Grimaldi Alexis

Master 2 BMC
From Molecular Developmental Biology To Biomedicine,
Evolution and Systems Biology

# Searching for the link between bacterial, archaeal and eukaryotic selenoproteomes

Abstract:

Selenoproteins are present in the three domains of life. Selenium is incorporated in these proteins as selenocysteine, which is considered the 21st amino-acid. Yet, many species, among bacteria, archaea and eukaryote seem to lack selenoproteins. In order to synthesize selenoproteins, bacteria, archaea and eukaryotes use a rather conserved machinery. Selenocysteine is encoded by the UGA codon, normally required for the termination of the translation. The production and incorporation of selenocysteine in selenoproteins in made in a co-translational manner, thanks to trans- and cys-acting elements. The main cys-acting element is the SElenoCysteine Insertion Sequence (SECIS), present downstream of the UGA codon in bacteria and in the 3'UTR in archaea and eukaryotes. The key enzymes involved in this process are a selenocysteine synthase, a specific elongation factor, a specific tRNA, and a selenophosphate synthetase. Other factor are required in archaea and eukaryotes such as a SECIS binding protein.

In this project, we propose to investigate with bioinformatic methods the characteristic features of the selenoproteomes and the selenium processing machinery of numerous species in the three domains of life. By analyzing bacterial, archaeal, and eukaryal specific traits, we hope to find remains of bacterial selenoproteome and machinery in archaea, as well as bacterial and archaeal selenoproteome and machinery in protists.

# Table of Contents

# 1. Status of the project

## *1.1 General overview of the biological context of selenoproteins*

Selenium has been considered for a long time as a toxic element for organisms. Selenium is incorporated in particular proteins called selenoproteins, in the form of selenocysteine ( Sec, U ).

### *1.1.1 Selenium incorporation*

As its name suggests Sec is basically a cysteine ( Cys ) in which sulphur is replaced by selenium. This confers to selenocysteine a higher reactivity ( $pKa_{Sec}$ = 5.2 ; $pKa_{Cys}$ = 8.3 ) [1]. Selenocysteine is incorporated in a co-translational manner, at the position of specific UGA codons [2], which normally encodes for stop. The insertion of Sec at a specific UGA location is achieved by a complex machinery. Although this machinery varies among distant phyla (most notably between prokaryotes and eukaryotes), the basics remain the same.

Mono-selenophosphate (SePh), the donor compound of selenium (Se) necessary for the synthesis of Sec, is formed by the selenophosphate synthetase (SPS), from either selenite (SeO3) or Se2- [3]. It has been shown that Sec is synthesized from serine in a multi-step reaction: Ser-tRNA[Sec] is phosphorylated by a phosphoseryl tRNA Kinase (PSTK) and converted to Sec-tRNA[Sec] by Sec synthetase (SecS) [4]. In eukaryotes, another partner seems to be involved in selenoprotein synthesis: Secp43. Although its exact role is still unclear, it has been shown that it binds to the Ser-tRNA[Sec]/SPS1/SecS complex [4,5]. Data suggests that Secp43 could be involved in promoting localization to the nucleus, or in regulating selenoproteins synthesis. [4]

### *1.1.2 Basic structure of selenoproteins*

The recognition between a Sec-UGA and a Stop-UGA is a key event in selenoprotein biosynthesis. Indeed, a stop codon should not be read-through by the machinery, and a Sec codon should not be interpreted as a stop codon. To achieve this kind of behavior, a particular structure called Sec insertion sequence ( SECIS ) is present in the RNA [6]. SECIS elements are basically stem-loop structures with either 1 or 2 loops depending of the type [6,7] (Fig 1) How exactly the SECIS element drives the reading of a Sec-UGA codon is still unknown, but several selenoprotein synthesis specific factors bind to it and / or to the translational machinery [5], making it a key element in decoding UGA codons.

## *1.2. Selenoproteomes*

Selenoproteins exist in all domains of life (prokaryotes, archaea and eukaryotes) However, it seems that higher plants and fungi have no selenoproteins. Yet, several selenoproteins have been found in the green algae *Chlamydomonas reinhardtii* [8]. Given the high nucleophilicity of Sec compared to Cys, many selenoproteins are in fact redox proteins, in which Sec is used to coordinate a redox-active metal (molybdenum, nickel or tungsten). Examples of selenoproteins are formate

dehydrogenase, glutathione peroxidase or thioredoxin reductase. Sec is then found almost exclusively in the catalytic sites of enzymes. It is interesting to note that a high number of selenoproteins have a homologue with Cys at the Sec position in other organisms [9], suggesting that not all organisms require selenium. The evolution of the Sec-encoding trait is subject to debate, but since this trait is present in organisms of the three domains of life, it is thought that the last universal common ancestor had this trait, which was subsequently lost in numerous lineages during evolution, and was regained by some lineages through gene transfer [10].

