Genomic and biochemical characterization of a bacterial homolog of Selenoprotein N, a selenium-containing protein involved in congenital muscular dystrophy

Nedaa Yousef Mahboub

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ ED414

UPR 9002 - ARN du CNRS

THÈSE présentée par :

Nedaa Yousef MAHBOUB

Pour l'obtention du Grade de: **DOCTEUR DE L'UNIVERSITÉ DE STRASBOURG**

Discipline : Sciences de la Vie et de la Santé

Spécialité : Aspects moléculaires et cellulaires de la biologie

Caractérisation génomique et biochimique d'un homologue bactérien de la Sélénoprotéine N, une protéine à sélénium impliquée dans différentes dystrophies musculaires congénitales

Soutenue publiquement le 31 mars 2023
devant la commission d'examen composée de:

Rapportrices externes
Dr. Valérie ALLAMAND Sorbonne Université, Inserm, Centre de Recherche en Myologie
Dr. Isabelle LIHRMANN BISSON Inserm U1239, Université de Rouen

Examinatrice interne
Dr. Christine ALLMANG-CURA CNRS - UPR 9002 ARN - Strasbourg

Thèse dirigée par:
Dr. Alain LESCURE Chargé de recherche, CNRS - UPR 9002 ARN - Strasbourg

PhD financé par le Bureau Culturel d'Arabie Saoudite à Paris

Acknowledgments

First of all, I thank my God who protected me and gave me the strength to complete my Ph.D journey.
After that,, I would like to thank Dr. Alain Lescure for giving me the opportunity to join his lab and work on this mysterious and bewildering project,
and to thank the team working on Selenoprotein N at the IBMC institute.
Pascale Romby, the director of the CNRS Unit "Architecture and Reactivity of RNA", thank you for your respect and caring.
I would also like to thank all the people involved in the collaboration project of sponge sampling and experiments at Observatoire Océanologique laboratoire de Banyuls-sur-Mer, Sorbonne Université for making my integration and working environment pleasant. Thanks to Marcelino Suzuki, the director of the Microbial Biodiversity and Laboratory "LBBM" and the compliment he gave to me. Nyree West, thank you for being like a friend more than a supervisor. Marie-line Escand, thank you for the tour of the Chapelle Notre-Dame de la Salette.
I leave my special thanks to:
My friend ''more than a brother'' Maaz Bin Nasim, Ph.D. holder, Université de Strasbourg, faculté de Pharmacie, for encouraging and persuading me to continue my higher education. He was always there for me during the ups and downs in my Ph.D. life, and always found the time for scientific discussions and sharing experiences.
My friend Mohammed Al-Arjah, and Taif AL-Sarhan,, thank you for your patience and support from the beginning of my Ph.D. journey to the end.
My mother, warm thanks and gratitude, you had paved my entire life journey since I was born. Thank you for your love and support.
My sweetheart, loving nephew Omar, you are the light of my life, my moon and sun, one day when you grow up you will read my words and realize how much am lucky to have you in my life, thank you for loving me and supporting me with your innocent smile.
Mr. Suliman Al-Jaboury, thank you for your continued support and academic-life consultations.
Finally, my grateful appreciation to the generosity of the Saudi Arabian cultural mission in France (Bureau culturel Saoudien en France et Swiss) for the scholarship grant to my Ph.D. program. Thank you for believing in me. Dr. Jamal ALSHEIBANI and Mr. Said MAZLOUM thank you for the academic supervision.

شُكر وامتنان

في البداية أشكُرُ إلهي الذي هداني ورزقني ورعاني في كل خُطوة وأنا بعيدة عن الوطن.. وفتح علي من واسِعِ فضلِه ,,

بعد ذلِك،، الشُكر للدكتور آلان ليسكيور لقبولي في مشروع الدكتوراة و الانضمامِ إلى مختبره والعمل على هذا المشروع الغامض والمحير،

وأشكُرُ فريقَ العمل على بروتين السيلينون في مركز أبحاث ال IBMC .

باسكال رومبي، مديرة وحدة CNRS الهندسة التفاعلية للرايبو نيوكليك أسيد، شكرًا لك على احترامك واهتمامك الدائم.

أود أيضًا أن أشكر جميع الأشخاص المشاركين في المشروع التعاوني لأخذ عينات الإسفنج والتجارب في Observatoire Océanologique laboratoire de Banyuls-sur-Mer، جامعة السوربون، شكراً على بيئة العمل المُمْتِعة. شكراً لمارسيلينو سوزوكي ، مدير التنوع البيولوجي الميكروبي ومختبر "LBBM" والثناء الذي قدمه لي. نيري ويست، أشكرك على كونكِ صديقة أكثر من كونكِ مشرفًا. ماري لاين إسكاند، شكرًا لك على جولة شابيل نوتردام دي لا ساليت.

أترك شكري الخاص لـ:

أخي و صديقي معاذ بن نسيم، حاصل درجة الدكتوراه، جامعة ستراسبورغ، كلية الصيدلة، شكراً لتشجيعي واقِناعي لمواصلة الدراسات العليا. شكراً لوجودك وقت الشدة والدعم النفسي و المعنوي الذي قدمته لي، شكراً للمناقشات العلمية وتبادل الخبرات.

أصدقائي محمد العرجة و طيف السرحان أشكُركُم على صبرِكُم ودعمِكُم لي من بداية الابتعاث حتى النهاية.

أمي الغالية ، أقدم لكِ شُكري وامتناني، لقد مهدت رحلة حياتي كلها منذ ولادتي. شكراً لكِ على حُبكِ ودعمكِ.

حبيبي عُمَر ، أنت نور حياتي ، قمري وشمسي ، يومًا ما عندما تكبُر ستقرأ كلامي وستُدرِك كم أنا محظُوظة لِوجودِك في حياتي ، أشكرُ لكَ صبْرك على سفرِي، شكراً على حُبِك و ابتِسامتك البريئة .

الأستاذ سُليمان الجبوري أشكُر لكم دعمكم الذي لم ينقطع رُغم الصُعُوبات الأكاديمية التي واجهتها.

أخيرًا ، أعرِبُ عن شُكري و امتناني لوزارة التعليم العالي، في المملكة العربية السعودية، على المنحة الدراسية من برنامج خادم الحرمين الشريفين و تمويل رسالة الدكتوراة، ومُمثِلتُها الملحقية الثقافية السعودية في سويسرا و فرنسا، ولا أنسى الدكتور جمال الشيباني و الأستاذ سعيد مظلوم على المتابعة الأكاديمية،، ونيابةً عن المبتعثين السعوديين في فرنسا أشكُر لكم دعمكُم، كُنتم لنا نِعم العوْن ودُمتم لنا ذخراً .

نداء يوسف أحمد محبوب

Abbreviations

2YT	Yeast Extract Tryptone growth medium
Amp	Ampicilline
APS	Ammonium peroxodisulfate
ATP	Adenosine Tri Phosphate
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
CHAPS	3-[(3-Cholamidopropyl)-Diméthylammonio]-1-Propane Sulfonate
Cys	Cysteine
DMSO	Dimetyhlsulfoxide
DNA	Deoxyribose nucleic acid
EDTA	Ethylene diamine tetra acetic
ER	Endoplasmic reticulum
h	hour
Ig	Immunoglobulin
IPTG	Isopropyl ß-D-Thiogalactopyranoside
kDa	Kilodalton
LB	Luria-Bertani broth
MES	2-(N-morpholino) ethanesulfonic acid
MOPS	3-(N-morpholino) propanesulfonic acid
NI-IDA	Imino-diacetic acid
OD	Optical dansity
PBS	Phosphate buffer solution
PDB	Protein data bank
RF1	Release factor 1
RyR	Ryanodine receptor
Se	Selenium
Sec	Selenocysteine
SECIS	selenocysteine insertion sequence
SELA	Sec synthase enzyme
SELB	A specialized dedicated elongation factor
SELC	Sec unique tRNA ($tRNA^{[Ser]Sec}$)
SELD	Selenophosphate, a selenium donor
Ser	Serine
SM	Streptomycine
TBE	Tris-borate EDTA buffer
TGS	Tris-glycerine-SDS buffer
TM	Transmembrane

Abbreviations

bSelenoN	Bacterial selenoN
CARD-FISH	Catalyzed reporter deposition-Fluorescence *in situ* hybridization
DNA	Deoxyribose nucleic acid
EDTA	Ethylene glycol tetraacetic acid
EFsec	selenocysteine-specific elongation factor
EM	Electron Microscopy
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated protein degradation
FT	Flow-through
Gpx	Glutathione peroxidase
GSH	Glutathione
hSelenoN	Human selenoN
ORF	Open reading frame
PSTK	phosphoserine-tRNA kinase
SDS	sodium dodecyl sulfate
SE	Soluble extract
SecS	Selenocysteine synthase
SELENON	Selenoprotein gene
SelenoN	Selenoprotein N
SeP	Selenophosphate
SerRS	Seryl-tRNA synthetase
SPS2	Selenophosphate synthetase
SR	Sarcoplasmic reticulum
TE	Total extract (crude extract)
Trx	Thioredoxin
TrxR	Thioredoxin reductase
UTR	untranslated region

