

Chapter 8

Evolution of Selenophosphate Synthetase

Marco Mariotti, Didac Santesmasses, and Roderic Guigó

Abstract Selenophosphate synthetase (SPS or SelD) provides the active selenium (Se) donor for the synthesis of selenocysteine (Sec), the 21st amino acid in the genetic code. In this chapter we summarize the distribution, phylogeny and function of all SPS genes across the tree of life. SPS is a selenoprotein itself in many prokaryotes (SelD) and eukaryotes (SPS2). As most other selenoproteins, SPS has orthologs with cysteine (Cys) in place of Sec. Although absent in many lineages, selenoproteins and SPS occur in bacteria, archaea and eukaryotes. In prokaryotes, SPS supports additional forms of Se utilization besides Sec, most notably the use of selenouridine in tRNAs. The study of selenophosphate synthetases, while serving a map of Se utilization across all sequenced organisms, also highlighted examples of functional diversification within this family. Within archaea, a few *Crenarchaeota* species exhibit a *SelD*-like gene. This is derived from *SPS*, but probably carries a different function, since it never co-occurs with other Se utilization genes. Within eukaryotes, many metazoan genomes, including humans, carry a paralog called *SPS1* in addition to *SPS2*, which replaces the Sec/Cys site with some other amino acid (i.e., threonine, arginine, glycine, or leucine). Strikingly, *SPS1* genes were generated through distinct gene duplication events of *SPS2* in several metazoan clades (e.g., vertebrates and insects). Their function is still unknown, but it appears to be

M. Mariotti (✉)

Division of Genetics, Department of Medicine, Brigham and Women's Hospital,
Harvard Medical School, Boston, MA 02115, USA

Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG),
The Barcelona Institute of Science and Technology, Barcelona 08003, Catalonia, Spain

Universitat Pompeu Fabra (UPF), Barcelona 08003, Catalonia, Spain

Institut Hospital del Mar d'Investigacions Mèdiques (IMIM),
Barcelona 08003, Catalonia, Spain

e-mail: marco.mariotti.mm@gmail.com

D. Santesmasses • R. Guigó

Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG),
The Barcelona Institute of Science and Technology, Barcelona 08003, Catalonia, Spain

Universitat Pompeu Fabra (UPF), Barcelona 08003, Catalonia, Spain

Institut Hospital del Mar d'Investigacions Mèdiques (IMIM),
Barcelona 08003, Catalonia, Spain

different from that of SPS2. Despite independently originated in parallel lineages, *SPS1* genes were shown to share a common function. Thus, this function was likely already present in the parental *SPS2* gene, driving the gene duplication events by sub-functionalization.

Keywords Evolution • Rescue • SeID • Selenocysteine • Selenium utilization • Selenophosphate synthetase • Selenoprotein • SPS • Sub-functionalization

8.1 Introduction

The main biological functions of selenium (Se) in human health are exerted by the non-standard amino acid selenocysteine (Sec), inserted co-translationally into selenoproteins [1]. Selenoproteins are found in organisms across the tree of life. Sec is encoded by in-frame UGA codons present in selenoprotein mRNAs and is incorporated through a dedicated machinery upon recognition of RNA structures acting as cis-signals (SECIS elements, for *Sec* insertion sequences). Sec synthesis occurs on the cognate tRNA (tRNA^{Sec}), in a multi-step process that requires monoselenophosphate. Although showing obvious homology, the mechanisms of Sec biosynthesis and insertion exhibit important differences between bacteria, archaea and eukaryotes (reviewed in [2–5]). Selenophosphate synthetase (SPS, or SeID in prokaryotes) is responsible for providing the active Se donor for the synthesis of Sec, activating selenide to generate selenophosphate at the expense of ATP [6]. *SPS* genes are conserved in all known Sec-encoding genomes, with a remarkable ~30% sequence identity between *Escherichia coli* and *Homo sapiens* ([7] and see Fig. 8.1).

The three-dimensional structure of SPS proteins [8–11] is highly conserved from bacteria to eukaryotes. SPS acts as a dimer and ATP is bound at the subunit interface, with the binding mediated by magnesium atoms. SPS hydrolyzes ATP subsequently to ADP and then AMP. A mobile N-terminal segment binds the substrate and holds it throughout the whole reaction. The residue responsible for this is thought to be a cysteine (Cys) located in a glycine-rich loop on the N-terminal domain. This same Cys residue is replaced by Sec in many organisms [7, 12], making SPS the only Sec machinery protein that is itself also a selenoprotein. As explained later in this chapter, several metazoans evolved a second, paralogous gene (*SPS1*), carrying other amino acids at this position. *SPS1* proteins appear to have a function distinct from selenophosphate synthesis. In all *bona fide* selenophosphate synthetase proteins (i.e., SeID, SPS2), this position is instead conserved with either Sec or Cys. Given their broad phylogenetic distribution, Sec utilization and thus SPS are arguably ancestral traits, already present in the last universal common ancestor. Selenophosphate synthetase, as both a constitutive part of Sec machinery and ancestral selenoprotein family, can be used as a proxy to follow the evolution of Sec utilization across the tree of life.