### *1.2.1 Bacterial Selenoproteome and Machinery*

Bacterial selenoproteins are mainly redox enzymes involved in energy-yielding pathways, since these processes heavily rely on oxido-reduction reactions [11,12]. Yet, selenoproteins are also found in other processes, such as antioxidant defense, redox homeostasis, or Sec biosynthesis [12]. (Table 1). It is worth pointing out that SelD homologues containing Sec have been identified in a few bacteria [12,13]. Though, when SelD is a selenoprotein, it has yet to be understood how Sec synthesis can be initiated. It seems a gram negative bacteria, *Heamophilus influenzae*, solves the problem by initially operating a Cys containing SelD, or by forming Sec-tRNA[Sec] in a selenophosphate independent manner [14].

Bacteria also incorporate Sec in a very specific manner. The SECIS element is found near the Sec-UGA codon, just downstream of it [12,15,16]. In this case the translation machinery is thought to pause at the site of the Sec-UGA in order to allow the other factors to insert the Sec. SelB in bacteria can bind directly to the SECIS and Sec-UGA [17], thus promoting elongation in order not to stop the translation (Fig. 2). Since in bacteria the SECIS element is found in the coding region, we can assume that it is the secondary structure of the SECIS element that matters, because it also has to keep its coding capacity.

### *1.2.2 Archaeal Selenoproteome and Machinery*

In archaea, selenoproteins seem to be a lot less common. It is quite striking that selenoproteins seem restricted to two genera ( based on the data gathered from 56 sequenced genomes, representing 43 genera ) [11], *Methanococcus* and *Methanopyrus*. As their name suggests, these are methanogenic organisms. It is also intriguing that the selenoproteins of these species, with the exception of SPS, are all involved in hydrogenotrophic methanogenesis (Table 2).

Strikingly, archaeal selenium metabolism seems to share both bacterial and eukaryal traits. A putative SPS has been highlighted, but has not been experimentally checked yet. Archaeal SECIS elements are not found downstream of the UGA in the coding sequence, but like in eukaryotes, in the 3'UTR [18]. Though, one case of SECIS like structure has be found in 5'UTR, which would be a unique feature only observed in this phylum [11,19]. Furthermore, as it will be developed in the next paragraph, it seems that archaeal SECIS elements do not possess non-Watson-Crick base pair associations as it is the case in eukaryotes [6]. Also, it is not known yet if Efsec binds directly to

the SECIS element as in bacteria, or if it does so indirectly as in eukaryotes [11] (Fig 3).

### *1.2.3 Eukaryal Selenoproteome and Machinery*

As in archaea, the SECIS element is found in the 3'UTR of the RNA. Being in an untranslated region, it is not under encoding constraints. Yet, it has been shown that factors interacting with the SECIS recognize non-Watson-Crick base pair associations in the stem [20]. The apparent importance of this primary structure actually drives the secondary structure (Fig 1). Indeed, thanks to those non-canonical base pair associations, the stem of the SECIS adopts a particular shape essential for the binding of specific factors [5]. The fact that the SECIS is not located near the Sec-UGA suggests a very complex and tightly regulated mechanism to insert Sec at the right position. In this situation, EFsec cannot directly bind to both the SECIS element and the RNA at the location of the Sec-UGA. Instead, it is recruited by SECIS binding protein 2 ( SBP2 ) (Fig 4), which binds to both the SECIS and the ribosome [5]. How exactly the SECIS/SBP2 complex recognizes a Sec-UGA is not fully understood. However, speculations can be made about the local context around the Sec-UGA, with some pattern driving the SECIS/SBP2 localization at the site of a Sec-UGA. We can also think that SBP2 binds the SECIS and the ribosome from the very beginning of the translation, recruits EFsec, thus preventing the ribosome from being released when a UGA is encountered. But then remains the question of how and under what conditions SBP2 would release EFsec, leading to ribosome release at the next UGA codon. Other data suggest that a viral RNA encoding for GPx4 possesses a SECIS element downstream of the UGA, in the coding sequence, and that it could be expressed in mammalian cells [5,21]. This raises the questions of how much specific the machinery of eukaryotes is, and what is the link between such apparently ancestral mechanism (until now specific to bacteria) and more complex ones.