Table of contents

List of Figures

I. Figures of The Introduction Part

II. Figures of The Results Part.

III. Figures of The Materials & Methods Part.

List of Tables

Genomic and biochemical characterization of a bacterial homolog of Selenoprotein N, a selenium-containing protein involved in congenital muscular dystrophy

Overview of the thesis project

Selenoprotein N (SelenoN), is a member of the selenium-containing proteins family. The *SELENON/SEPN1* gene dysfunction was linked to different inherited forms of muscular diseases in humans, and the SelenoN protein was shown to be implicated in muscle development and regeneration. So far, the pathogenic mechanisms behind the muscular disorders are poorly understood since the catalytic function of SelenoN remains elusive, and no treatment is available yet. *SELENON* gene homologs were identified in most animals, including vertebrates and invertebrates, but recently, bioinformatics screening also identified multiple *SELENON* genes in one single group of yet-to-be cultured bacteria, *Candidatus Poribacteria,* that live in symbiosis with marine sponges. Therefore, we got interested in the function of *SELENON* in the bacteria and its possible role in the symbiotic interaction between the *Poribacteria* and its sponge host. From identifying SelenoN function in bacteria, we anticipate that we can make a link with its function in eukaryotes, and toward a better understanding of the muscle physiopathology in animals.

This overview acts as a guide throughout the different parts of the manuscript. The first part includes the introduction that is divided into two parts: First, Introduction I that will be talking about selenium utilization in life, evolutions of selenoproteins in prokaryotes and eukaryotes, and highlighting different important selenoprotein functions. Introduction II, will provide information about marine sponge taxonomy, the importance of its associated microbial community and our targeted bacteria *Candidatus Poribacteria.*

Proceeding to the next sections of the Results that include three parts: bioinformatic studies of the bacterial *SELENON* gene, amplification and expression of the related SelenoN protein, and finally microscopy studies to investigate the distribution of the *Candidatus Poribacteria* and SelenoN expression in a sponge model. A discussion section will contrast the results with the challenges we faced in the project, and present future perspectives. The materials and methods section will be the last part in this manuscript.

INTRODUCTION I

I. Selenium overview

1. Selenium and its physiological importance

The trace element selenium (Se) was discovered in 1817 by the Swedish chemist, Jöns Jacob Berzelius. Se is the 34th element in the periodic table with a molecular weight of 78.971 Da and was named after the Greek goddess of the moon ‘’Selene’’ as a metaphor for the mysterious side of the moon reviewed in (Tsuji et al., 2022). Inorganic Se is preserved in nature in four oxidation states: selenate (SeO_4^{2-}) and selenite (SeO_3^{2-}), both are highly soluble in water; elemental selenium (Se) is essentially non-toxic and highly insoluble in water; conversely selenide (Se_2^-) is a highly toxic and reactive (Mangiapane et al., 2014).

Se has been known for many years as a toxic element, as Se toxicity was reported already in the 13th century in China and later in the 20th century in the USA when a group of cattle and horses were diagnosed with necrotic hoof disorder accompanied by loss of hair in the mane and tail after feeding on plants rich in Se (Tsuji et al., 2022). However, Se gained its good reputation in 1951, when Se was observed to prevent dietary liver necrosis in rats and pigs and multiple necrotic degenerations in different organs in mice (Schwarz & Foltz, 1958). Several years later, it was shown to aid in the anaerobic growth of *Escherichia coli* (Enoch & Lester, 1972). Therefore, Se importance as an essential trace element in all forms of life was appreciated, nevertheless in a constrained dose.

Se supplementation in the diet of livestock was shown to benefit different conditions such as enhanced fertility in cattle, alleviation of white muscle disease in lambs and calves, and pancreatic degeneration. Even though, Se is available in the diet including seafood, eggs, meat, poultry, broccoli, and certain grains and nuts, the Se content in these different dietary sources varies accordingly and is mainly dependent on its reservoir in soil (Franke, 1934) (Tsuji et al., 2022).

In the case of Human health, pathogenic conditions such as Keshan disease, cardiomyopathy primarily identified in children (Ge et al., 1983), and Kashin–Beck disease, deforming osteoarthritis (Yu et al., 2019), were found in certain rural regions in China where the soil is deficient in Se reservoir, and almost the Se status of the individuals living therein is low. However, Keshan disease was eradicated in China after Se supplementation in the diet (Tsuji et

al., 2022). Later, Se was identified as an essential supply micronutrient for patients with chronic malnutrition and especially after long-term total parenteral nutrition (Stockhausen, 1988). In addition to having roles in preventing heart disease and other muscle disorders, as well as enhancing male fertility, Se was found to be important as a chemo-preventive agent in certain cancers, in boosting immune function, and in inhibiting HIV and other viral replication. In addition, Se possibly can slow the aging process that results from a prolonged accumulation of molecular damage to cells and tissues in response to oxidative stress, mainly through the modulation of oxidative damage and DNA repair capacity by some of the selenium-containing proteins such as Glutathione Peroxidase (GPx) (Mangiapane et al., 2014).

The shift between Se deficiency and toxicity is remarkably narrow, and balanced dietary Se intake is important to health. Even more, the form of dietary Se supply, for example, mineral versus organic forms, affects its incorporation in different tissues. In the USA, the Food and Nutrition Board of the National Academies of Medicine defines the recommended dietary allowance (RDA) for Se as 55 µg per day for adults, dose increasing for pregnant or lactating as 60 or 70 µg per day (Peters et al., 2016).

2. Selenium biological forms

Selenium is present in multiple different biological forms (Figures 1 and 2), the major one being **selenocysteine** (Sec, U) the 21st amino acid present in selenium-containing proteins and found in the three domains of life, eukaryotes, prokaryotes, and archaea. Sec insertion is an expansion of the genetic code since it is encoded by the UGA codon that is classically recognized as a stop and is specifically co-translationally incorporated into the growing polypeptide chain through a complex insertion machinery (Tsuji et al., 2022).

The second major form is the **2-selenouridine** (Se^2U), an analog of uridine, that can be further modified into 5-methylaminomethyl-2-selenouridine (mnm^5Se^2U) and 5-carboxymethyl-aminomethyl-2-selenouridine ($cmnm^5Se^2U$). These naturally modified nucleotides are other Se forms found at the wobble position of the anticodons of bacterial Glu, Lys, and Gln tRNAs. Selenouridine is synthesized from a 2-thiouridine (S^2U) precursor by 2-selenouridine synthase (YbbB or SelU). However, why nature introduced selenium into the wobble uridines of tRNA is

not fully understood. It was shown that the oxidized state of Se^2U is more reversible than its thiol-equivalent S^2U (Kulik et al., 2022). Therefore, Se^2U was suggested to play a quality control role in protein translation (Zhang et al., 2022).

Selenolate of Sec, a derivative of selenol where a metal atom replaces the hydrogen attached to Se. it is found as a ligand in some of the bacterial and archaeal coenzymes such in (1) the molybdenum atom of molybdopterin guanine dinucleotide in formate dehydrogenase, (2) the nickel atom of NiFeSe hydrogenases, and (3) iron in a putative iron−sulfur cluster in the methionine sulfoxide reductase from Metridium senile (sea anemone) (Khangulov et al., 1998; Reich & Hondal, 2016; Yamazaki, 1982). Two proteins are involved in these reactions YqeB and YqeC (Peng et al., 2016).

Selenoneine is another biological form of Se and is an analog of ergothioneine, seleno-, and sulfur-derived histidine respectively. Selenoneine is a naturally occurring amino acid found in certain organisms such as *Cyanobacteria*. Interestingly, it was shown that **selenoneine** possesses strong antioxidant activity and is the major form of organic selenium in the blood, muscles, tuna, and several other types of fish. In addition, it was shown to have a role in mercury detoxification. (Yamashita et al., 2013; Yamashita, (山下 & Yamashita, (山下, 2010). Recently the biosynthetic pathway for Selenoneine in different microorganisms was identified, demonstrating the use of an unexpected selenosugar as a biosynthetic intermediate, different from the selenophosphate used for other Se-containing compounds (Kayrouz et al., 2022).

Selenomethionine (SeMet). SeMet is a selenium analog of methionine (Met) and is incorporated into proteins in a random non-specific manner in place of Met using the regular $tRNA^{Met}$ since there is no $tRNA^{SeMet}$ in nature. The SeMet substitution to Met is dependent on SeMet to Met cellular ratio, which results from the non-discriminating activity between selenium and sulfur of methionine synthesizing enzymes (Mangiapane et al., 2014). SeMet is the major form of selenium in plants, therefore it is an important source of selenium in diet. Animals can convert SeMet dietary source to Sec via the transsulfuration pathway (Reich & Hondal, 2016; Weekley & Harris, 2013).

