

Chapter 9

Selenoproteins and Selenoproteomes

Vadim N. Gladyshev

Abstract Recent progress in high-throughput sequencing and analysis allowed efficient identification of selenocysteine-containing proteins in sequence databases, including full sets of selenoproteins in organisms, designated selenoproteomes. Information is currently available on selenoproteomes from all major model organisms as well as humans, which have 25 selenoprotein genes. This chapter gives an overview of selenoproteins at the level of individual proteins, protein families, and entire selenoproteomes. Comparative genomic analyses of selenoproteins offer exciting avenues for studying selenoprotein function and evolution, provide insights into the biological functions of the trace element selenium, and even allow addressing important biological questions unrelated to selenium.

9.1 Introduction

Selenium (Se) is found in biological molecules in three specific forms. It occurs in the form of selenouridine in the wobble position of certain bacterial tRNAs [1]. In addition, in some bacterial Se-containing molybdoproteins it is present as a labile cofactor that contains a Se–Mo bond that is directly involved in catalysis [2–4]. However, the major form of Se in biological systems is represented by selenocysteine (Sec), the 21st amino acid in the genetic code. It is encoded by the UGA codon and has been found in each domain of life (i.e., bacteria, archaea, and eukaryotes). It is now clear that the essential roles of Se in biology, as well as its beneficial functions in human health, are due to its presence in proteins in the form of Sec.

V.N. Gladyshev (✉)

Division of Genetics, Department of Medicine, Brigham and Women's Hospital,
Harvard Medical School, Boston, MA 02115, USA
e-mail: vgladyshev@rics.bwh.harvard.edu

In contrast to the 20 common amino acids in proteins, Sec is utilized only when it is required for protein function. Accordingly, it is normally a key functional (and almost always catalytic) group in proteins, wherein selenoproteins use Sec in redox catalysis. Therefore, information on identities and functions of selenoproteins is a key to the understanding of biological and biomedical roles of Se.

9.2 Bioinformatics Tools for Selenoprotein Identification

Over the years, researchers in the Se field developed very convenient tools for selenoprotein analyses. Selenoproteins can be analyzed by following the presence of Se in protein fractions, e.g., by inductively coupled plasma mass-spectrometry (ICP-MS) [5]. Sec-containing proteins can also be metabolically labeled with ^{75}Se , a convenient γ -emitter that remains covalently bound to proteins during the SDS-PAGE procedure and can be visualized on gels and membranes with a PhosphorImager [6]. With these techniques, a number of proteins have been identified in both prokaryotes and eukaryotes [6–8]. We suggest, however, that the approach that in recent years benefited the Se field the most with regard to selenoprotein discovery is the analysis of sequence databases.

Remarkable progress in genome sequencing and analyses offered an excellent resource for selenoprotein discovery and analysis of selenoprotein functions. All selenoprotein genes have two characteristic features: a Sec-encoding TGA codon and a Sec insertion sequence (SECIS) element. The TGA triplet that codes for Sec does not provide sufficient information at the nucleotide sequence level to identify Sec sites computationally. However, SECIS elements are amenable to these techniques as these structures are highly specific for selenoprotein genes, have conserved sequences, and possess a sufficiently complex secondary structure. Therefore, many bioinformatics analyses focused on SECIS elements, and selenoprotein discovery used the following strategy: (1) finding candidate SECIS elements; (2) analyzing upstream regions to identify coding regions; and (3) testing candidate selenoproteins for insertion of Se by metabolically labeling cells with ^{75}Se . The first selenoproteins identified using this technique were mammalian selenoproteins R (now known as methionine-*R*-sulfoxide reductase 1), N, and T [9, 10]. These searches were initially restricted to small nucleotide sequence databases, but later could be adapted to searching entire genomes [11–13]. Currently, to aid in these analyses, groups of closely related genomes are analyzed in order to identify evolutionarily conserved SECIS elements that belong to selenoprotein orthologs in these organisms [14].

A separate approach (independent of SECIS elements) was also developed that searched for in-frame TGA codons by analyzing TGA flanking sequences [13–16]. This approach is possible because the majority of selenoprotein genes have homologs (most often in organisms with reduced or lost Sec utilization), in which Sec is replaced with Cys. Thus, a strategy was developed wherein protein databases (large sets of overlapping reading frames, nonredundant protein databases, etc.) are searched

against nucleotide sequences from organisms that contain selenoprotein genes (genomes, expressed sequence tags, metagenomic projects) to identify nucleotide sequences that, when translated, align with Cys-containing protein sequences from the protein database, such that Cys residues align with candidate Sec and these pairs are flanked by conserved sequences. Although SECIS predictions could be used to guide the computational gene predictions, the Sec/Cys homology approach is completely independent of the searches for SECIS elements. Therefore, this method provides a SECIS-independent tool for selenoprotein identification. Both Sec/Cys and SECIS-based algorithms identify very similar sets of selenoprotein genes in organisms, suggesting that both tools show excellent performance and that all, or almost all, selenoproteins can be identified by these programs in completely sequenced genomes and large sequence databases.