It is also worth noting that Sec synthesis in eukaryotes relies on SPS2, which is itself a selenoprotein [22]. Although it has the same role as SelD (SPS) in bacteria, an homologue called SPS1 is present, containing either a threonine or an arginine . SPS1 does not seem to play a role in selenoprotein synthesis, though, but is essential for the viability of the organisms [5,23], and has been shown to be implicated in developmental processes in *drosophila* [24,25].

The size of selenoproteomes varies greatly among eukaryotes [26]. It seems that the largest ones are those of aquatic organisms, whether they are plants ( algae ) or animals [27]. It has been hypothesized that blue and green algae need an additional antioxidant protection provided by the selenoproteins. Furthermore, the high availability of Se in sea water could have made possible the integration of this element in oxidoreductase enzymes. An example of very scattered distribution among species is SelU. It is present in fishes, birds, and some unicellular organisms, but is present as a Cys-containing homologue in mammals and other eukaryotes [9]. A recently characterized family, SelJ has also such a scattered distribution among eukaryotes [28].

Some selenoproteins seem to be implicated in the development. Indeed, the antioxidant function of selenoproteins could play an important role in keeping the mice developing tissues ( mainly the

muscles and the nervous system ) devoid of free radicals [29]. The importance of keeping the low level of reactive oxygen in the developing embryo is also pointed out by the fact that early stages of mice develop in an almost oxygen free environment. It has been suggested that SelW, a selenoprotein of unknown function, could play such a role in the developing brain, even in post-natal stages [29]. SelN, another selenoprotein whose function remains elusive, has been shown to be expressed in mice embryos, but not in adults. Interestingly, in this case, SelN transcripts were still present in adults, but no actual protein, suggesting subtle regulation of the translation of selenoproteins [30]. Another selenoprotein, SelJ, could play a role in the development of the eye [28].

### *1.4. Evolution of selenoproteomes*

Selenoproteomes and Se processing machinery are often compared between more or less distant species, and can explain for example the divergence of SPS1 and SPS2 in eukaryotes. However, such comparisons have not been made between ancestral groups. It has yet to be elucidated how this machinery and the subsequent selenoproteins have evolved, from the Last Universal Common Ancestor to the higher eukaryotes.

## 2. Project Description

### *2.1. Aim of the project*

When having a close look at selenoproteomes in different species, be it in prokaryotes, archaea or eukaryotes, they seem to be quite heterogeneous in size as well as in quality [11,12,26]. Strong divergence in selenoproteomes can explain these data. However, one cannot rule-out the hypothesis that genome misannotation and search criteria might lead to missing some selenoproteins.

The selenoproteome is not the only data that shows both links and differences between different species. All the selenoprotein synthesis machinery is also both very conserved and divergent at the same time. Core factors and structures are present in the three domains of life, but less restrictive mechanisms may diverge a lot.

Understanding how selenoproteins are related to their homologues in different species, and to what extent the underlying bio-synthesis mechanisms are conserved during evolution might enrich our knowledge about selenoproteins in general, enabling us to pinpoint their implication in various processes ( development, endocrine regulations … ) as well as finding new selenoprotein candidates. Furthermore, hints about the role of Sec free homologues could be elucidated (i.e. the role of SPS1). Current gaps in phylogenetic trees might be filled after this project is completed.

In this project, we hope to outline selenoproteome and selenium processing machinery evolution from bacteria to eukaryotes.

This project will thus be mainly focused on two aspects:

### *2.1.1 Creating profiles of bacteria and archaea.*

The creation of these profiles will be the starting point of the project, since they will be used to perform the other tasks of this project.

Selenoprofiles (unpublished), a recently developed pipeline to automate the search of selenoproteins, uses profiles, witch are multiple alignments of protein families, as an input. This allows to capture more precisely the intrinsic variation within a protein family. The extensive use of this tool will enable us to identify all members of the families that are present in a genome.

Such profiles have already been made for eukaryal selenoprotein families, but remain to be done in bacteria and archaea

### *2.1.2 Looking for a link between selenium processing machinery and selenoproteins in bacteria, archaea and eukaryotes*

Looking for a link between i. the selenoproteome ii. the selenium processing machinery in bacteria, archaea and eukaryotes will be made by several approaches.