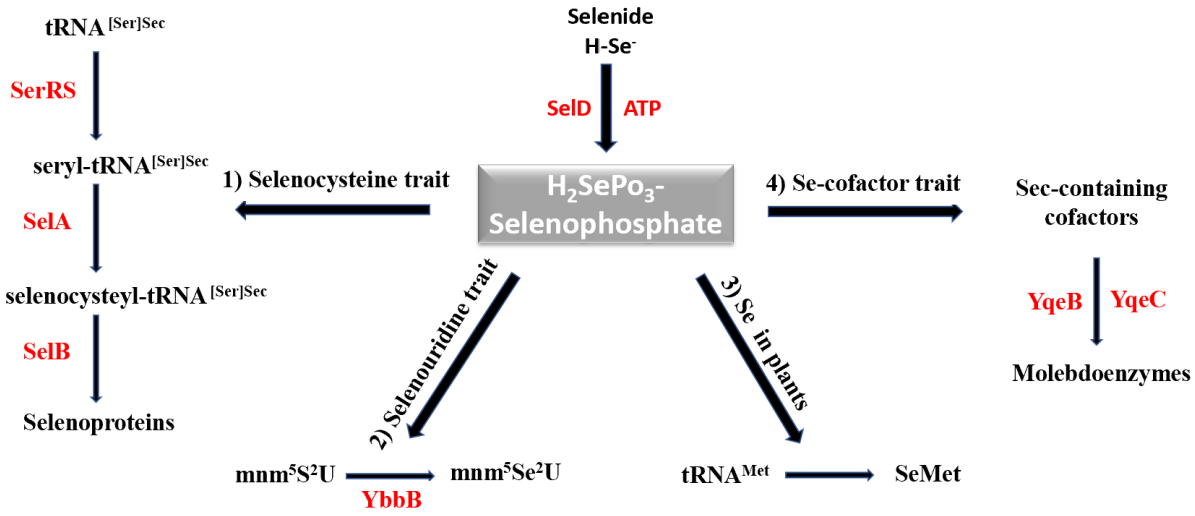


Figure 1. A schematic representation of the major Se-utilization traits in prokaryotes and plants. Each trait has unique genes. The selenophosphate acts as the active Se donor in all three traits, which is synthesized from selenide and ATP by selenophosphate synthetase (SelD) which serves as a general signature for all Se-utilization traits. 1) The seryl-tRNA synthetase (SerRS) charges the $tRNA^{[Ser]Sec}$ with serine to yield seryl-$tRNA^{[Ser]Sec}$, this will be converted to selenocysteyl-$tRNA^{[Ser]Sec}$ by SelA, to be incorporated into selenoproteins. 2) In the selenouridine trait the selenouridine synthase (YbbB) forms the selenouridine from its sulfur homolog. 3) selenomethionine is the major form of Se in plants. 4) Se-cofactor traits are utilized in metal-selenoproteins. Created by Microsoft PowerPoint and adapted from (Zhang et al., 2022). Note that for animals only the selenocysteine trait is represented.

3. Physiochemical properties of selenocysteine versus cysteine

The structure of Sec is homologous to cysteine (Cys), with a Se atom instead of a sulfur atom (Figure 2). Sec insertion machinery is an exhausting process on the cells, but what are the catalytic advantages provided by this energetically costly mechanism? Sec has several unique chemical and physical properties provided by the selenium atom that might have determined its evolutionary selection over Cys in catalytic sites of some selenoproteins and make it more favorable in thiol-disulfide exchange reaction rather than Cys.

The characteristics of Se atom make it a better nucleophile and electrophile than the sulfur atom. Moreover, the selenol group of Sec is more easily ionized at physiological pH, and is much more acidic than the thiol group in Cys, since its pKa is 5.2 compared to the 8.5 pKa of a thiol group. In addition, Se covalent radius is higher than the sulfur radius (1.16 Å vs. 1.03 Å), thus a sulfur-selenium bond is more stable than a sulfur-sulfur bond. Altogether these properties favor the stabilization and the acceleration of a thiol-disulfide exchange reaction in presence of Sec (Arnér, 2010). Therefore, due to the above-mentioned characteristics, Sec is more reactive than Cys, and Sec-containing enzymes employ a wide range of substrates and can function in different pH conditions.

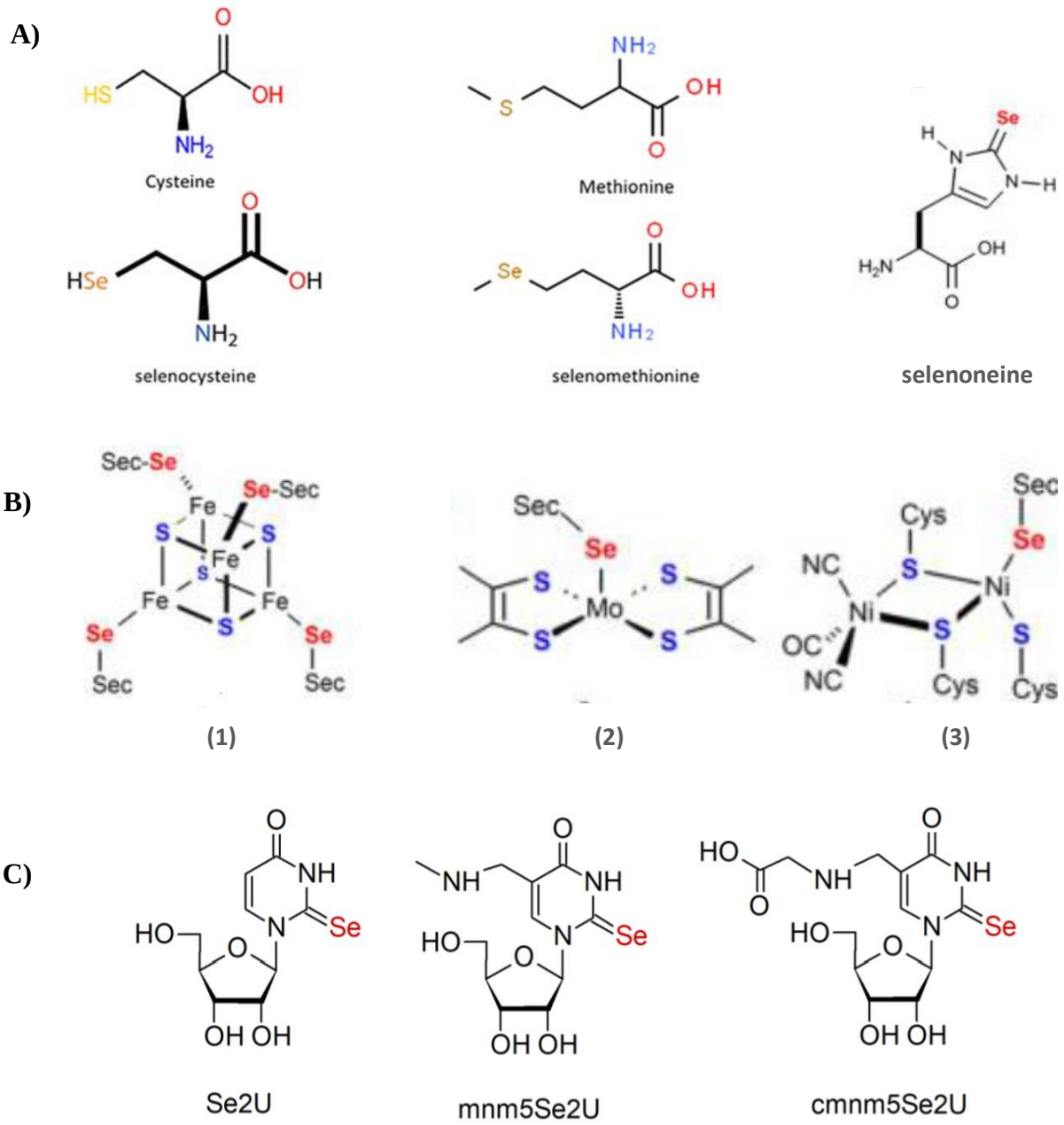


Figure 2. Structures of Se-biological forms. A) Structure of cysteine and methionine amino acids in comparison to seleno-amino acids. (Arshad et al., 2021; Reich & Hondal, 2016). **B) Selenium-containing cofactors**. 1) Selenium, as selenocysteine, is a putative ligand for iron in an iron−sulfur cluster. 2) Selenium, as selenocysteine, is a ligand for the molybdopterin guanine dinucleotide cofactor of formate dehydrogenase. 3) Selenium, as selenocysteine, is a ligand for nickel in [NiFeSe] hydrogenases (Reich & Hondal, 2016). **C) The structures of 2-selenouridine Se2U and its substituted counterparts mnmSe2U and cmnmSe2U present in wobble position of bacterial tRNAs** (Kulik et al., 2022).