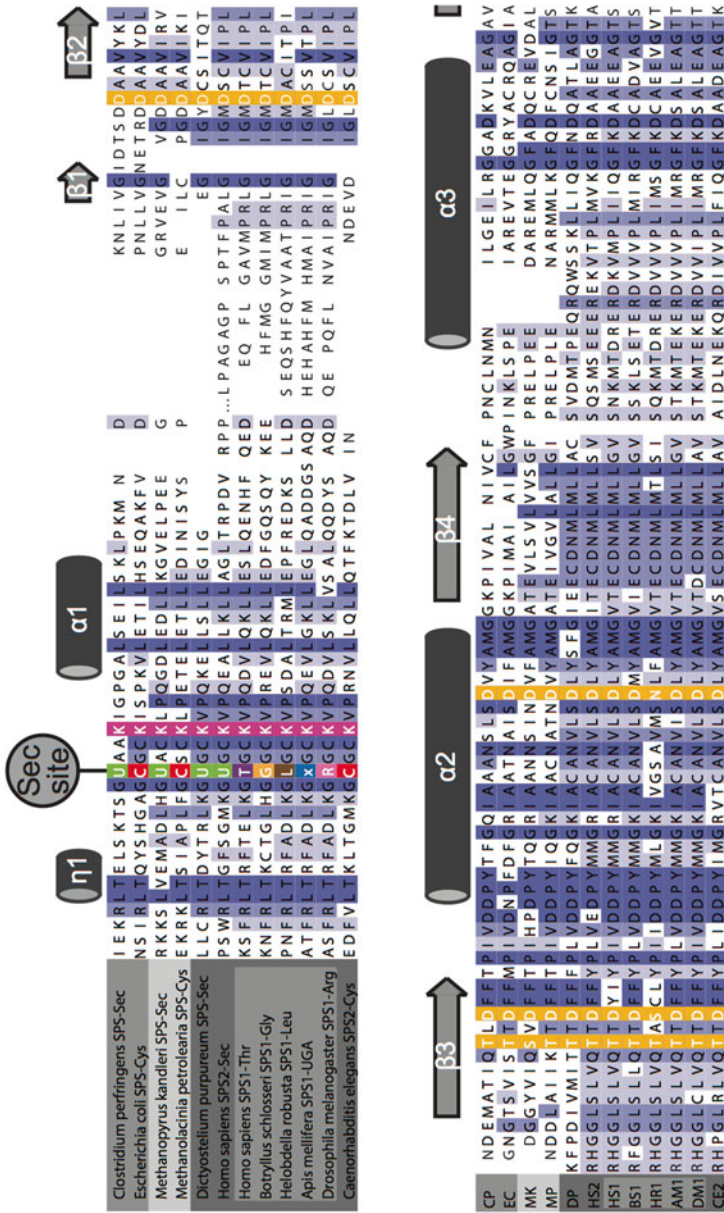
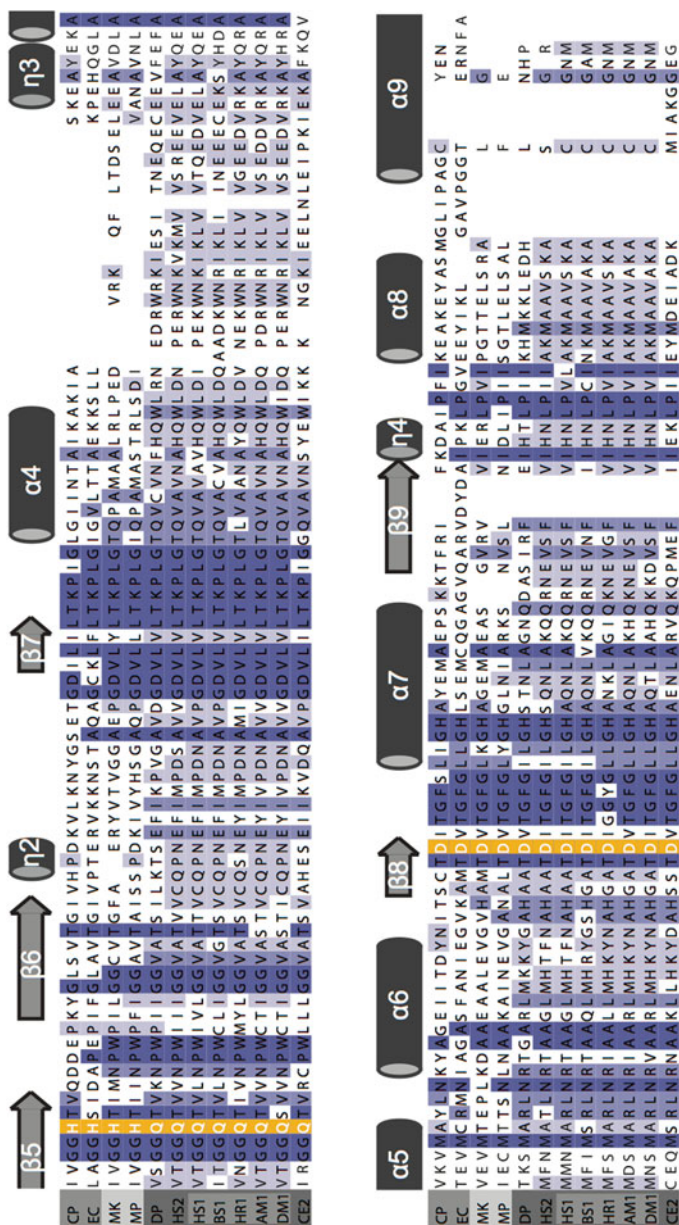