For both proteins implicated in selenoproteins bio-synthesis, and selenoproteins, we will search for bacterial homologues in archaea and eukaryotes, as well as archaeal homologues in bacteria and eukaryotes. A higher variability can be observed in the protists phylum than in the other eukaryotes. This is why protists will be the eukaryotes we will focus on.

We will also look for bacterial selenoprotein specific traits in archaea and protists and archaeal selenoprotein specific traits in bacteria and eukaryotes. Such key traits will be the structure of SECIS elements and their position along the RNA and the presence of factors that, until now, seemed specific to subsets of the three domains of life (For instance the use of SBP2 to make the link between Efsec and the SECIS element).

## *2.2. Tools used*

In order to conduct this project, several bioinformatic tools will be used. Some are databases which will mainly be used as a starting point for the further comparison of the selenoproteomes and the selenium processing machinery. Computational tools will then be used to actually perform the comparison. Finally, existing pipelines will be used and hopefully improved, as well as new ones will be designed, in order to automatize as many steps as possible.

### *2.2.1. Pfam & COG databases*

As a starting point, selenoprotein families have to be gathered from all three domains of life in order to build profiles. Basically, these profiles are multiple alignments of one family of proteins. The found homologues can thus be in the same organism or in others. The selenoproteins families will be gathered from [31].

To build these profiles two databases of protein families will be used, using data from . The first is

the Pfam database, a large collection of protein families, each represented by a multiple sequence alignment and a Hidden Markov Model (HMM) [32]. The second is the COG (Clusters of Orthologous Groups of proteins) database [33]. The clusters were delineated by comparing protein sequences encoded in complete genomes, representing major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogs from at least 3 lineages and thus corresponds to an ancient conserved domain.

### *2.4.2. BLAST*

Nowadays, blast (Basic Local Alignment and Search Tool) is a very common tool for quickly finding homologies between biological sequences [34]. Yet, several options of this tool makes it suitable to perform other kind of tasks. This project will make an intensive use of psiblast (Position Specific Iterated Blast) which is an extension of the regular blast. Instead of using a a standard scoring matrix such as BLOSUM or PAM, psiblast uses a PSSM (Position Specific Scoring Matrix) when performing the alignments. Since the PSSM is obtained from a previous alignment, psiblast can use its own alignments to compute new PSSM and re-perform the alignment in an iterative process which enable an increased sensitivity as recursion depth grows.

### *2.2.3. Alignment tools*

Third party alignment tools will be used. T_coffee [35], for instance, is a multiple alignment package which will allow us to trim the alignments during the profiles building stage.

### *2.2.4. SECISearch*

SECISearch is a tool for searching SECIS elements in a given sequence [36]. Specific patterns are searched in a target sequence. SECISearch will then look at the secondary structure adopted by the candidates. A score based on the stability of the predicted secondary structure is attributed, filtering steps, outputs the candidates likely to be SECIS.

### *2.2.3. Selenoprofiles*

Selenoprofiles is a pipeline written in python providing an efficient way to predict selenoproteins by performing a homology based search against whole genomes (unpublished). It makes extensive use of psitblastn to align the searched sequence , exonerate [37] to extend the hits, genewise [38] to check the extensions or perform a comparison if exonerate didn't work, and SECISearch to finally check the presence of SECIS elements.

### *2.2.6. SelenoDB*

SelenoDB is a joint effort between experimental and computational laboratories to create, maintain and update a database of eukaryal selenoprotein genes, proteins, SECIS elements and related molecules [39]. Although the predictions are manually curated, they are initially obtained by gene comparison and computational RNA secondary structure predictions.

## 3. Expected results

From previous analyses on the protists, we expect a very diversified panorama in terms of selenoproteins and selenium processing machinery. We expect to find species in which the complete eukaryal machinery system is not present, or is very divergent. We hope to find species in which a system similar to the archaeal one is used. We also expect to find transition states of the position of the SECIS from the coding region to the UTR.

Comparing selenoproteins, we expect to identify some that are conserved (and identifiable) from bacteria to eukaryotes, while most of them are specific to some lineages (intra-kingdom).

We also expect that in archaea, the role of SBP2 is carried out by some other, yet, unknown protein. Although this may be difficult, we may be able to come up with candidates for this role.

This work will contribute to shed some light on the dynamics of evolution of the selenoprotein trait, with focus on the transition from archaea to eukarya.