Major currently known selenoprotein families are shown in Fig. 9.1. The majority of these proteins were discovered using bioinformatics approaches and subsequently verified, at least in the case of eukaryotic selenoproteins, experimentally. Below, the best studied selenoproteins are described, with the focus on eukaryotic selenoproteins. Additional and more detailed information on various selenoproteins can be found in various chapters throughout the book.

9.3 Mammalian Selenoproteins

9.3.1 *Glutathione Peroxidases*

Mammals have eight glutathione peroxidases (GPx1–GPx8), of which five are Sec-containing enzymes (GPx1, GPx2, GPx3, GPx4, and GPx6). However, GPx6 reverted back to a Cys-containing protein in many rodents, including mice and rats, so these organisms have only four selenoprotein GPxs [14]. GPx1 is the first animal selenoprotein identified [17] and it is also the most abundant one in mammals, especially in the liver and kidney. It catalyzes glutathione-dependent hydroperoxide reduction. In recent years, another GPx, GPx4, received much attention due to its essential status during embryonic development in mice and role in regulation of phospholipid hydroperoxide levels [18]. Moreover, the mitochondrial form of this protein serves a structural role in mature sperm and was implicated in disulfide bond formation during spermiogenesis [19]. Whereas GPx1 and GPx4 are expressed in all cells, GPx2 is gastrointestinal and GPx3 is primarily made in the kidney and secreted to the blood stream. It localizes to the basement membrane of the proximal tubules in the kidney [20]. However, it remains unclear how it can function in the extracellular milieu in the absence of sufficient levels of thiol reductants. Besides mammals, selenoprotein GPx homologs were identified in most animals as well as various single-celled eukaryotes and even bacteria. However, the ancestral form of these proteins is the Cys-containing form, and it is thought that Cys was replaced with Sec during evolution to make these enzymes better catalysts.

