-mediated native chemical ligation. *Helv. Chim. Acta.* **84**, 1191–1206.

17. Giesselman M.D., Xie L., van der Donk W.A. 2001. Synthesis of a selenocysteine-containing peptide by native chemical ligation. *Org. Lett.* **3**, 1331–1334.

18. Gray H.B., Winkler J.R. 1996. Electron transfer in proteins. *Annu. Rev. Biochem.* **65**, 537–561.

19. Shao Y., Kent S.B. 1997. Protein splicing: occurrence, mechanisms and related phenomena. *Chem. Biol.* **4,** 187–194.

20. Berry S.M., Giesselman M.D., Nilges M.J., van der Donk W.A., Lu Y. 2002. An engineered azurin variant containing a selenocysteine copper ligand. *J. Am. Chem. Soc.* **124**, 2084–2085.

21. Ralle M., Berry S.M., Nilges M.J., Giesselman M.D., van der Donk W.A., Lu Y., Blackburn N.J. 2004. The selenocysteine-substituted blue copper center: spectroscopic investigations of Cys112SeCys *Pseudomonas aeruginosa* azurin. *J. Am. Chem. Soc.* **126**, 7244–7256.

22. Roelfes G., Hilvert D. 2003. Incorporation of selenomethionine into proteins through selenohomocysteine-mediated ligation. *Angew Chem. Int. Ed. Engl.* **42**, 2275–2277.

23. Hondal R.J., Nilsson B.L., Raines R.T. 2001 Selenocysteine in native chemical ligation and expressed protein ligation. *J. Am. Chem. Soc.* **123**, 5140–5141.

24. Eckenroth B.E., Harris K., Turanov A.A., Gladyshev V.N., Raines R.T., Hondal R.J. 2006. Semisynthesis and characterization of mammalian thioredoxin reductase. *Biochemistry.* **45**, 5158–5170.

25. Kim H.Y., Gladyshev V.N. 2005. Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol.* **3**, e375.

Translated by T. Tkacheva