## 4. Figures and Tables

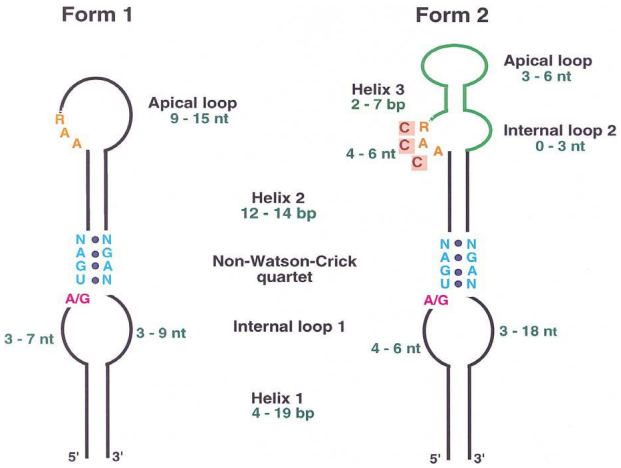


*Fig 1:* ***The two types of SECIS elements found in eukaryotes****. The Non-Watson-Crick quartet gives the particular secondary structure to the SECIS elements. Adapted from [6]*

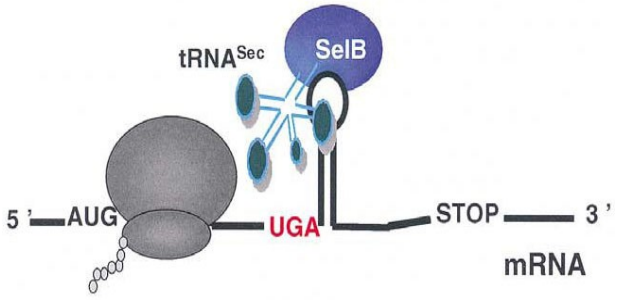


*Fig 2:* ***Selenocysteine insertion during bacterial translation****. SelB, the elongation factor, binds to the SECIS element and recruits tRNASec at the position of the UGA. Adapted from [6]*

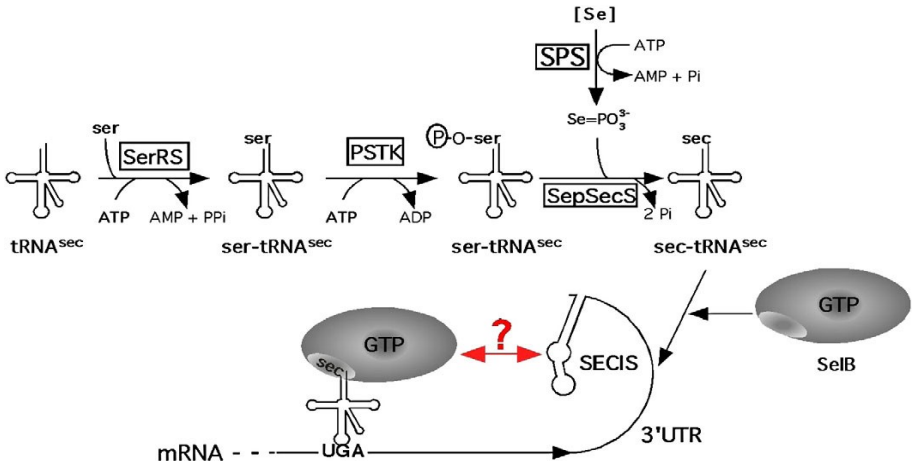


*Fig 3:* ***Selenium processing and Sec insertion in archaea****. The SECIS element is in the 3'UTR, but no factor has been found to make the link between SelB and and the SECIS element. Adapted from [11]*