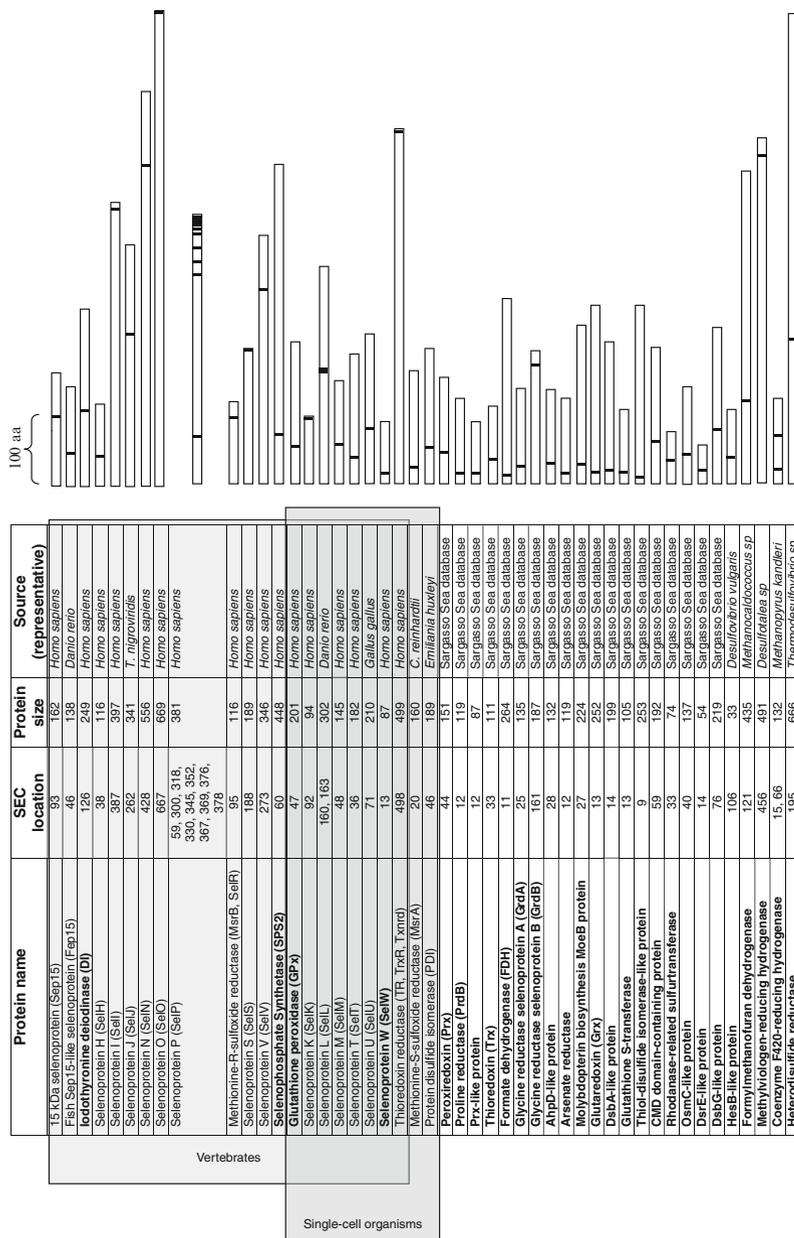


Fig. 9.1 Selenoprotein families. Selenoproteins that occur in vertebrates or single-celled eukaryotes are highlighted by *shaded boxes*, and selenoproteins found in prokaryotes are shown in *bold*. On the *right*, relative sizes of selenoproteins are shown (relative to a 100 amino acid scale) and the location of Sec within protein sequence is shown by a *black line*

9.3.2 *Thyroid Hormone Deiodinases*

There are three deiodinases (DI1, DI2 and DI3) in mammals, which activate and/or inactivate thyroid hormones by reductive deiodination. Deiodinases also occur in other vertebrates, and their homologs were even detected in unicellular eukaryotes and bacteria, although their function must be different in these organisms. Like GPxs and the majority of other selenoproteins, deiodinases are thioredoxin-fold proteins. These enzymes are extensively reviewed in this book.

9.3.3 *Thioredoxin Reductases*

The entire family of mammalian thioredoxin reductases (TRs) is dependent on Se as all three TRs in mammals are selenoproteins. Sec in TRs is located in the C-terminal penultimate position. These enzymes evolved from glutathione reductases by adding a C-terminal Sec-containing extension that became an intraprotein substrate for the classical N-terminal active center of pyridine nucleotide disulfide oxidoreductase family members [21–24]. TR1 (also known as TrxR1, TxnRd1) is a cytosolic and nuclear protein. Its main function is to control the reduced state of thioredoxin. However, it exhibits broad substrate specificity, especially with regard to low molecular weight compounds [25] and occurs in the form of at least six isoforms generated by alternative transcription initiation and alternative splicing [26–28]. A close homolog of TR1 is thioredoxin/glutathione reductase (TGR, also known as TR2, TxnRd3 and TrxR3) that, compared to other animal TRs, has an additional N-terminal glutaredoxin domain [22]. This protein was implicated in the formation/isomerization of disulfide bonds during sperm maturation [29]. TGR can catalyze many reactions specific for thioredoxin and glutathione systems. TR3 (also known as TxnRd2 and TrxR2) is a mitochondrial protein, which keeps mitochondrial thioredoxin and glutaredoxin 2 in the reduced state. TR1 and TR3 are essential for embryonic development in mammals [30, 31], while the consequences of TGR knockout have not been examined thus far.

9.3.4 *Methionine-R-Sulfoxide Reductase 1 (MsrB1)*

MsrB1 is the first selenoprotein identified through bioinformatics approaches. It was designated as Selenoprotein R [9] and Selenoprotein X [10], but after it was found to catalyze stereospecific reduction of methionine-*R*-sulfoxide residues in proteins, it was renamed as MsrB1 [32]. Mammals have two additional MsrBs (MsrB2 and MsrB3), which contain catalytic Cys in place of Sec and reside in mitochondria and the endoplasmic reticulum, respectively [33]. At least in liver and kidney of mammals, MsrB1 has the highest activity of all MsrBs, so the protein reductive repair

function is dependent on Se in mammals. MsrB1 is located in the cytosol and nucleus [33]. MsrB1 knockout mice are viable, but are characterized by oxidative stress [34].

9.3.5 15 kDa Selenoprotein (Sep15)

Sep15 is a conserved eukaryotic selenoprotein that occurs in most animals as well as in some unicellular eukaryotes, such as algae [7]. It resides in the endoplasmic reticulum where it binds UDP-glucose:glycoprotein glucosyltransferase, a sensor of protein folding [35]. Sep15 is composed of a N-terminal ER signal peptide, a Cys-rich domain responsible for binding UDP-glucose:glycoprotein glucosyltransferase, and a C-terminal domain characterized by the thioredoxin-like fold. Sep15 is implicated in the cancer prevention effect of dietary Se [36, 37]. Sep15 knockout mice are viable, but develop cataracts (MV Kasaikina and VN Gladyshev, unpublished).

9.3.6 Selenophosphate Synthetase 2 (SPS2)

By analogy to bacterial selenophosphate synthetase SelD [38], SPS2 was thought to synthesize selenophosphate, a Se donor compound. It was recently shown to be essential for selenoprotein biosynthesis in mammals, whereas the function of SPS1, a paralog of SPS2, remains unknown [39].