Fig. 8.1 Alignment of SPS protein sequences. This figure contains representative sequences from bacteria, archaea and eukaryotes, including various metazoan *SPS1* genes. *B. lue* is used to represent the degree of conservation at each position. Above, the secondary structure inferred from the solved crystal structure in [10] is displayed. Additionally, a few columns of the alignment are highlighted: the Sec site is colored differently after each residue found here, following the color scheme of Figs. 8.2 and 8.3; the nearby essential lysine is in magenta; the residues involved in metal-mediated ATP interactions are in orange. Sequences were trimmed at N and C termini, and a few additional positions were removed to improve visualization (e.g., a short fragment was removed from the bacterial sequences after helix $\alpha 9$, which contained the β -sheet and helix named $\beta 10$ and $\alpha 10$ in [10])



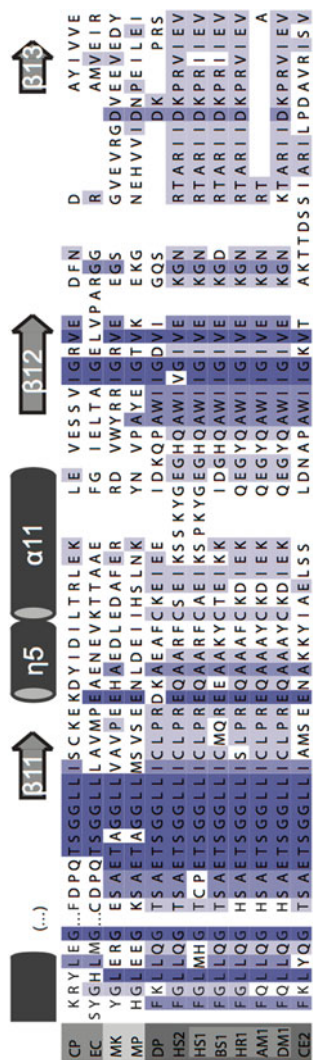


Fig. 8.1 (continued)

8.2 SPS Supports Diverse Se Utilization Traits in Prokaryotes

8.2.1 Bacteria

Those bacteria that utilize Se are characterized by multiple pathways (see Fig. 8.2). SPS proteins, found in 26–38% of sequenced bacteria [7, 13], are considered a fundamental genetic indicator of any of the Se utilization traits. In literature, all SPS-dependent pathways (Sec, SeU, etc.) are referred to as “Se utilization traits” [13–17]. We will follow this nomenclature, but the reader should be aware that this excludes other more “passive” Se pathways in which there is no activation to selenophosphate (e.g., Se detoxification in plants). In the last several years, numerous studies exploited the presence of SPS in combination with other gene markers to profile the various forms of Se utilization across genomes [7, 13–17].

The most common form of Se utilization is the insertion of Sec in selenoproteins (18–25% of sequenced bacteria [7, 13]). Besides *SPS*, Sec-containing organisms possess also the rest of the Sec machinery genes (*tRNA^{Sec}* or *SelC*, *SelA*, *SelB*), as well as one or more selenoprotein genes. The second most common form of Se utilization (16–22% of sequenced bacteria [7, 13]) is the use of 5-methylaminome thyl-2-selenouridine (SeU) at the anticodon wobble position of certain tRNAs [18]. SeU is synthesized by 2-selenouridine synthase (YbbB), which together with SPS acts as marker of SeU usage in genomes. A third Se utilization trait was recently discovered through the analysis of genomes with an “orphan” *SPS* gene, i.e., not coupled with any other marker of known Se utilization [16, 19]. It appears that Se is used by these organisms as a cofactor to certain molybdenum-containing hydroxylases [20]. Two genes, *YqeB* and *YqeC*, seem to be required in this process, and thus can be used in conjunction with *SPS* as genetic markers for this form of Se utilization [13, 17]. An estimated eight percent of bacteria possess this trait. It is entirely possible that other forms of Se utilization exist besides these three, and they may be uncovered as more and more sequences become available. As a matter of fact, a recent study showed that, even taking into account the markers for Sec, SeU and Se-cofactor, there are a few species that possess an orphan SPS [13]. These findings, if confirmed, would point to the existence of additional forms of Se utilization.

The known Se utilization traits significantly overlap: many bacterial species possess more than one trait. *SPS* genes are found in all genomes with at least one of these traits. Only in Sec-encoding bacteria, SPS is sometimes present as a selenoprotein. In all other cases, SPS contains Cys at the Sec homologous site. When it contains Sec, the *SPS* gene includes a SECIS element typical of bacteria, which is a small hairpin-loop structure overlapping the Sec-coding UGA.

In many bacterial lineages, SPS is found fused with other genes, a frequent feature of bacterial proteins. Two types of SPS fusions are particularly common with: i) a NADH-dehydrogenase-like domain; and ii) a NifS-like domain (Cys sulfinate