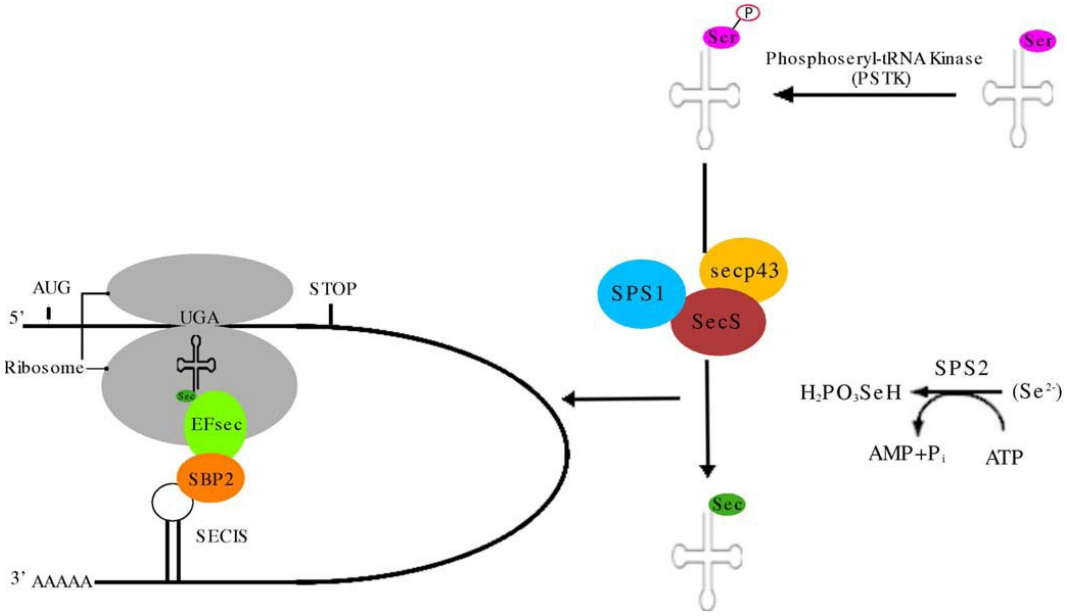


*Fig 4:* ***Selenocystein incorporation in eukaryotes****. Two selenophosphate synthetases (SPS1 and SPS2) are present. EFsec is recruited to the SECIS element by SBP2. Adapted from [11]*

| Protein | Characteristic organism | Verified? |
|---|---|---|
| Sec-containing | | |
| Formate dehydrogenase | *M. thermoacetica* | Yes |
| Glycine reductase $P_A$ | *C. sticklandii* | Yes |
| Glycine reductase $P_B$ | *E. acidaminophilum* | Yes |
| Sarcosine reductase $P_B$ | *E. acidaminophilum* | Yes |
| Betaine reductase $P_B$ | *E. acidaminophilum* | Yes |
| Proline reductase $P_B$ | *Clostridium sticklandii* | Yes |
| SPS | *E. acidaminophilum* | Yes |
| Peroxiredoxin | *E. acidaminophilum* | Yes |
| Methionine sulfoxide reductase | *Clostridium* OhILAs | Yes |
| not Sec-containing | | |
| Xanthine dehydrogenase | *C. acidiurici* | Yes |
| Purine hydroxylase | *C. purinolyticum* | Yes |
| Nicotinic acid dehydrogenase | *C. barkeri* | Yes |

*Table 1: **Non exhaustive bacterial selenoproteome**. Many selenoproteins have a role in oxido-reduction reactions. Adapted from [11]*

| Selenoprotein | Subunit | Characteristic organism | Verified? |
|---|---|---|---|
| Formate dehydrogenase | FdhA | *Methanococcus vannielii* | Yes |
| Formyl-methanofuran dehydrogenase | FwuB | Methanopyrus kandleri | Yes |
| $F_{420}$-reducing hydrogenase | FruA | *Methanococcus voltae* | Yes |
| $F_{420}$-non-reducing hydrogenase | VhuD VhuU | *Methanococcus voltae* | Yes |
| Heterodisulfide reductase | HdrA | *Methanococcus jannaschii* | Yes |
| Selenophosphate synthetase | homomeric | *Methanococcus jannaschii* | No |
| HesB-like protein | unknown | *Methanococcus jannaschii* | No |

*Table 2: **Non exhaustive archaeal selenoproteome**. The majority of selenoproteins are involved in methanogenesis. Adapted from [11]*

# References


[1] D.E. Fomenko, W. Xing, B.M. Adair, D.J. Thomas, and V.N. Gladyshev, "High-throughput identification of catalytic redox-active cysteine residues," *Science (New York, N.Y.)*, vol. 315, Jan. 2007, pp. 387-389.

[2] T.C. Stadtman, "Selenocysteine," *Annual Review of Biochemistry*, vol. 65, 1996, pp. 83-100.

[3] C. Allmang and A. Krol, "Selenoprotein synthesis: UGA does not end the story," *Biochimie*, vol. 88, Nov. 2006, pp. 1561-1571.

[4] J.E. Squires and M.J. Berry, "Eukaryotic selenoprotein synthesis: mechanistic insight incorporating new factors and new functions for old factors," *IUBMB Life*, vol. 60, Apr. 2008, pp. 232-235.