9.3.7 Selenoprotein P (SelP)

SelP is the only selenoprotein with multiple Sec residues [40]. For example, human and mouse SelP have ten Sec residues and zebrafish SelPa has 17 [41]. However, the number of Sec residues in SelP homologs varies greatly (e.g., 7–15 in mammals) [41]. SelP is the major plasma selenoprotein, which is synthesized primarily in the liver and delivers Se to certain other organs and tissues [42, 43]. The SelP knockout mouse was particularly useful in examining Se metabolism in mammals as discussed elsewhere in this book.

9.3.8 Selenoproteins W (SelW) and V (SelV)

SelW is the smallest mammalian selenoprotein [44]. Although it was one of the first identified (more than 20 years ago), its function remains unknown. SelW homologs were identified in lower eukaryotes and even bacteria, but these findings did not help identify SelW function [16]. A SelW paralog, SelV, is a larger protein due to an additional N-terminal sequence of unknown function [14]. This protein is expressed exclusively in testes. Its function is also not known.

9.3.9 *Selenoproteins T (SelT), M (SelM), and H (SelH)*

Functions of these three proteins are not known. They are clustered here because they belong to a group of thioredoxin-like fold proteins (together with Sep15, SelW and SelV). SelT is among the first selenoproteins identified through bioinformatics [9]. SelM is a distant homolog of Sep15 and, like Sep15, it resides in the endoplasmic reticulum [37, 45]. SelH was first identified as BthD in fruit flies [12, 14]. It resides in the nucleus. Several studies have found that knockdown of these proteins leads to oxidative stress suggesting roles, at least partially, as antioxidants.

9.3.10 *Selenoproteins O (SelO) and I (SelI)*

SelO is a widely distributed protein with homologs in animals, bacteria, yeast and plants, but the functions of any members of this protein family are not known [14]. Only vertebrate homologs of SelO have Sec, which is located in the C-terminal penultimate position. In SelO homologs from other organisms, Sec is replaced with Cys. SelI is a recently evolved selenoprotein specific to vertebrates [14]. This membrane selenoprotein has no known function.

9.3.11 *Selenoprotein K (SelK) and S (SelS)*

SelK and SelS are unusual among selenoproteins in that they do not have a pronounced secondary structure [14]. These small selenoproteins contain a single transmembrane helix in the N-terminal sequence that targets them to the ER membrane. SelK homologs were detected in many eukaryotes, but no information is available on the function of any of these proteins. In contrast, recent studies revealed the role of SelS in retrotranslocation of misfolded proteins from the ER to the cytosol, where these proteins are further degraded [46]. SelS binds Derlin 1, an ER membrane-resident protein. In addition, SelS was implicated in inflammation and the immune response. A SelK knockout mouse model was recently developed [47] and is discussed elsewhere in the book.

9.3.12 *Selenoprotein N (SelN)*

One of the first selenoproteins discovered through bioinformatics approaches [10], SelN remains a selenoprotein of unknown function. This protein was implicated in the role of Se in muscle function through biochemical and genetic analyses, as well as through analyses of knockout mice [48], and was found to serve as a cofactor for the ryanodine receptor [49].

9.4 Additional Selenoproteins in Eukaryotes

The following selenoproteins that are absent in mammals were identified in eukaryotes: methionine-*S*-sulfoxide reductase (MsrA), protein disulfide isomerase (PDI), selenoproteins U (SelU), L (SelL), J (SelJ), Fep15, MCS, plasmodial selenoproteins Sel1, Sel2, Sel3 and Sel4, and a selenoprotein SelTryp in *Trypanosoma*. MsrA is a widely distributed protein family, whose function is to repair methionine residues in proteins. Like MsrB, it catalyzes a stereospecific reduction of methionine sulfoxides, but is specific for methionine-*S*-sulfoxides. MsrA was initially found in the green algae, *Chlamydomonas* [8], but later was also identified in other eukaryotes as well as in some bacteria. PDI is also very narrowly distributed in eukaryotes [50], in contrast to Cys-containing PDIs, which are essential for formation of disulfide bonds in the ER of eukaryotic cells. SelU [51], SelJ [52], Fep15 [53], and SelL [54] were only found in fish and/or invertebrates. The four *Plasmodium* selenoproteins (Sel1–Sel4) show no detectable homology to any other proteins [55]. However, Sel1 and Sel4 have Sec in the C-terminal regions and may be related to SelK and SelS.

9.5 Prokaryotic Selenoproteins

Several selenoproteins discussed above, including selenophosphate synthetase, deiodinase homologs, glutathione peroxidase and SelW, occur in both prokaryotes and eukaryotes. Below, we briefly discuss selenoproteins specific for prokaryotes.