4. Selenocysteine biosynthesis and insertion machinery

Sec has unique insertion machinery, in contrast to the 20 canonical amino acids: Sec is co-translationally inserted into the growing polypeptide through the aid of multiple auxiliary proteins to decode the opal stop codon UGA into Sec. The pathway of Sec biosynthesis and insertion shows some differences between prokaryotes, eukaryotes, and archaea reviewed in (Allmang et al., 2009; Labunskyy et al., 2014; Mariotti et al., 2016; Vindry et al., 2018)**.**

The main cis- and trans-acting factors involved in Sec biosynthesis are:

- *cis*-acting Sec insertion sequence (SECIS), (Allmang & Krol, 2006; Berry et al., 1991; Tujebajeva et al., 2000), a highly conserved stem-loop structure on the selenoprotein coding mRNAs. In prokaryotes this motif is located about 17 nucleotides downstream of the UGA Sec codon, as in eukaryotes this RNA structure is present in the 3' untranslated region. SECIS elements are not necessarily conserved in all bacterial species, and there is no evidence that all bacterial selenoproteins possess a SECIS element (Gursinsky et al., 2000; Mangiapane et al., 2014).
- *cis*-acting selenocysteine codon redefinition element (SRE) (Howard et al., 2007), an additional stem-loop structure that starts just 6 nucleotides downstream the UGA codon in a subset of the eukaryotic selenoprotein mRNAs.
- Sec unique tRNA ($tRNA^{[Ser]Sec}$) encoded by *SelC* gene in prokaryotes (Hatfield et al., 1994; B. J. Lee et al., 1989). The crystal structures of bacterial and human $tRNA^{[Ser]Sec}$ have been solved (Itoh et al., 2009) as shown in (Figure 3.1). Is characterized by its unique secondary structure and nucleotide length, although with some differences between the different domains of life (Ishii et al., 2013; Meng et al., 2022; Sturchler et al., 1993) (Figure 3.2). Two isoforms of $tRNA^{[Ser]Sec}$ were identified in mammals (Diamond et al., 1993), reviewed in (Labunskyy et al., 2014).
- Sec synthase enzyme (SecS/SepSecS) encoded by *SelA* gene in prokaryotes (Forchhammer & Böck, 1991).
- A specialized dedicated elongation factor, called eEFSec in eukaryotes (Fagegaltier et al., 2000; Small-Howard et al., 2006) encoded by *SelB* gene in prokaryotes, reviewed in (Labunskyy et al., 2014).

- Selenophosphate (SeP), a selenium donor synthesized from selenide and ATP in a reaction catalyzed by selenophosphate synthetase in both prokaryotes and eukaryotes. SPS2 or SEPHS2 in eukaryotes is encoded by *SelD* gene in prokaryotes. In eukaryotes, there are two homologs of the SPS, but only the SPS2 is active in selenophosphate synthesis(Guimarães et al., 1996; I. Y. Kim & Stadtman, 1995) , reviewed in (Labunskyy et al., 2014).
- SECIS binding protein (SBP2) in eukaryotes, an RNA binding protein that binds to the eukaryotic SECIS element (Allmang et al., 2002; Low et al., 2000; Takeuchi et al., 2009).
- Phosphoseryl-tRNA kinase (PSTK) specific to eukaryotes and archaea.
- Other SECIS-RNA binding proteins, such as the ribosomal protein L30, translation initiation factor 4A3 (eIF4A3) and nucleolin, have been identified as modulators of selenoprotein synthesis in eukaryotes (Budiman et al., 2009; Chavatte et al., 2005; R. Wu et al., 2000)

Sec is synthesized on its $tRNA^{[Ser]Sec}$ (X.-M. Xu, Carlson, Mix, et al., 2007) in multiple successive steps:

1) $tRNA^{[Ser]Sec}$ in presence of Serine (Ser) moiety is initially aminoacylated by the canonical seryl-tRNA synthetase (SerRS) to form Seryl-$tRNA^{[Ser]Sec}$ (Ser-$tRNA^{[Ser]Sec}$) as the Ser moiety provides the carbon backbone for the Sec synthesis.

2) Phosphoseryl-tRNA kinase (PSTK) catalyzes the formation of phosphoseryl $tRNA^{[Ser]Sec}$ (PSer-$tRNA^{[Ser]Sec}$)as an intermediate, this phosphorylation step is not present in prokaryotes.

3) Sec incorporation through the Sec synthase (SecS/ SelA) collaboration with the Se donor SPS enzyme (SelD) and catalyzes the conversion of Ser to Sec on the tRNA to synthesize the selenocysteinyl-tRNA (Sec-$tRNA^{[Ser]Sec}$).

4) The molecular mechanism of eukaryotic Sec incorporation is as follows; in prokaryotes SelB binds simultaneously to GTP, the SECIS motif, and the Sec-$tRNA^{[Ser]Sec}$. While in eukaryotes the SECIS element binds SBP2 and eEFSec that recruits GTP and the Sec- $tRNA^{[Ser]Sec}$. By forming these complexes, the UGA stop codon is decoded into Sec during the elongation step of protein translation (de Jesus et al., 2006; Howard et al., 2012; Small-Howard et al., 2006; Vindry et al., 2018). The translation machinery shown in (Figure 4) is taken from (Mariotti et al., 2016) and shows Sec biosynthesis and insertion in the three domains of life.

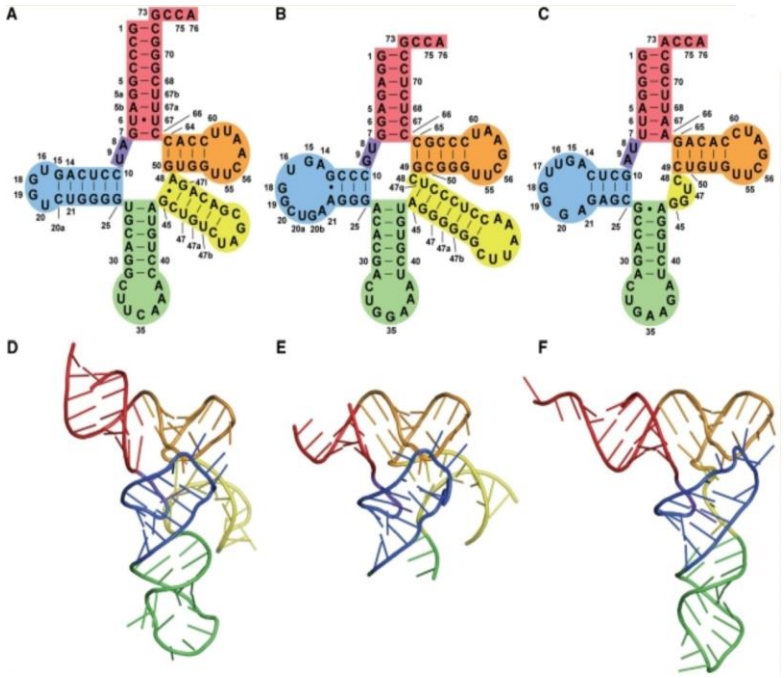


Figure 3.1 Structure of human tRNA$^{[Ser]Sec}$, and its comparison with other tRNAs. Human tRNA$^{[Ser]Sec}$ is shown as a cloverleaf model (**A**), a ribbon model (**D**). For comparison, *T. thermophilus* tRNA^{Ser} [PDB ID: 1SER (38)] (**B**, **E**) and *S. cerevisiae* tRNA^{Phe} [PDB ID: 4TNA (47)] (**C**, **F**) are shown. The acceptor arm, AD linker, D arm, anticodon arm, extra arm and T arm are colored red, purple, blue, green, yellow and orange, respectively (Itoh et al., 2009).

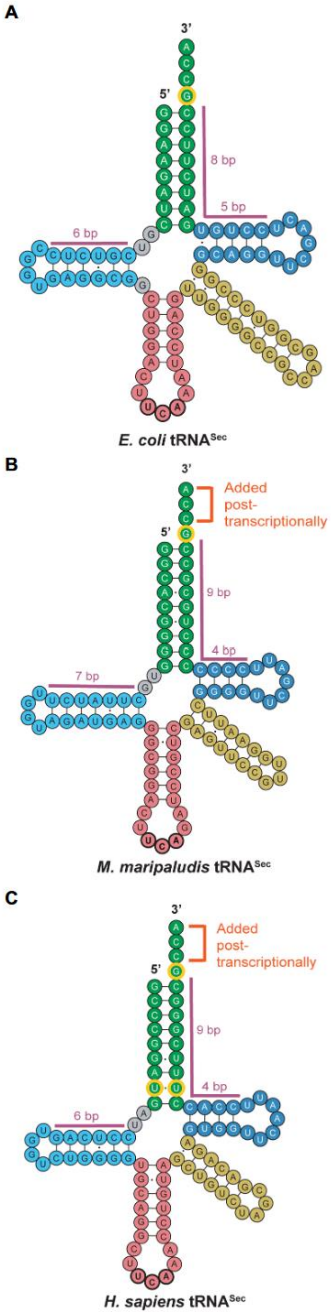


Figure 3.2 tRNA$^{[Ser]Sec}$ secondary structures across the three domains of life; (A) bacterial *E. coli*, (B) archaeal *M. maripaoudis*, and (C) eukaryotic *H. sapiens* tRNA$^{[Ser]Sec}$.

The tRNAs$^{[Ser]Sec}$ are the longest tRNAs through eukaryotes, archaea, and bacteria. It has unique features, such as a long variable arm, long acceptor arm, and long D arm, that prevent its recognition by the general translation factors EFTu or EF but conversely recruit the factors required for Sec synthesis and translation.