[5] C. Allmang, L. Wurth, and A. Krol, "The selenium to selenoprotein pathway in eukaryotes: More molecular partners than anticipated," *Biochimica Et Biophysica Acta*, vol. 1790, Nov. 2009, pp. 1415-1423.

[6] A. Krol, "Evolutionarily different RNA motifs and RNA-protein complexes to achieve selenoprotein synthesis," *Biochimie*, vol. 84, Aug. 2002, pp. 765-774.

[7] C.E. Chapple, R. Guigó, and A. Krol, "SECISaln, a web-based tool for the creation of structure-based alignments of eukaryotic SECIS elements," *Bioinformatics (Oxford, England)*, vol. 25, Mar. 2009, pp. 674-675.

[8] S.V. Novoselov, M. Rao, N.V. Onoshko, H. Zhi, G.V. Kryukov, Y. Xiang, D.P. Weeks, D.L. Hatfield, and V.N. Gladyshev, "Selenoproteins and selenocysteine insertion system in the model plant cell system, Chlamydomonas reinhardtii," *The EMBO Journal*, vol. 21, Jul. 2002, pp. 3681-3693.

[9] S. Castellano, S.V. Novoselov, G.V. Kryukov, A. Lescure, E. Blanco, A. Krol, V.N. Gladyshev, and R. Guigó, "Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution," *EMBO Reports*, vol. 5, Jan. 2004, pp. 71-77.

[10] Y. Zhang, H. Romero, G. Salinas, and V.N. Gladyshev, “Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues,” *Genome Biology*, vol. 7, 2006, p. R94.
[11] T. Stock and M. Rother, “Selenoproteins in Archaea and Gram-positive bacteria,” *Biochimica Et Biophysica Acta*, vol. 1790, Nov. 2009, pp. 1520-1532.
[12] G.V. Kryukov and V.N. Gladyshev, “The prokaryotic selenoproteome,” *EMBO Reports*, vol. 5, May. 2004, pp. 538-543.
[13] T. Gursinsky, J. Jäger, J.R. Andreesen, and B. Söhling, “A selDABC cluster for selenocysteine incorporation in Eubacterium acidaminophilum,” *Archives of Microbiology*, vol. 174, Sep. 2000, pp. 200-212.
[14] R. Wilting, K. Vamvakidou, and A. Böck, “Functional expression in Escherichia coli of the Haemophilus influenzae gene coding for selenocysteine-containing selenophosphate synthetase,” *Archives of Microbiology*, vol. 169, Jan. 1998, pp. 71-75.
[15] F. Zinoni, J. Heider, and A. Böck, “Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, Jun. 1990, pp. 4660-4664.
[16] Z. Liu, M. Reches, I. Groisman, and H. Engelberg-Kulka, “The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in Escherichia coli,” *Nucleic Acids Research*, vol. 26, Feb. 1998, pp. 896-902.
[17] M. Thanbichler and A. Böck, “Functional analysis of prokaryotic SELB proteins,” *BioFactors (Oxford, England)*, vol. 14, 2001, pp. 53-59.
[18] M. Rother, A. Resch, R. Wilting, and A. Böck, “Selenoprotein synthesis in archaea,” *BioFactors (Oxford, England)*, vol. 14, 2001, pp. 75-83.
[19] R. Wilting, S. Schorling, B.C. Persson, and A. Böck, “Selenoprotein synthesis in archaea: identification of an mRNA element of Methanococcus jannaschii probably directing selenocysteine insertion,” *Journal of Molecular Biology*, vol. 266, Mar. 1997, pp. 637-641.
[20] R. Walczak, P. Carbon, and A. Krol, “An essential non-Watson-Crick base pair motif in 3'UTR to mediate selenoprotein translation,” *RNA (New York, N.Y.)*, vol. 4, Jan. 1998, pp. 74-84.
[21] H. Mix, A.V. Lobanov, and V.N. Gladyshev, “SECIS elements in the coding regions of selenoprotein transcripts are functional in higher eukaryotes,” *Nucleic Acids Research*, vol. 35, 2007, pp. 414-423.
[22] M.J. Guimarães, D. Peterson, A. Vicari, B.G. Cocks, N.G. Copeland, D.J. Gilbert, N.A. Jenkins, D.A. Ferrick, R.A. Kastelein, J.F. Bazan, and A. Zlotnik, “Identification of a novel selD homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism?,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, Dec. 1996, pp. 15086-15091.
[23] C.E. Chapple and R. Guigó, “Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes,” *PloS One*, vol. 3, 2008, p. e2968.
[24] B. Alsina, F. Serras, J. Baguñá, and M. Corominas, “patufet, the gene encoding the Drosophila melanogaster homologue of selenophosphate synthetase, is involved in imaginal disc morphogenesis,” *Molecular & General Genetics: MGG*, vol. 257, Jan. 1998, pp. 113-123.
[25] M. Morey, F. Serras, J. Baguñà, E. Hafen, and M. Corominas, “Modulation of the Ras/MAPK signalling pathway by the redox function of selenoproteins in Drosophila melanogaster,” *Developmental Biology*, vol. 238, Oct. 2001, pp. 145-156.
[26] A.V. Lobanov, D.L. Hatfield, and V.N. Gladyshev, “Eukaryotic selenoproteins and selenoproteomes,” *Biochimica Et Biophysica Acta*, vol. 1790, Nov. 2009, pp. 1424-1428.
[27] A.V. Lobanov, D.E. Fomenko, Y. Zhang, A. Sengupta, D.L. Hatfield, and V.N. Gladyshev, “Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life,” *Genome Biology*, vol. 8, 2007, p. R198.
[28] S. Castellano, A.V. Lobanov, C. Chapple, S.V. Novoselov, M. Albrecht, D. Hua, A. Lescure, T.