9.5.1 Formate Dehydrogenase (FDH)

FDH is the most widespread prokaryotic selenoprotein. Sec in this protein is coordinated to molybdenum and directly involved in the oxidation of formate to carbon dioxide [56, 57]. In many bacteria, FDH is the only selenoprotein, which may be responsible for maintaining the Sec trait in these organisms [58].

9.5.2 Hydrogenase

Several hydrogenases are known that contain Sec. In these proteins, Sec is bound to nickel and is directly involved in catalysis [59]. Two different hydrogenase subunits may contain Sec, including one which may have two Sec residues [60].

9.5.3 *Formylmethanofuran Dehydrogenase (FMDH)*

FMDH is a distant homolog of FDH and catalyzes a similar reaction (with formylmethanofuran as the substrate) [61]. As in FDH, Sec in FMDH is coordinated to molybdenum in the enzyme active site.

9.5.4 *Selenoproteins A (GrdA) and B (GrdB)*

GrdA is a selenoprotein component of a multiprotein glycine reductase complex in certain bacteria [62]. This is currently the only known prokaryotic selenoprotein for which no Cys homologs have been detected [38]. GrdB is a selenoprotein component of multiprotein complexes involved in the reduction of glycine, sarcosine, betaine, and other substrates [63–65]. GrdB proteins are substrate-specific and bind a single GrdA.

9.5.5 *Thioredoxin-Like Selenoproteins*

Peroxiredoxins (Prxs), thioredoxins (Trxs), and glutaredoxins (Grxs) are abundant Cys-containing proteins that are present in essentially all organisms. However, some bacteria contain Sec-containing forms of these proteins [15, 16, 66–68]. Especially in bacteria, there are a variety of selenoproteins of the thioredoxin fold with distant homology to Prx, Trx, or Grx.

9.5.6 *HesB-Like*

This distant homolog of HesB proteins (also known as IscA) is a selenoprotein only present in certain archaea and bacteria [15]. HesB/IscA proteins are involved in iron-sulfur cluster biosynthesis, but the function of their selenoprotein homolog has not been characterized.

9.5.7 *Additional Prokaryotic Selenoproteins*

Additional prokaryotic selenoproteins are listed in Fig. 9.1. Most of these proteins are homologs of thiol-dependent oxidoreductases, in which the catalytic Cys is replaced with Sec [15, 16]. There are also numerous predicted bacterial selenoproteins of unknown function [68, 69].

9.6 Selenoprotein Functions

From the brief description of selenoprotein functions, it is apparent that selenoproteins for which functions are known are oxidoreductases. In these proteins, Sec is the catalytic residue that is employed because it is superior to Cys in this function [38]. In selenoproteins, Sec reversibly changes its redox state during catalysis. Functions of many selenoproteins, particularly those found in vertebrates, are not known. However, by analogy to proteins with known functions, it may be expected that the majority of these uncharacterized selenoproteins are also oxidoreductases.

All selenoproteins may be loosely clustered into three protein groups. The most abundant selenoprotein group includes proteins containing Sec in the N-terminal region or in the middle of the protein. Many of these selenoproteins exhibit thioredoxin or thioredoxin-like folds, but some proteins (e.g., SelD, MsrA) show different folds. In these proteins, Sec is the catalytic group, which often works in concert with a resolving Cys.

In the second group, Sec is located in the C-terminal sequences. These proteins so far have been described only in eukaryotes and include selenoproteins K, S, O, I, and TRs. Except for TRs, the function of Sec in selenoproteins in this group is not known.

Selenoproteins in the third group utilize Sec to coordinate redox metals (Mo, W, Ni) in the active sites of these proteins. This protein class includes hydrogenase, FDH, and FMDH.

However, non-catalytic functions of Sec, while rare, may be expected. Known examples include Sec residues in the C-terminal region of SelP (they function to transport Se) and recently evolved Sec residues in the *Metridium senile* MsrB homolog (the function of these Sec residues is not known) [70].

9.7 Selenoproteomes

Bioinformatics analyses allowed the identification of all or almost all selenoproteins in a variety of organisms [71]. The data involving full sets of selenoproteins in organisms (selenoproteomes) provide an opportunity to address numerous questions relevant to the biology of Se. This information helps explain the biological and biomedical effects of dietary Se. This is because it is now possible to link individual selenoproteins or selenoprotein groups with the specific effects of dietary Se. In this respect, Se is ahead of the studies involving other trace elements (as well as vitamins and other biofactors) where new proteins are still discovered biochemically (and often by accident) and where full sets of proteins dependent on a particular biofactor is difficult to ascertain.