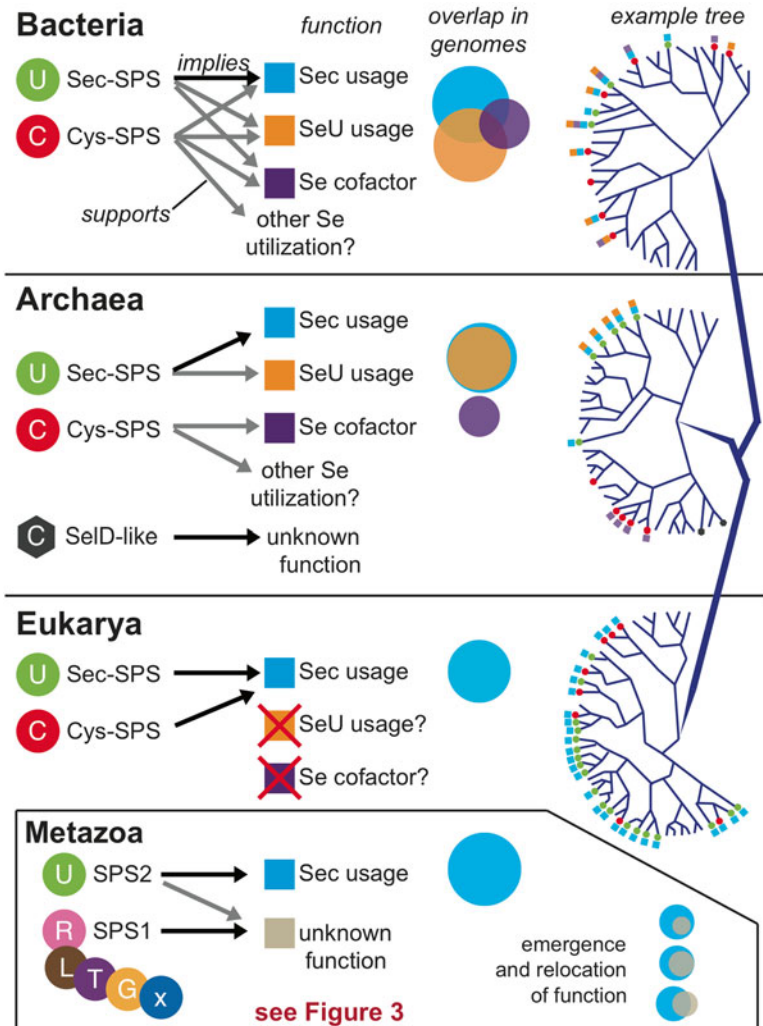


Fig. 8.2 SPS and Se utilization in bacteria, archaea and eukaryotes. The schema summarizes the genome distribution and function of SPS proteins. Each SPS gene is represented by a colored circle enclosing one letter, which indicates the amino acid found at the Sec site (e.g., C: Cys; U: Sec). SPS supports diverse ways of Se utilization in prokaryotes. In bacterial genomes, these traits show a complex mosaic pattern of overlap. The same traits are found in archaea, but their distribution appears limited to specific lineages. In eukaryotes, Sec/Cys SPS genes are always accompanied by the rest of Sec machinery, suggesting that Sec usage is the only eukaryotic Se utilization trait. Some archaea possess a protein family derived from SPS, called SelD-like [17] (grey hexagon). Although functionally uncharacterized, there are indications that SelD-like may be involved in a process other than selenophosphate synthesis. In metazoans, SPS1 genes (see Fig. 8.3) are known to carry a function different from selenophosphate synthesis, which is still unknown

desulfhinase) [7, 16]. In all of these cases, SPS is found at the C-terminal side of the fused polyprotein. Also, in all SPS fusions, SPS does not carry Sec, with a single known exception (SPS-NifS of *Geobacter sp. FRC-32*).

8.2.2 Archaea

Se utilization appears to be less common in archaea than it is in bacteria, and is strictly limited to specific lineages, in contrast to the mosaic pattern in bacteria (see Fig. 8.2). SPS proteins are found in only 12% of sequenced archaea [17]. To date, Sec utilization has been observed only in the orders of *Methanococcales* (15 genome sequences available) and *Methanopyrales* (with *Methanopyrus kandleri* being its only sequenced species). All archaea in these two orders possess a specific set of selenoproteins involved in hydrogenotrophic methanogenesis, plus a Sec-containing SPS protein [4, 7, 13, 17]. SPS genes contain an archaeal SECIS element in the 3'UTR, as expected for archaeal selenoproteins [4]. All *Methanococcales* possess a peculiar bipartite form of YbbB and utilize SeU in tRNAs [21], while *M. kandleri* lacks YbbB and therefore does not use SeU. In fact, SeU utilization appears to be absent from all archaeal lineages other than *Methanococcales* investigated thus far. SPS genes are found also in some archaeal species belonging to the order of *Halobacteriales* [13, 17]. The co-occurrence of the genes *YqeB* and *YqeC* in these genomes suggests that these organisms use Se as a cofactor. In all of these species, SPS is present in a Cys form. Another archaeal genome, *Methanolacinia petrolearia* (previously known as *Methanoplanus petrolearius*), belonging to the order of *Methanomicrobiales*, also possesses a Cys form of SPS, but this is not coupled to any other Se utilization marker [13, 17]. This orphan SPS may hint at yet another undiscovered form of Se utilization in archaea.

A recent study [17] reported a novel protein family related to SPS, named SelD-like. These proteins are similar in sequence to *bona fide* SPS, but they form a separate phylogenetic cluster. SPS and SelD-like most likely share a common ancestor, and possibly a common catalytic mechanism. The SPS Sec/Cys site is conserved in SelD-like with Cys. Furthermore, conservation of critical residues suggests that most likely SelD-like proteins bind certain metal atoms (e.g., magnesium) as well as ATP or its analogues, similarly to SPS. *SelD*-like genes were found uniquely in certain species within the *Crenarchaeota* phylum, and more specifically in the orders of *Thermoproteales* and *Sulfolobales*. No Se utilization marker is found in these genomes, or in any other sequenced *Crenarchaeota*. In *Thermoproteales*, SelD-like is fused to an acylphosphatase-like protein, located at the C-terminus. Although acylphosphatase genes are found throughout the tree of life, fusions with SPS were never observed. To date, the biological function of SelD-like proteins remains unknown, but based on gene co-occurrence, it was speculated that it may be involved in sulfur metabolism [17]. It cannot be excluded, however, that SelD-like proteins are just divergent selenophosphate synthetases implicated in yet another form of Se utilization.