A few differences are noted:
1)The canonical archaeal and eukaryotic tRNA$^{[Ser]Sec}$ share the post-transcriptionally added CCA tail and the 9/4 configuration of the acceptor stem (green) and T-arm (dark blue).

2) The bacterial tRNA$^{[Ser]Sec}$ possesses a longer variable arm (yellow) than the other two domains of life.

3) The archaeal tRNA$^{[Ser]Sec}$ possesses a longer D-arm (light blue) than both bacterial and eukaryotic tRNA$^{[Ser]Sec}$.

4) The tRNA$^{[Ser]Sec}$ for all three domains share the G73 discriminator base as an identity element (circled in yellow), but eukaryotic tRNA$^{[Ser]Sec}$ also has a conserved U6:U67 base pair. (Meng et al., 2022).

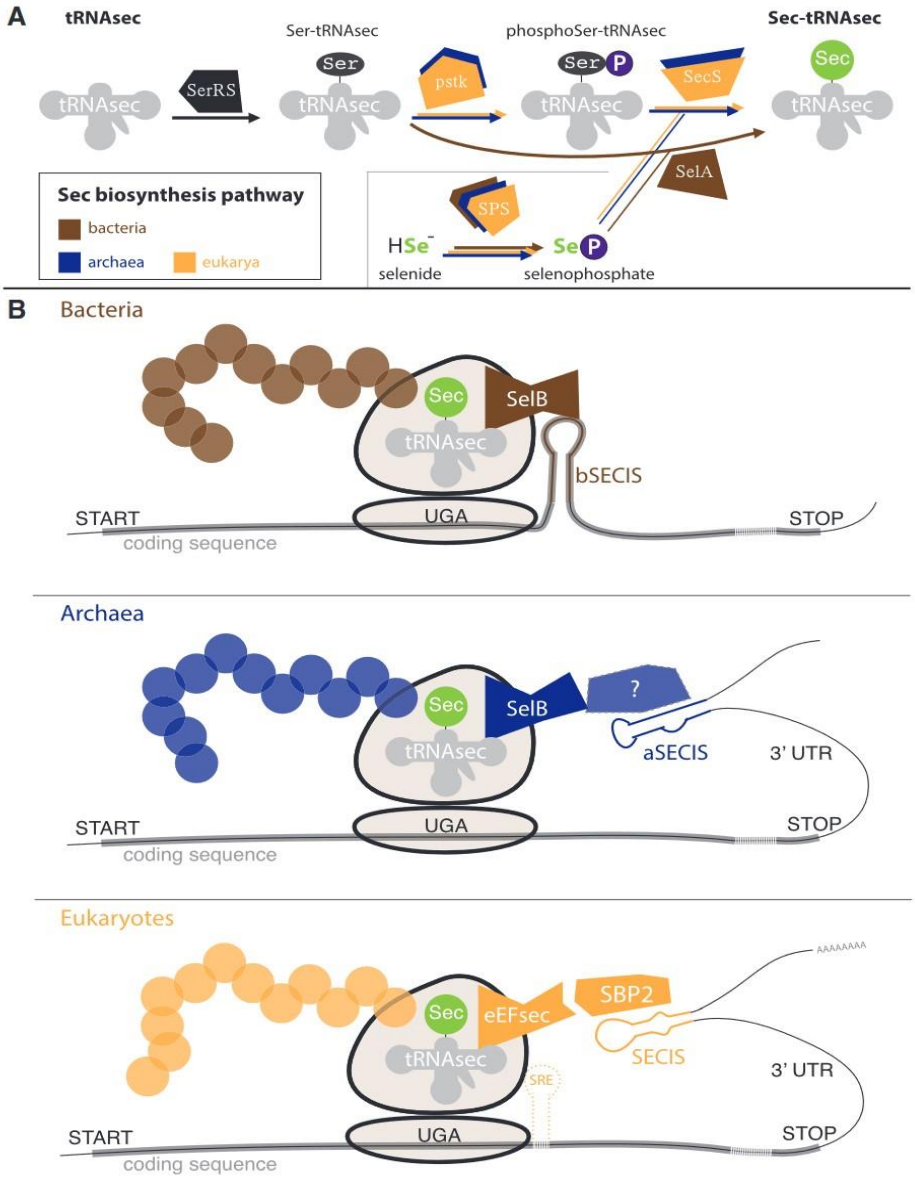


Figure 4. Mechanism of Sec biosynthesis and insertion in the three domains of life. A) Sec biosynthesis. B) Sec insertion into the mRNA of bacteria, archaea, and eukaryotes highlighted in different colors (Stock & Rother, 2009; Yoshizawa & Böck, 2009). (SerRS) seryl-tRNA synthetase, (pstk) phosphoseryl-tRNA kinase, (SecS) archaeal/eukaryotic Sec synthase, (SelA) bacterial Sec synthase, (SPS) selenophosphate synthetase, (SelB) bacterial/archaeal Sec elongation factor, (eEFsec) eukaryotic Sec elongation factor, (SBP2) SECIS binding protein 2, (bSECIS) bacterial SECIS, (aSECIS) archaeal SECIS, (SECIS) eukaryotic SECIS, (SRE) Sec redefinition element (Mariotti et al., 2016).

Note that importantly, the structure of the mammalian ribosome decoding the UGA codon was solved at an atomic resolution (Hilal et al., 2022) and it showed that the SECIS acts as a bridge between the SBP2 and the eEFSec·Ser-tRNA^{Sec} to form eEFSec·Ser-tRNA^{Sec}·SECIS·SBP2 complex that directly positions the anticodon into the ribosome decoding center (Figure 5).

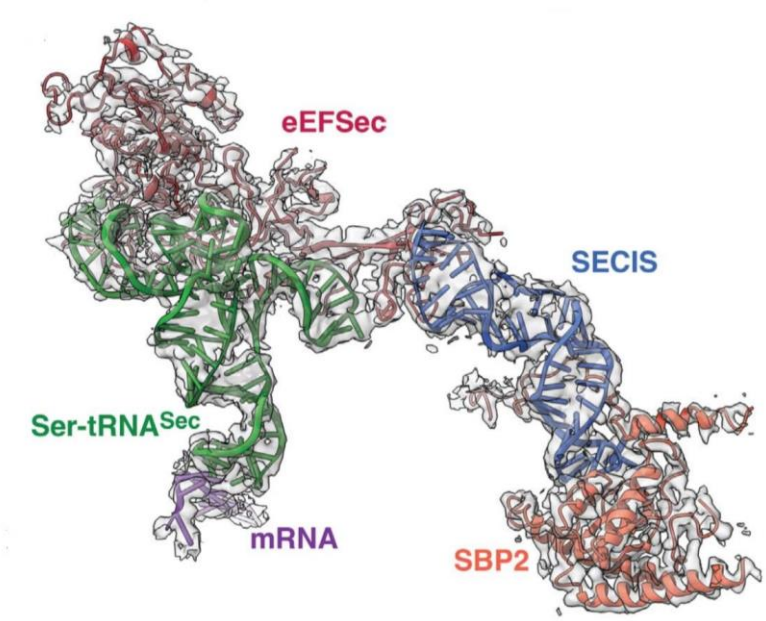


Figure 5. CryoEM map for eEFSec·Ser-tRNA^{Sec}·SECIS·SBP2 complex

II. Selenoproteome

1. Selenoproteomes Evolution

Sec-insertion is a complex mechanism and similar although with slight differences in the three domains of life. Upon these insertion characteristics, Sec has been considered alternatively as remaining of the primitive genetic code, or as an addition to the current genetic code that was next evolved due to its superior redox properties over Cys in biological functions. Even though, it was observed that the Sec insertion mechanism was lost in some lineages across the evolution and cannot be restored. Therefore, understanding the evolution of selenoproteomes as well as Se-utilization patterns in living organisms is a central question to selenium biology. (Gladyshev & Kryukov, 2001; Romero et al., 2005)

1.1. Evolution of Sec insertion mechanism in Prokaryotes.

As mentioned previously Se is utilized in living organisms through two major traits, Sec-decoding and SeU-synthesis. Se-utilization by both pathways was analyzed in two consecutive studies, the first one involved completely sequenced genomes of 151 different prokaryotic species (Romero et al., 2005), and the second one involved completely and incompletely sequenced bacterial genomes of 349 species (Zhang et al., 2006). The key factors in the analysis were Sec-decoding genes (*selA, selB, selC, selD*) and 2-selenouridine synthesis genes (*ybbB, selD*), as selD is the Se donor in both traits. Another additional criterion in the case of the Sec-decoding trait is the presence of at least one known selenoprotein in the genome. In the first study, they identified 29 *selA*, 32 *selB*, 41 *selD*, and 21 *ybbB* sequences from the total analyzed genomes of the 151 prokaryotic species. While in the second study, out of 349 different bacteria species, seventy-five selenoprotein-containing bacteria were found and 84 *selA*, 75 *selB*, 127 *selD*, and 88 *ybbB* genes were identified. In both studies, *selD* was found represented in both traits as a representation of overall Se utilization.