Searches of the nematode selenoproteomes revealed that *C. elegans* and *C. briggsae* have only a single UGA codon that codes for Sec in their genomes [72]. This codon is present in the TR1 gene, and phylogenetic analyses suggested that

other selenoprotein genes were lost in these nematodes during evolution. Recently, the first selenoproteinless animals were identified, all of which are arthropods (mostly insects, such as beetles and silkworms) [73, 74]. Information about such animals (or other organisms that lost selenoproteins, such as yeast and higher plants) helps explain the changing requirements for Se during evolution. For example, selenoproteinless insects lost the entire Sec insertion machinery, but preserved SPS1, suggesting that this protein does not provide a Se intermediate for Sec biosynthesis [73]. It is an exciting possibility that SPS1 is involved in some other Se-dependent pathway.

Recent characterization of selenoproteomes of nematodes [71], fruit flies [12, 13], mammals [14], other vertebrates [51, 52], Apicomplexan parasites [54], and numerous other organisms including bacteria and archaea [15, 68–75], provided many clues with regard to the use of Se in these organisms. A recent study identified 57 selenoproteins in the harmful alga *Aureococcus anophagefferens*, which is currently the eukaryote with the largest number of selenoprotein genes [76]. Among prokaryotes, this record belongs to a symbiotic deltaproteobacterium of the gutless worm *Olavius algarvensis*, which also has 57 selenoprotein genes [77].

Rapid progress in genome sequencing should allow application of bioinformatics tools to many additional genome projects. It should be noted that environmental genome projects are also amenable to these applications [16]. For example, one recent study characterized the selenoproteome of the microbial marine community derived from the Global Ocean Sampling expedition [69]. More than 3,600 selenoprotein gene sequences belonging to 58 protein families were detected. Geographic location had little influence on Sec utilization, but higher temperature and marine (as opposed to freshwater and other aquatic) environments were associated with the increased use of this amino acid. This study provided insights into global trends in microbial Se utilization in marine environments.

Selenoproteome analyses also are capable of uncovering trends in the use of Sec [78], although some limitations of this approach have been described [79]. An analysis of selenoproteomes of several model eukaryotes detected 26–29 selenoprotein genes in two species of *Ostreococcus*, 5 in the social amoebae *Dictyostelium discoideum*, and 16 in the diatom *Thalassiosira pseudonana*, including several new selenoproteins [78]. Further analyses identified massive, independent selenoprotein losses in land plants, fungi, nematodes, insects, and some protists. Comparative analyses of selenoprotein-rich and -deficient organisms revealed that aquatic organisms generally have large selenoproteomes, whereas several groups of terrestrial organisms reduced their selenoproteomes through loss of selenoprotein genes and replacement of Sec with Cys. These observations suggested that many selenoproteins originated at the base of the eukaryotic domain and showed that the environment may play a role in selenoproteome evolution. In particular, aquatic organisms apparently retained and sometimes expanded their selenoproteomes, whereas the selenoproteomes of some terrestrial organisms were reduced or completely lost. It is an interesting possibility that aquatic life supports Se utilization, whereas terrestrial habitats lead to reduced use of this trace element [78].

In a separate study involving vertebrates, reconstruction of evolutionary changes in the Se transport domain of SelP revealed a decrease in the Sec content specifically in the mammalian lineage via replacement of Sec with Cys [41]. Compared to mammals, fish showed higher Sec content of SelP, larger selenoproteomes, elevated SelP gene expression, and higher levels of tissue Se. In addition, mammals replaced Sec with Cys in several proteins and lost several selenoproteins altogether, whereas such events were not found in fish. These data suggested that evolution from fish to mammals was accompanied by decreased use of Sec and that analyses of SelP, selenoproteomes, and Sec/Cys transitions provide a genetic marker of utilization of this trace element in vertebrates. The evolved reduced reliance on Se raises questions regarding the need to maximize selenoprotein expression by Se dietary supplements in situations when pathology is not imminent, a currently accepted practice.

9.8 Applications of Selenoproteome Analyses to Biology

9.8.1 *Genetic Code Supports Targeted Insertion of Two Amino Acids by One Codon*

Strict one-to-one correspondence between codons and amino acids was thought to be an essential feature of the genetic code. However, a recent selenoproteome analysis of the ciliate *Euplotes crassus* revealed that one codon can code for two different amino acids with the choice of the inserted amino acid determined by a specific 3'-untranslated region and location of the dual-function codon within the mRNA [80]. It was found that the codon UGA specified insertion of Sec and Cys in *E. crassus*, that the dual use of this codon could occur even within the same protein, and that the structural arrangements of *Euplotes* mRNA preserved the location-dependent dual-function of UGA when expressed in mammalian cells. Thus, the genetic code supports the use of one codon to code for multiple amino acids. This finding challenged one of the foundations of the code, i.e., that a genetic codeword is used only for one amino acid in an organism.