8.3 SPS in Eukaryotes

In eukaryotes, *SPS* genes are always accompanied by the rest of the Sec machinery and by selenoprotein genes in the same genome (see Fig. 8.2). The only exceptions are certain *SPS1* genes, which are explained below. This suggests that, in eukaryotes, the insertion of Sec in selenoproteins may be the only Se utilization trait supported by SPS. The gene, *YbbB*, is in fact generally missing from eukaryotic genomes, with a few exceptions attributed to gene transfer from bacteria (unpublished data). Most likely, Sec utilization was directly transmitted through the line of descent from bacteria to archaea to eukaryotes, while the other Se utilization traits were not. Although several lineages are devoid of selenoproteins (e.g., fungi, land plants), Sec utilization is spread across the eukaryotic phylogenetic tree, and it is highly represented in most metazoans [22, 23]. As in prokaryotes, eukaryotic Sec-encoding genomes contain a *SPS* gene carrying either Sec or Cys at its homologous site. Since SPS is a selenoprotein in Sec-utilizing archaea, and since a Cys to Sec conversion has necessarily to be considered an event with very low probability, the most parsimonious explanation is that the last common ancestor of eukaryotes had a Sec-containing SPS.

The reconstructed tree of *SPS* genes from the three domains of life broadly follows their known phylogenetic relationships, which supports the scenario of direct inheritance of the Sec trait. There are few exceptions, however, as observed in a number of protist organisms, such as green algae, most of alveolates and certain amoebas, exhibit a bacterial-like SPS protein. These were likely acquired by horizontal gene transfer in relatively recent times [7]. Many of these transferred genes are fusions of SPS with other proteins. Giving support to the horizontal gene transfer hypothesis, the most common SPS fusions in protists are with a NADH-dehydrogenase-like domain, and with a NifS-like domain, as found in bacteria. Two additional cases of extended SPS were observed, which are unique to protists. First, the heterolobosean amoeba, *Naegleria gruberi*, has SPS fused to a methyltransferase domain [24]. Strikingly, the genome of this species also contains a second *SPS* gene, fused with a NifS-like domain. Second, all *Plasmodia* species have a single *SPS* gene with an additional large domain at its N-terminus (>500 residues). This domain shows no homology to any known proteins, and its function remains unknown.

8.3.1 Emergence of *SPS1* Genes in Metazoa

Almost every available eukaryotic genome outside metazoa possesses either a single *SPS* gene or none at all, and, when present, it contains Sec or Cys. In contrast, many metazoans, including *Homo sapiens* and *Drosophila melanogaster*, possess two distinct genes, *SPS1* and *SPS2*. The *SPS2* gene carries Sec at the usual site, and it was shown to synthesize selenophosphate from selenide and ATP, constituting the true

functional homologue of SPS in non-metazoans [25–27]. Instead, SPS1 proteins carry neither Sec nor Cys at this homologous site, whereas human SPS1 has threonine (SPS1-Thr) and fruit fly SPS1 carries arginine (SPS1-Arg). Diverse lines of evidence suggest that SPS1 proteins do not catalyze the canonical SPS function [26–29], although their molecular function has not been characterized thus far.

Notably, *SPS1-Thr* and *SPS1-Arg* are not phylogenetic homologues. Instead, they were generated through two distinct gene duplication events in different lineages. Accordingly, other similar cases have been observed. To date, we detected a total of four gene duplications across parallel lineages in metazoans [7] (see Fig. 8.3). In each case, the duplication involved the ancestral Sec-containing *SPS2* gene, and generated a novel *SPS* gene carrying some amino acid different from Sec or Cys at the homologous site. For convenience, we used the term SPS2 to designate all *bona fide* SPS genes, carrying Sec or Cys, and we use SPS1 to designate the other metazoan genes with any other amino acid. Despite their phylogenetically independent origin, various *SPS1* genes can partially rescue the phenotypic effect of *Drosophila* SPS1 knockout. This indicates that the diverse SPS1 proteins share a common function, yet distinct from that of SPS2 [7]. Thus, *SPS1* genes in different metazoan lineages seem to be functional homologues, despite the lack of direct phylogenetic orthology. We hypothesized that this may be explained by a process of parallel sub-functionalization. In this scenario, the ancestral metazoan *SPS2* gene had acquired a second function, on top of its canonical selenophosphate synthesis activity. Later in metazoan evolution, there was selective pressure to separate the two functions to two distinct genes. This resulted in the observed pattern of gene duplications across parallel lineages, with *SPS1* genes emerging to assume this second function, while *SPS2* retained just the original, canonical activity. Upon duplication, *SPS2* exhibited in each case a decrease of selective pressure on protein sequence, which we interpret as a signature of sub-functionalization. In contrast, the newly generated SPS1 proteins show a tight level of conservation in every case. The metazoan *SPS2* genes that never duplicated, which are expected to possess both SPS1 and SPS2 functions, consistently show a high level of selective pressure. The various metazoan *SPS* duplications, although analogous in outcome, occurred through diverse molecular mechanisms, as detailed below (see Fig. 8.3).