In addition, some species were found to possess the two Sec-utilization traits. A phylogenetic tree was constructed based on the 16S rRNA of the 151 microbial species and identified several horizontal gene transfers (HGT) events involving both traits. This phylogenetic analysis also proved that these traits evolved across all the taxonomical levels of the microbial species as a result of speciation and differential gene loss or HGT.

Not so far, in 2016, a study repeated the analysis of Se-utilization machinery on more than 5200 sequenced bacterial genomes (Peng et al., 2016). In this study not only the Sec-decoding and SeU-synthesis traits were analyzed, but also the Se-cofactor trait with its key components (*SelD*/*YqeB*/*YqeC*). The result of the study identified 21.5% of Sec-utilizing, 18.8% of SeU-utilizing, and 6.0% of Se-cofactor-utilizing organisms. This study generated the largest Se-utilization map in bacteria till today (Figure 6, 7) showing the distribution of key components involved in each Se-utilization trait. In addition, environmental factors of studied organisms have been taken into account, and favorable conditions for some Se-utilization traits over the others, such as O2 exposure, temperature, and habitat. All studies of Se-utilization traits have added to the understanding of the evolution of the genetic code and the Se-utilization machinery in bacteria. And as shown above, SelD is required for all three traits and is considered SelD-based Se-utilization dependent. Another interesting finding is the identification of an orphan *SelD* gene in the last two studies with no accompanying of any of the Se-utilization traits keys (Figure 6). This thought to be either due to the presence of a new SelD-based Se-utilization trait, or a loss of Se-utilization traits in these organisms. However, a very recent bioinformatics analysis identified this *SelD* cluster as the enzyme involved in the synthesis of the precursor of selenoneine, therefore defining a new Se-utilization pathway (Kayrouz et al., 2022).

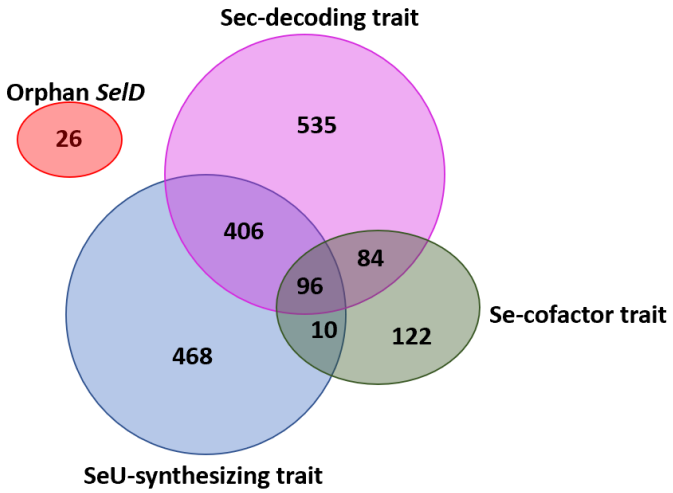


Figure 6. Venn diagram representation of the distribution of Se-utilization traits in prokaryotes. The number of organisms containing the corresponding Se traits is indicated. Created by Microsoft PowerPoint and adapted from (Peng et al., 2016)

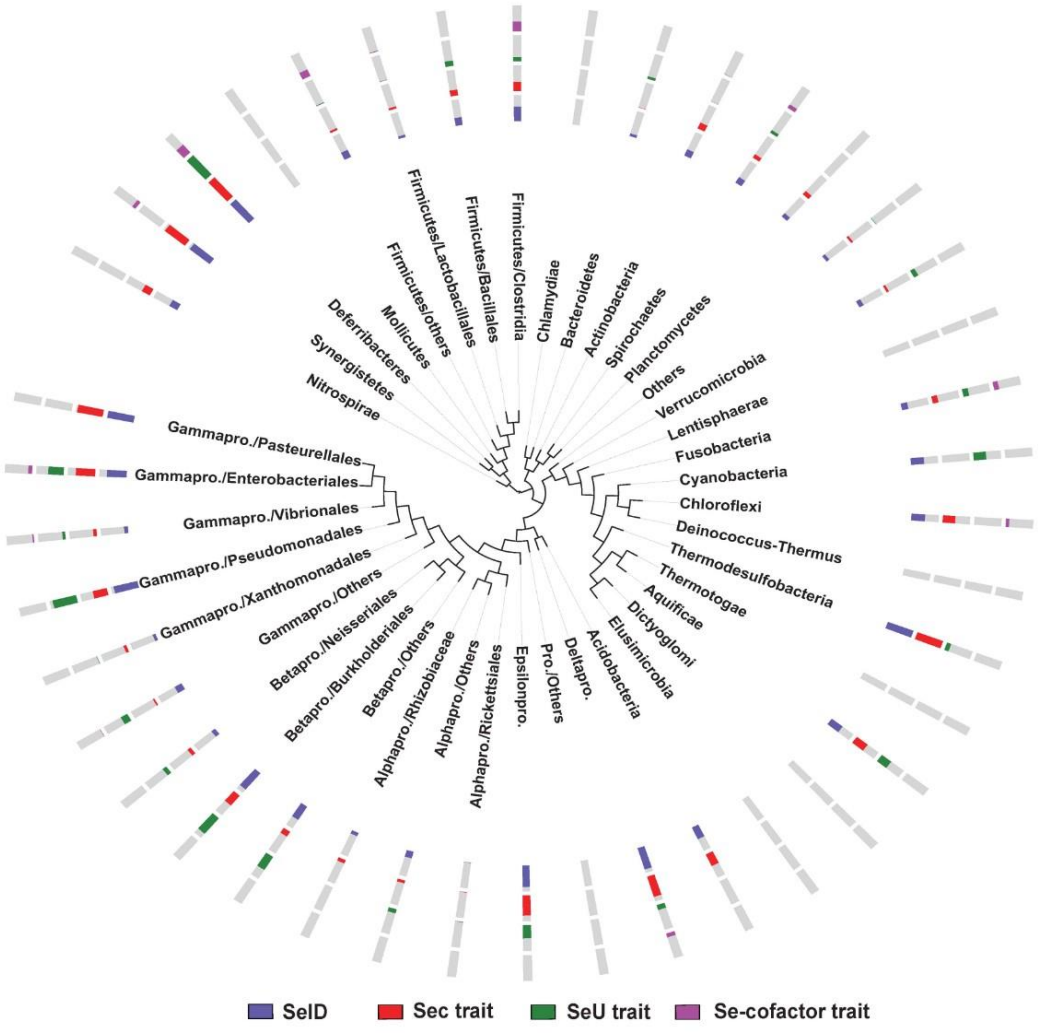


Figure 7. Phylogenetic tree of the distribution of *SelD* and Se-utilization traits in bacteria. For simplicity, only major bacterial taxa and branches are shown. The length of the colored section of each column represents the percentage of organisms that have either *SelD* or the corresponding Se trait among all sequenced organisms in this phylum (Peng et al., 2016)

1.2. Evolution of Selenoproteomes in Eukaryotes

The conservation of genes involved in Sec biosynthesis, including the TGA-Sec codon sequence, in all three domains of life eukaryotes, prokaryotes, and archaea, suggests that Sec-decoding machinery dates back to the last common ancestor. To appreciate the evolution in eukaryotes, the selenoproteome of six different eukaryotic model organisms was analyzed and compared to the mammalian selenoproteome. This comparison revealed that 20 out of the 25 Human selenoproteins have homologs in many unicellular organisms, suggesting that some selenoprotein families might be originated in singe-cell eukaryotes (Lobanov et al., 2007).

Moreover, the distribution of selenoproteins across eukaryotes lineages indicates a highly dynamic evolutionary history (Figure 8) (Santesmasses et al., 2020). First, many selenoproteins were found to be conserved between single-cell eukaryotes and vertebrates, but some replaced Sec by Cys homologs (Lobanov et al., 2007, 2009; Santesmasses et al., 2020). Secondly, the selenoproteome size represented in many selenoproteins varies among different species. *Aureococcus anophagefferens,* an algae that possess 59 selenoproteins, represents the largest selenoproteome to date, in contrast to the nematode *Caenorhabditis elegans* genome that only encodes one single selenoprotein, Thioredoxin Reductase (TrxR). Thirdly, many eukaryotic selenoprotein genes were lost in different lineages in the course of evolution (Santesmasses et al., 2020).