9.8.2 *High-Throughput Identification of Catalytic Redox-Active Cysteine (Cys) Residues*

Cys residues often play critical roles in proteins; however, identification of their specific functions has been limited to case-by-case experimental approaches. A recent study developed a procedure for high-throughput identification of catalytic redox-active Cys in proteins by searching for sporadic Sec/Cys pairs in sequence databases [81]. This method was independent of protein family, structure, and taxon. It was used to selectively detect the majority of known proteins with redox-active Cys and

to make additional predictions, one of which was verified. Rapid accumulation of sequence information from genomic and metagenomic projects, coupled with selenoproteome analyses, should allow detection of many additional oxidoreductase families as well as identification of redox-active Cys in these proteins.

9.9 Conclusions

Fifteen years ago, only several selenoproteins were known. Largely due to remarkable progress in genomics, we now know approximately 100 selenoprotein families. This information allows researchers to study various aspects of Se biology and selenoprotein functions and address questions, not even imaginable until only recently, such as geographical distribution of selenoprotein utilization or expansion of the genetic code. In selenoproteins with known functions, Sec is a key functional group that carries out redox catalysis. Further studies of selenoproteins and selenoproteomes should help explain known biological and biomedical effects of Se and identify new biological processes and pathways dependent on this trace element.

Acknowledgements This work was supported by National Institutes of Health grants.

References

1. Wolfe MD, Ahmed F, Lacourciere GM et al (2004) *J Biol Chem* 279:1801
2. Dilworth GL (1982) *Arch Biochem Biophys* 219:30
3. Gladyshev VN, Khangulov SV, Stadtman TC (1994) *Proc Natl Acad Sci USA* 91:232
4. Self WT, Stadtman TC (2000) *Proc Natl Acad Sci USA* 97:7208
5. Ballihaut G, Mounicou S, Lobinski R (2007) *Anal Bioanal Chem* 388:585
6. Behne D, Kyriakopoulos A, Meinhold H et al (1990) *Biochem Biophys Res Commun* 173:1143
7. Gladyshev VN, Jeang KT, Wootton JC et al (1998) *J Biol Chem* 273:8910
8. Novoselov SV, Rao M, Onoshko NV et al (2002) *EMBO J* 21:3681
9. Kryukov GV, Kryukov VM, Gladyshev VN (1999) *J Biol Chem* 274:33888
10. Lescure A, Gautheret D, Carbon P et al (1999) *J Biol Chem* 274:38147
11. Gladyshev VN, Kryukov GV, Fomenko DE et al (2004) *Annu Rev Nutr* 24:579
12. Martin-Romero FJ, Kryukov GV, Lobanov AV et al (2001) *J Biol Chem* 276:29798
13. Castellano S, Morozova N, Morey M et al (2001) *EMBO Rep* 2:697
14. Kryukov GV, Castellano S, Novoselov SV et al (2003) *Science* 300:1439
15. Kryukov GV, Gladyshev VN (2004) *EMBO Rep* 5:538
16. Zhang Y, Fomenko DE, Gladyshev VN (2005) *Genome Biol* 6:R37
17. Rotruck JT, Pope AL, Ganther HE et al (1973) *Science* 179:588
18. Seiler A, Schneider M, Förster H et al (2008) *Cell Metab* 8:237
19. Ursini F, Heim S, Kiess M et al (1999) *Science* 285:1393
20. Olson GE, Whitin JC, Hil KE et al (2010) *Am J Physiol Renal Physiol* 298:F1244
21. Gromer S, Wissing J, Behne D et al (1998) *Biochem J* 332:591
22. Sun QA, Kirnarsky L, Sherman S et al (2001) *Proc Natl Acad Sci USA* 98:3673
23. Sandalova T, Zhong L, Lindqvist Y et al (2001) *Proc Natl Acad Sci USA* 98:9533