The *SPS1-Thr* gene is found only in vertebrates and is present in all available genomes in this lineage, with the species in the *Cyclostomata* group (jawless vertebrates, such as lampreys) being the sole exception. *SPS1-Thr* most likely originated through one of reported rounds of genome duplication at the root of vertebrates (Fig 8.3, duplication #1). *SPS1-Thr* maintained the overall gene structure of its parental *SPS2*, which features several introns in conserved positions. The *SPS2* gene later went through additional evolutionary events in mammals [30]. It appears that placentals replaced the original *SPS2* gene (*SPS2a*) with one of its retrotransposed copies (*SPS2b*). As a result, the only extant *SPS2* gene in placentals (including human and mouse) has no introns. Certain non-placental mammals, such as the marsupials, *Macropus eugenii* and *Monodelphis domestica*, still retain the two *SPS2* copies in their genome; i.e., *SPS2a* with the ancestral gene structure, and intronless *SPS2b*. It is not known whether both copies are functional in these organisms.

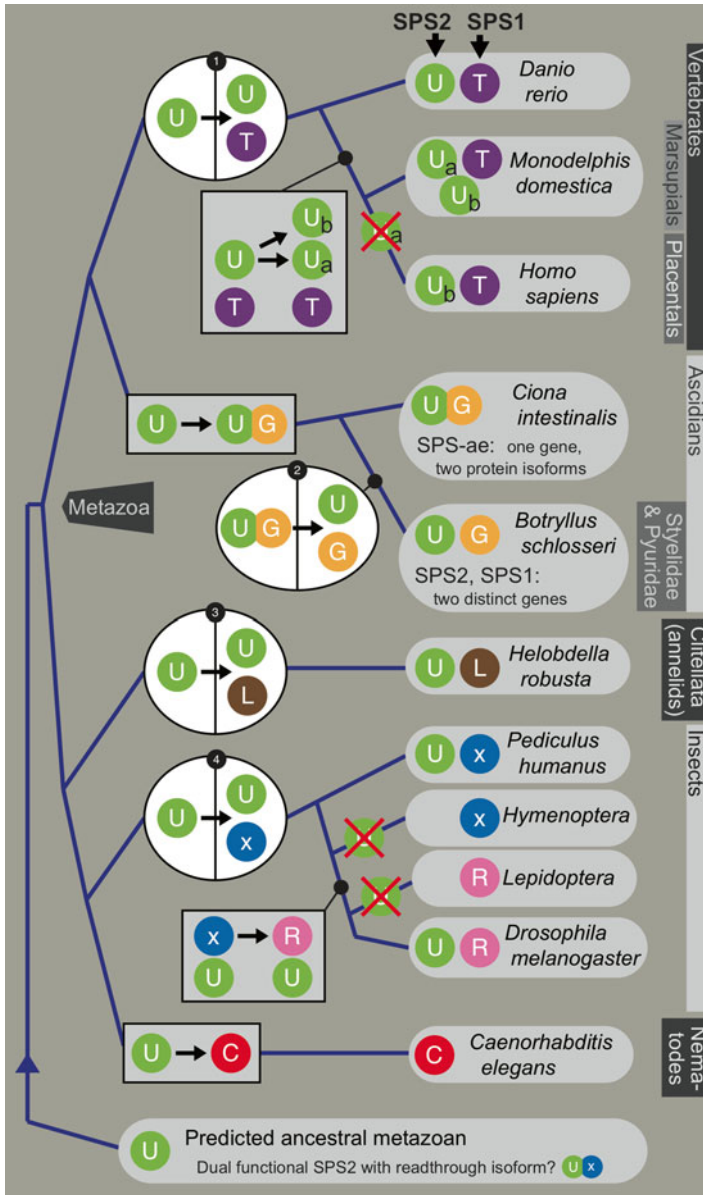


Fig. 8.3 Emergence of metazoan *SPS1* genes by parallel gene duplications of *SPS2*. The schema summarizes the reconstructed phylogeny of SPS proteins described in [7]. *SPS* genes are represented as in Fig. 8.2. *SPS2* genes carry Sec (U) or Cys (C, uniquely in nematodes), while *SPS1* genes carry other amino acids at this site (T: threonine, G: glycine, L: leucine, R: arginine; x: unknown residue). Gene duplications are highlighted in the *bipartite white circles*

Another *SPS1* gene emerged within tunicates (Fig 8.3, duplication #2). Here, the ascidians in the *Styelidae* and *Pyuridae* sister families exhibit a *SPS1* gene with a glycine residue aligned to the Sec site (*SPS1-Gly*). This gene presents the ancestral metazoan gene structure, featuring several introns. The same organisms also possess a *SPS2* gene with Sec, which, in contrast, has no introns. The analysis of other tunicate genomes allowed reconstructing how this situation came about [7]. While non-ascidian tunicates, such as *Oikopleura dioica*, display a single “standard” *SPS2* gene with the ancestral intron structure, the same gene has a peculiarity in the model species *Ciona intestinalis* and other ascidians. This gene (*SPS-ae*) produces two different transcript isoforms that differ in their 5' end, resulting in two distinct protein isoforms, a Sec-containing *SPS2*-like isoform, and a *SPS1*-like isoform with glycine. Whereas this dual gene configuration is found in most ascidians, at the root of *Styelidae* and *Pyuridae*, the *SPS2*-like transcript retrotransposed to the genome, generating a novel intronless *SPS2* gene. Hence, the *SPS-ae* gene then specialized into the *SPS1*-like isoform only, becoming *SPS1-Gly*.