Lengauer, A. Krol, V.N. Gladyshev, and R. Guigó, “Diversity and functional plasticity of eukaryotic selenoproteins: identification and characterization of the SelJ family,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, Nov. 2005, pp. 16188-16193.

[29] Y.W. Chung, D. Jeong, O.J. Noh, Y.H. Park, S.I. Kang, M.G. Lee, T. Lee, M.B. Yim, and I.Y. Kim, “Antioxidative role of selenoprotein W in oxidant-induced mouse embryonic neuronal cell death,” *Molecules and Cells*, vol. 27, May. 2009, pp. 609-613.

[30] P. Castets, S. Maugenre, C. Gartioux, M. Rederstorff, A. Krol, A. Lescure, S. Tajbakhsh, V. Allamand, and P. Guicheney, “Selenoprotein N is dynamically expressed during mouse development and detected early in muscle precursors,” *BMC Developmental Biology*, vol. 9, 2009, p. 46.

[31] Y. Zhang and V.N. Gladyshev, “Trends in selenium utilization in marine microbial world revealed through the analysis of the global ocean sampling (GOS) project,” *PLoS Genetics*, vol. 4, Jun. 2008, p. e1000095.

[32] R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L.L. Sonnhammer, and A. Bateman, “The Pfam protein families database,” *Nucleic Acids Research*, vol. 36, Jan. 2008, pp. D281-288.

[33] R.L. Tatusov, M.Y. Galperin, D.A. Natale, and E.V. Koonin, “The COG database: a tool for genome-scale analysis of protein functions and evolution,” *Nucleic Acids Research*, vol. 28, Jan. 2000, pp. 33-36.

[34] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, “Basic local alignment search tool,” *Journal of Molecular Biology*, vol. 215, Oct. 1990, pp. 403-410.

[35] C. Notredame, D.G. Higgins, and J. Heringa, “T-Coffee: A novel method for fast and accurate multiple sequence alignment,” *Journal of Molecular Biology*, vol. 302, Sep. 2000, pp. 205-217.

[36] G.V. Kryukov, S. Castellano, S.V. Novoselov, A.V. Lobanov, O. Zehtab, R. Guigó, and V.N. Gladyshev, “Characterization of mammalian selenoproteomes,” *Science (New York, N.Y.)*, vol. 300, May. 2003, pp. 1439-1443.

[37] G.S.C. Slater and E. Birney, “Automated generation of heuristics for biological sequence comparison,” *BMC Bioinformatics*, vol. 6, 2005, p. 31.

[38] E. Birney and R. Durbin, “Using GeneWise in the Drosophila annotation experiment,” *Genome Research*, vol. 10, Apr. 2000, pp. 547-548.

[39] S. Castellano, V.N. Gladyshev, R. Guigó, and M.J. Berry, “SelenoDB 1.0 : a database of selenoprotein genes, proteins and SECIS elements,” *Nucleic Acids Research*, vol. 36, Jan. 2008, pp. D332-338.