24. Sun QA, Wu Y, Zappacosta F et al (1999) *J Biol Chem* 274:24522
25. Arner ES, Holmgren A (2000) *Eur J Biochem* 267:6102
26. Sun QA, Zappacosta F, Factor V et al (2001) *J Biol Chem* 276:3106
27. Rundlof AK, Janard M, Miranda-Vizuete A et al (2004) *Free Radic Biol Med* 36:641
28. Su D, Gladyshev VN (2004) *Biochemistry* 43:12177
29. Su D, Novoselov SV, Sun QA et al (2005) *J Biol Chem* 280:26491
30. Conrad M, Jakupoglu C, Moreno SG et al (2004) *Mol Cell Biol* 24:9414
31. Jakupoglu C, Przemecck GK, Schneider M et al (2005) *Mol Cell Biol* 25:1980
32. Kryukov GV, Kumar RA, Koc A et al (2002) *Proc Natl Acad Sci USA* 99:4245
33. Kim HY, Gladyshev VN (2004) *Mol Biol Cell* 15:1055
34. Fomenko DE, Novoselov SV, Natarajan SK et al (2009) *J Biol Chem* 284:5986
35. Korotkov KV, Kumaraswamy E, Zhou Y et al (2001) *J Biol Chem* 276:15330
36. Kumaraswamy E, Malykh A, Korotkov KV et al (2000) *J Biol Chem* 275:35540
37. Ferguson AD, Labunskyy VM, Fomenko DE et al (2005) *J Biol Chem* 281:3536
38. Stadtman TC (1996) *Annu Rev Biochem* 65:83
39. Xu XM, Carlson BA, Irons R et al (2007) *Biochem J* 404:115
40. Burk RF, Hill KE (2005) *Annu Rev Nutr* 25:215
41. Lobanov AV, Hatfield DL, Gladyshev VN (2008) *Genome Biol* 9:R62
42. Hill KE, Zhou J, McMahan WJ et al (2003) *J Biol Chem* 278:13640
43. Schomburg L, Schweizer U, Holtmann B et al (2003) *Biochem J* 370:397
44. Vendeland SC, Beilstein MA, Yeh JY et al (1995) *Proc Natl Acad Sci USA* 92:8749
45. Korotkov KV, Novoselov SV, Hatfield DL et al (2002) *Mol Cell Biol* 22:1402
46. Ye Y, Shibata Y, Yun C et al (2004) *Nature* 429:841
47. Verma S, Hoffmann FW, Kumar M et al (2011) *J Immunol* 186:2127
48. Moghadaszadeh B, Petit N, Jaillard C (2001) *Nat Genet* 29:17
49. Juryneć MJ, Xia R, Mackrill JJ et al (2008) *Proc Natl Acad Sci USA* 105:12485
50. Obata T, Shiraiwa Y (2005) *J Biol Chem* 280:18462
51. Castellano S, Novoselov SV, Kryukov GV et al (2004) *EMBO Rep* 5:71
52. Castellano S, Lobanov AV, Chapple C et al (2005) *Proc Natl Acad Sci USA* 102:16188
53. Novoselov SV, Hua D, Lobanov AV et al (2005) *Biochem J* 394:575
54. Shchedrina VA, Novoselov SV, Malinouski MY et al (2007) *Proc Natl Acad Sci USA* 104:13919
55. Lobanov AV, Delgado C, Rahlfs S et al (2006) *Nucl Acids Res* 34:496
56. Gladyshev VN, Khangulov SV, Axley MJ et al (1994) *Proc Natl Acad Sci USA* 91:7708
57. Boyington JC, Gladyshev VN, Khangulov SV et al (1997) *Science* 275:1305
58. Romero H, Zhang Y, Gladyshev VN et al (2005) *Genome Biol* 6:R66
59. Garcin E, Vernede X, Hatchikian EC et al (1999) *Structure* 7:557
60. Wilting R, Schorling S, Persson BC et al (1997) *J Mol Biol* 266:637
61. Vorholt JA, Vaupel M, Thauer RK (1997) *Mol Microbiol* 23:1033
62. Garcia GE, Stadtman TC (1991) *J Bacteriol* 173:2093
63. Wagner M, Sonntag D, Grimm R et al (1999) *Eur J Biochem* 260:38
64. Andreesen JR, Wagner M, Sonntag D et al (1999) *Biofactors* 10:263
65. Schrader T, Rienhofer A, Andreesen JR (1999) *Eur J Biochem* 264:862
66. Sohling B, Parther T, Rucknagel KP et al (2001) *Biol Chem* 382:979
67. Hatfield DL, Gladyshev VN (2002) *Mol Cell Biol* 22:3565
68. Zhang Y, Romero H, Salinas G et al (2006) *Genome Biol* 7:R94
69. Zhang Y, Gladyshev VN (2008) *PLoS Genet* 4:e1000095
70. Lee BC, Lobanov AV, Marino SM et al (2011) *J Biol Chem* 286:18747
71. Lobanov AV, Hatfield DL, Gladyshev VN (2009) *Biochim Biophys Acta* 1790:1424
72. Taskov K, Chapple C, Kryukov GV et al (2005) *Nucleic Acids Res* 33:2227
73. Lobanov AV, Hatfield DL, Gladyshev VN (2008) *Protein Sci* 17:176
74. Chapple CE, Guigó R (2008) *PLoS One* 3:e2968

75. Zhang Y, Gladyshev V (2005) *Bioinformatics* 21:2580
76. Gobler CJ, Berry DL, Dyhrman ST et al (2011) *Proc Natl Acad Sci USA* 108:4352
77. Zhang Y, Gladyshev VN (2007) *Nucleic Acids Res* 35:4952
78. Lobanov AV, Fomenko DE, Zhang Y et al (2007) *Genome Biol* 8:R198
79. Castellano S, Andrés AM, Bosch E et al (2009) *Mol Biol Evol* 26:2031
80. Turanov AA, Lobanov AV, Fomenko DE et al (2009) *Science* 323:259
81. Fomenko DE, Xing W, Adair BM et al (2007) *Science* 315:387