Selenoproteins are very rich in metazoans. Three selenoproteins were found to be specific to vertebrates: SELENOI (EPT1 or SelenoI), SELENOV (SelenoV), and SELENOE (Fep15) (Mariotti et al., 2012; Santesmasses et al., 2020). Studies on selenoproteomes in vertebrates showed that evolution in selenoproteins is under the influence of two environmental factors, selenium availability, and aquatic environmental conditions. Indeed, the largest vertebrate selenoproteomes were identified in fish with 38 selenoproteins, and it was hypothesized that it is linked to the highest Se accessibility in water (Mariotti et al., 2012; Santesmasses et al., 2020). Another characteristic example of this evolution among vertebrates is Selenoprotein P (SelenoP) which is characterized by the presence of multiple Sec residues in its polypeptide chain and two SECIS elements; the number of Sec residues in SelenoP was found to be higher in fish than in mammals, and two forms of SelenoP were detected in fishes (Mariotti et al., 2012). In addition,

most of the mammalian selenoproteins are shared with fish except for selenoprotein E, selenoprotein J, and selenoprotein L (Fep15, SelenoJ, SelenoL) which are fish-specific (Lobanov et al., 2009). Fish also possess a Sec-containing selenoprotein U (SelenoU), which is found as a Cys-homologue in mammals, that could be indicative of a possible ancient selenoprotein and was lost during evolution (Santesmasses et al., 2020). Thus, the importance of the environment in shaping the evolution of selenoproteomes in organisms was illustrated in the rich selenoproteome in aquatic life compared to land life (Lobanov et al., 2009; Santesmasses et al., 2020). In Humans, 25 selenoproteins were identified so far and they are quite conserved among mammals. Compared to Humans, rodent genomes contain one additional selenoprotein, Glutathione Peroxidase 6 or Gpx6, which is a Cys ortholog in Humans (Lobanov et al., 2009).

In the case of invertebrates, the available information revealed that the ecdyzoa clade formed by insects, arthropods, and nematodes have low selenoprotein composition due to multiple genes loss through evolution (Chapple & Guigó, 2008; Mariotti, 2016; Santesmasses et al., 2020). An example of selenoprotein gene loss is SelenoP which is absent in tunicates (invertebrate sea squirts), crustaceans, Platyhelminthes (flatworms), most nematodes, and most insects (Santesmasses et al., 2020). Unfortunately, no extensive study has been reported for this group yet, otherwise, this will be of interest to study the evolution of selenoproteins and Sec-utilization in invertebrates in comparison to vertebrates.

Fungal phyla were originally identified as selenoprotein-free, as an example, the yeast *Saccharomyces cerevisiae* is free of selenoproteins (Lobanov et al., 2007). However, recent genomic investigations identified selenoprotein-coding genes and Sec-utilization machinery in nine species belonging to several early-branching fungal phyla (Mariotti et al., 2019; Santesmasses et al., 2020). Interestingly, most fungal mRNAs coding for selenoproteins lack a consensus SECIS element but exhibit other alternative RNA structures, suggesting a divergent Sec insertion mechanism in fungi.

Similarly, selenoproteins and Sec insertion machinery are lost in higher plants. However, a recent study identified thousands of selenoprotein genes in 137 species of green algae, indicative of an early presence of selenoproteins and evolutive disappearance in this phylum as well (Jiang et al., 2020). These genes belong to 42 selenoproteins families, including three novel families. Of note, plants selectively absorb both selenate and selenite and accumulate Se from the soil. They reduce selenate to selenide, in which the latter can be converted to either SeMet or methylselenocysteine MeSeCys (Lobanov et al., 2007; Mangiapane et al., 2014), which constitute an important reservoir for Se nutrition supply in animals.

In Protists, a group of unicellular eukaryotes, the selenoproteins are highly distributed among the members of this group. As mentioned above the harmful pelagophyte, *Aureococcus anophagefferens,* possess the largest selenoproteome known to date, although in the same group, many selenoprotein-less species were identified (Gobler et al., 2013). The decrease of selenoproteomes size in some eukaryotes may be attributed to the loss of entire selenoprotein genes or the substitution of Sec with its Cys orthologs.

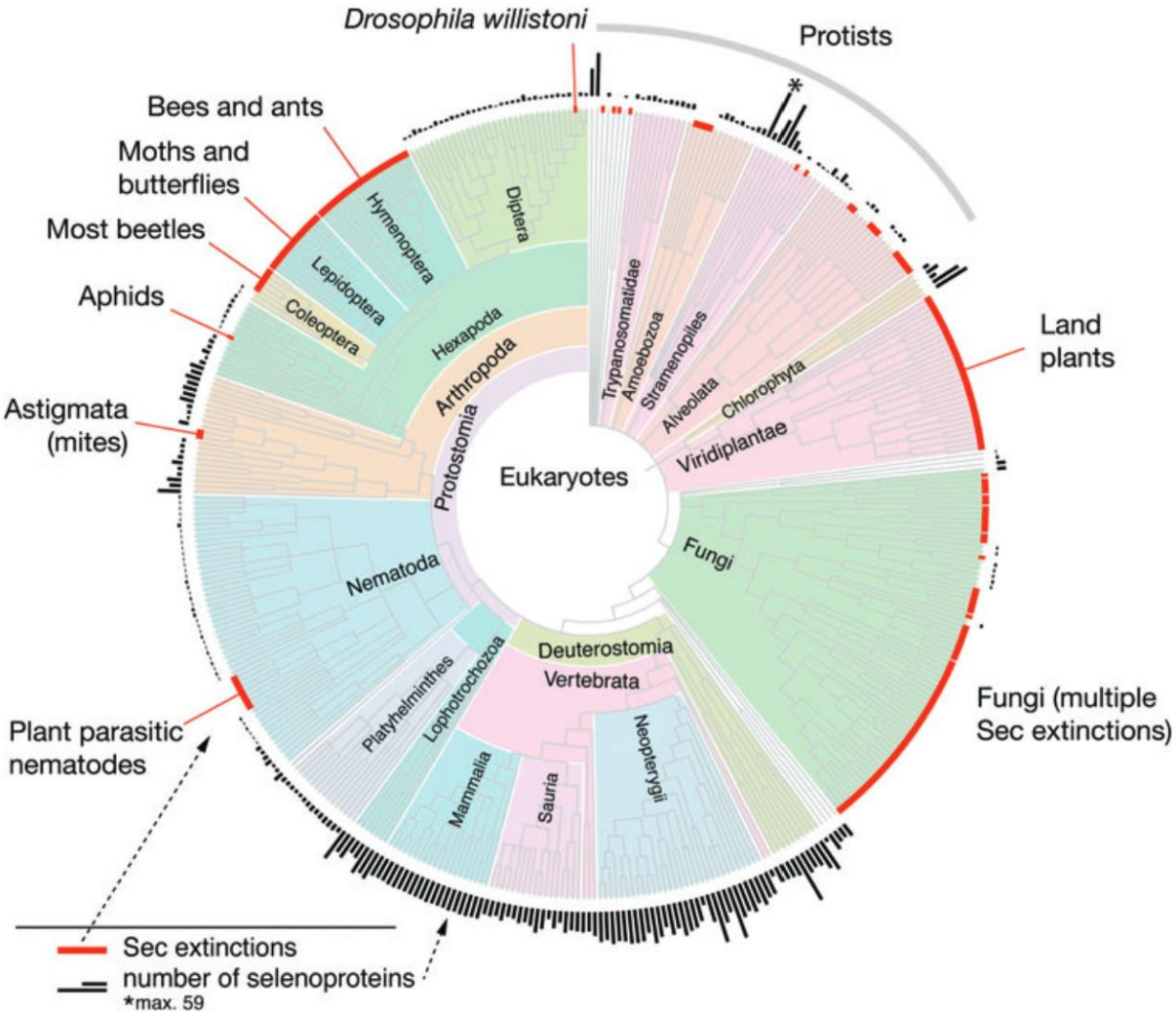


Figure 8. Phylogenetic tree of the distribution of selenoproteins and their extinctions in eukaryotes. The black bars represent the selenoproteome size (number of selenoproteins in the genome) for each species. The reported Sec extinctions are indicated in red. Plant, Fungal, and insect species show a high rate of Sec extinctions. Protists show a highly scattered distribution of selenoproteins (Santesmasses et al., 2020).

2. Selenoproteins.

Before we talk about selenoproteins, we will discuss the redox mechanism of the biological cell. Oxidative stress is caused by an accumulation of reactive oxygen species (ROS). Even though ROS are naturally generated in cells, excess ROS can damage proteins, DNA, and membrane lipids, and eventually lead to cell death. Most degenerative disorders, including cancer, cardiovascular diseases, diabetes, stroke, Alzheimer's, and Parkinson's disease have been linked to oxidative damage. Oxidation of proteins triggers cell damage and dysfunction leading to apoptosis and necrosis; oxidative damages to DNA are sources of somatic mutations; and lipid peroxidation modifies membrane fluidity and permeability. As a protection, cells have developed a highly complex ROS scavenging network to control the intracellular redox balance (Seiler et al., 2008). In addition, a dynamic equilibrium between cellular non-radical oxidative and reductive molecules constitutes an elaborate signaling pathway regulating many cellular activities in response to environmental modifications or challenges. (Jones & Sies, 2015). The founder specialized protein in protecting the cell against oxidative damages is the thioredoxin, a small ubiquitous protein, very stable at high temperatures, reduces protein disulfides, and is important in maintaining the cytoplasm in a reducing state (Pan & Bardwell, 2006). The thioredoxin protein contains the catalytic site of (CXXC), in which C corresponds to cysteine and X to any amino acid. The thioredoxin has evolved via gene duplication events and mutation which resulted in the alteration of function of many members in the thioredoxin superfamily while some structural features of many thioredoxins and thioredoxin-like proteins such as thioredoxin-fold are remained conserved. Each thioredoxin-like protein has adopted some changes to accommodate a particular reactivity either by manipulating the protein active site, changing the protein oligomeric state, or adding new domains. There are two oxidoreductase systems that can reduce protein disulfides in the cells, the thioredoxin (Trx) and the glutaredoxin (Grx) systems. The Trx reduces the disulfide bonds and is recycled by thioredoxin reductase enzyme (TrxR). While The Grx system is divided into dithiol, containing two Cys residues (CXXC), and a monothiol system containing one Cys residue (CXXS) (Pan & Bardwell, 2006). The classic thioredoxin/glutaredoxin fold is characterized by four β-sheets in the core center and surrounded by three α-helices (Figure 9), the original article is (Eklund et al., 1992). The dithiol Grxs reduces disulfides in proteins using reduced glutathione GSH as a reductant, whereas the glutathione reductase (GR) reduces the oxidized glutathione GSSG to GSH using NADPH as an