Another *SPS1* gene emerged within annelids, in the class of *Clitellata* (Fig 8.3, duplication #3). This gene carries a leucine residue aligned to the homologous Sec position (*SPS1-Leu*). Since *SPS1-Leu* possesses the ancestral *SPS2* gene structure, it must have been generated by duplication of *SPS2* through a conservative mechanism, such as homologous recombination.

Lastly, another *SPS1* gene emerged in insects (Fig 8.3, duplication #4). The phylogenetic history of this gene is particularly complex, and it is complicated by the phenomenon of Sec extinctions in this lineage [31]. Insects, in fact, went through a progressive depletion of selenoprotein genes, culminating with the complete loss of the Sec trait in multiple lineages (see [32]). Several events of Sec extinction occurred independently across parallel insect lineages. Some of them affected entire taxonomic orders (e.g., *Hymenoptera*, *Lepidoptera*), while others happened more recently and affected only certain species (e.g., *Drosophila willistoni* and *Acyrtosiphon pisum* [33]). In each selenoprotein-less insect, *SPS2* was lost together with other parts of the Sec machinery, while *SPS1* was maintained. We traced the origin of the insect *SPS1* gene prior to all Sec extinctions, as far back as the last common ancestor of all insects. In our reconstruction [7], the insect *SPS1* gene first emerged by duplication of *SPS2* in a rather bizarre form, as it maintained its in-frame UGA codon (translated as Sec in *SPS2*), but it lost its SECIS element. We believe that this gene (*SPS1-UGA*) is translated by a read-through mechanism that does not involve insertion of Sec or Cys, though we still do not know which amino acid is used at this site. The *SPS1-UGA* gene can be observed in the genomes of all extant *Hymenoptera*, which cannot synthesize Sec, as is also found in some paraneopterans, such as *Pediculus humanus* and *Rhodnius prolixus*. These latter organisms instead encode *SPS2* and other selenoproteins. After its genesis, the UGA codon in *SPS1-UGA* was then converted to an arginine codon, yielding *SPS1-Arg*. This event occurred independently in at least two lineages in *A. pisum* and in the entire order of *Diptera* that includes *D. melanogaster*.

In terms of *SPS* presence, we can summarize metazoan genomes as either: (1) carrying one *SPS2* gene with Sec UGA and SECIS, which is presumably

dual-functional; (2) carrying two distinct *SPS2* and *SPS1* genes following duplication or two transcript isoforms (e.g., see *Ciona*); and (3) carrying only a *SPS1* gene following duplication and then *SPS2* gene loss (e.g., see insects).

Only the nematode lineage escapes all these categories. In this phylum, most organisms have a single *SPS* gene with Cys. This gene appears to be the direct descendant of metazoan *SPS2* that underwent a Sec to Cys conversion. However, nematodes, similar to insects, went through a progressive selenoproteome depletion, with certain parasitic nematodes enduring a complete Sec loss [34]. In selenoprotein-less nematodes, *SPS2* was also lost, making them the only known metazoans devoid of any *SPS* proteins. In the sub-functionalization scenario, the nematode *SPS2* is expected to have lost the *SPS1* function at some point.

The evolution of *SPS* genes may provide additional hints about how the *SPS1* function came into existence. We observed that *SPS1* proteins perform a novel function, distinct but derived from the original *SPS2* function. The most relevant difference between *SPS2* and *SPS1* is the replacement of the Sec/Cys site by some other amino acid. In *SPS2*, Sec is inserted in response to a UGA codon through a recoding mechanism. If such a mechanism is not fully specific to Sec, and random amino acids are sometimes inserted instead, then a UGA containing *SPS2* gene would produce a *SPS1*-like protein. We think that such a non-Sec recoding mechanism may have been the key to the origin of *SPS1* function, and then to its selected maintenance in the dual-function ancestral metazoan *SPS2*. Indeed, *SPS2* genes contain conserved stem structures overlapping the Sec-UGA, previously reported in several selenoprotein genes and named SRE (Sec redefinition elements). The SRE of human selenoprotein N was shown to promote a non-Sec read-through activity in the absence of a SECIS element downstream and of its whole genomic context [35, 36]. Plausibly, the SRE in *SPS2* genes is expected to possess a similar activity. We believe that in the ancestral metazoa, SRE in *SPS2* was selected to support a minor non-Sec read-through activity, while at the same time maintaining the production of its main Sec-containing isoform. The dual-function state of this gene may be explained by the presence of these two protein isoforms; i.e., the Sec isoform carrying the canonical *SPS2* selenophosphate synthesis activity, and the non-Sec isoform carrying the *SPS1* function. In support of this view, the insect *SPS1-UGA* genes all contain stable structures similar to SRE elements, which we named hymenopteran read-through element (HRE). We think that HRE derived from SRE elements, which specialized in some yet uncharacterized form of non-Sec UGA recoding upon gene duplication and sub-functionalization.

Today, the molecular function of *SPS1* genes remains unknown. Structural comparisons suggest that *SPS1* proteins catalyze a phosphorylation reaction similar to *SPS2*, but likely acting on a substrate different from selenide. Various indirect evidences have linked *SPS1* to diverse biological pathways. Human *SPS1* was reported to interact with Sec synthase [37] and has been proposed to function in Sec recycling due to rescue experiments in *E. coli* growing on L-Sec [38]. In fruit flies, *SPS1* knockout mutants were reported to have defects in selenoprotein expression [39]. However, conservation of *SPS1* function in selenoprotein-less insects suggests that it is unrelated to Sec synthesis [29]. Based on differential gene expression analyses

performed upon SPS1 knockdown, *Drosophila* SPS1 has been proposed to be involved in vitamin B6 metabolism [40]. Previously, a role in redox homeostasis was proposed, based on the observation that heterozygous fruit flies with a compromised SPS1 copy are more sensitive than wild type flies to oxidative stress [41]. Hopefully, the enzymatic function of SPS1 proteins will be resolved in the future, clarifying the many aspects of their evolution that today remain elusive and necessarily subject to speculation.

8.4 Concluding Remarks

Selenophosphate synthetases (SelD/SPS) are required for the synthesis of Sec, and thus are present in every organism utilizing this amino acid. SPS are peculiar in that, besides participating in the Sec pathway, they are themselves often selenoproteins. The study of *SPS* genes allowed delineating the Se utilization traits across genomes. Meanwhile, it also uncovered classes of SPS proteins that apparently evolved novel functions, distinct from selenophosphate synthesis and likely unrelated to Sec synthesis, yet still to be defined (i.e., SPS1 in metazoans, SelD-like in *Crenarchaeota*). We can see the SPS family as a case in point of how comparative genomics can help in unraveling the complex nature of genes and their functions. Yet, this often results in further questions. What is the molecular and biological function of SPS1? What residue is used at the Sec homologous site of *SPS1-UGA* in Hymenoptera, and why is it inserted by a read-through stop codon? Are there additional means of natural Se utilization supported by SPS in prokaryotes and eukaryotes? What is the role of gene fusions in bacterial and protist *SPS* genes? What is the function of archaeal SelD-like proteins? These are just few of the questions, which we hope to see answered in the upcoming years of SPS research.

Acknowledgements The research leading to these results has received funding from the EMBO fellowship ASTF 289-2014, Spanish Ministry of Economy and Competitiveness under grant number BIO2011-26205 and Centro de Excelencia Severo Ochoa 2013–2017, SEV-2012-0208.

References

1. VM Labunsky et al 2014 *Physiol Rev* 94:739
2. A Krol 2002 *Biochimie* 84:765
3. D Yoshizawa, A Böck 2009 *Biochim Biophys Acta* 1790:1404
4. T Stock, M Rother 2009 *Biochimica Biophys Acta* 1790:1520
5. C Allmang et al 2009 *Biochim Biophys Acta* 1790:1415
6. Z Veres et al 1994 *J Biol Chem* 269:10597
7. M Mariotti et al 2015 *Genome Res* 25:1256
8. E Matsumoto 2008 *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64:453
9. KT Wang et al 2009 *J Mol Biol* 390:747
10. Y Itoh et al 2009 *J Mol Biol* 385:1456

11. N Noinaj et al 2012 *J Bacteriol* 194:499
12. MJ Guimarães et al 1996 *Proc Natl Acad Sci U S A* 93:15086
13. J Lin et al 2015 *Genome Biol Evol* 7:664
14. H Romero et al 2005 *Genome Biol* 6:R66
15. Y Zhang 2006 *Genome Biol* 7:R94
16. Y Zhang et al 2008 *BMC Genomics* 9:251
17. GP Li et al 2014 *BMC Genomics* 15:908
18. WM Ching et al 1985 *Proc Natl Acad Sci U S A* 82:347
19. DH Haft, WT Self 2008 *Biol Direct* 3:4
20. M Srivastava 2011 *J Bacteriol* 193:1643
21. D Su et al 2012 *FEBS Lett* 586:717
22. AV Lobanov et al 2009 *Biochim Biophys Acta* 1790:1424
23. L Jiang et al 2012 *BMC Genomics* 13:446
24. MT da Silva et al 2013 *Mol Biochem Parasitol* 188:87
25. M Hirosawa-Takamori et al 2000 *EMBO Rep* 1:441
26. XM Xu et al 2007 *Biochem J* 404:115
27. XM Xu et al 2007 *PLoS Biol* 5:e4
28. BC Persson et al 1997 *J Mol Biol* 274:174
29. AV Lobanov et al 2008 *Protein Sci* 17:176
30. M Mariotti et al 2012 *PLoS One* 7:e33066
31. CE Chapple, R Guigó 2008 *PLoS One* 3:e2968
32. M Mariotti 2016 in *Short Views on Insect Genomics and Proteomics* (Springer International Publishing) p 113
33. S Richards et al 2010 *PLoS Biol* 8:e1000313
34. L Otero et al 2014 *RNA* 20:1023
35. MT Howard et al 2005 *EMBO J* 24:1596
36. MT Howard et al 2007 *RNA* 13:912
37. A Small-Howard et al 2006 *Mol Cell Biol* 26:2337
38. T Tamura et al 2004 *Proc Natl Acad Sci U S A* 101:16162
39. B Alsina et al 1999 *J Cell Sci* 112:2875
40. KH Lee et al 2011 *BMC Genomics* 12:426
41. M Morey et al 2003 *FEBS Lett* 534:111