electron donor. Both system are responsible for reducing ribonucleotide reductase (RNR), the enzyme that is essential in the synthesis of deoxyribonucleotides. Therefore, Trx and Grx systems are two intricated redox processes controlling redox status in the cell (Hanschmann et al., 2013). Some selenoproteins have thioredoxin-like domain and posses oxidoreductase activities. Selenoproteins are distributed among the three domains of life as shown in a schematic representation of selenoproteins in (Figure 10).

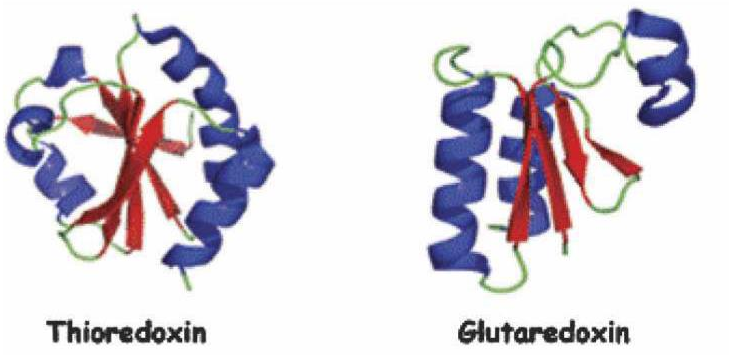


Figure 9. Ribbon structure of the thioredoxin and glutaredoxin that is characterized by four β-sheets surrounded by three a-helices (Pan & Bardwell, 2006).

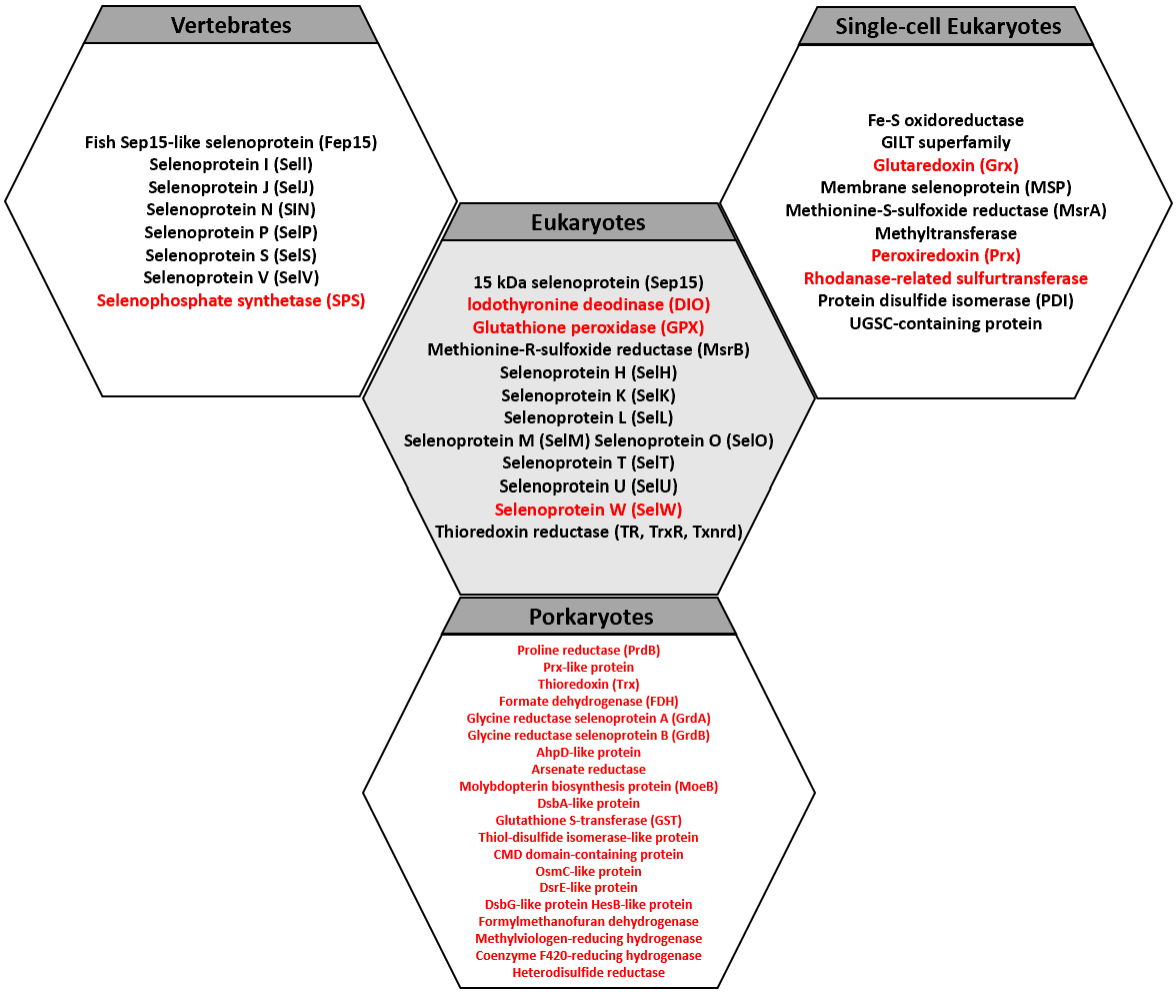


Figure 10. An illustration of selenoproteins distributed among the three domains of life. The middle hexagonal represents Selenoproteins that are shared among all eukaryotes, in red Selenoproteins in prokaryotes. Created by Microsoft PowerPoint, and adapted from (Labunskyy et al., 2014)

2.1. Selenoproteins shared in eukaryotes and prokaryotes.

Here are some of the most common selenoproteins that are shared among eukaryotes and prokaryotes.

2.1.1 Glutathione Peroxidases (Gpxs)

Gpxs constitute a very large family that is widely distributed across all three domains of life. The active site in all selenoprotein Gpxs is conserved, the catalytic mechanisms are identical, but the substrates and the physiological roles are very diversified. There are eight Gpxs paralogs in mammals, Sec-containing selenoproteins (Gpx1-4, Gpx6), and its Cys-containing homologs (Gpx5, Gpx7, and Gpx8). The functions of the Gpxs include detoxification and scavenging of hydroperoxides, regulation of ferroptosis and hydrogen peroxide (H_2O_2) signaling, and maintenance of cellular redox homeostasis in general (Arthur, 2001; Labunskyy et al., 2014; Tsuji et al., 2022).

a) **Gpx1** was the first glutathione peroxidase to be identified and is distributed in the cytosol of most cells. It is a 22 kDa protein, assembled in homotetramer and it is the most abundant selenoprotein in mammals. It uses a two-glutathione GSH redox system to catalyze the reduction of H_2O_2 (Burk & Hill., 2010).

In addition, Gpx1 activity regulates H_2O_2 signaling response. H_2O_2 has an important signaling role in the regulation of many biological processes and pathways, such as cell proliferation, apoptosis, stress response, and mitochondria-related functions. The level of Gpx1 can affect the level of H_2O_2 in the cells: Gpx1 deficiency leads to elevated levels of H_2O_2, while a high level of Gpx1 can interfere with the H_2O_2 signaling pathway (Labunskyy et al., 2014).

In addition, Gpx1 was predicted to have dual roles as an antioxidant and pro-oxidant under certain conditions. For example, GSH acts as a major scavenger of peroxynitrite ($ONOO^-$), and increased Gpx1 activity in mice led to depleting intracellular levels of reduced GSH. Accordingly, knockout of Gpx1 in mice protected primary hepatocytes against peroxynitrite, which suggests that GPx1 activity promotes the accumulation of reactive nitrogen species (RNS) and therefore is involved in oxidative stress in cells (Labunskyy et al., 2014; Zhu & Lei, 2006). Same to the type 2 diabetes-like phenotypes in Gpx1-overexpressing mice, the increase in Gpx1 expression level was linked to increased insulin resistance and hyperglycemia due to the

decreased level of intracellular H_2O_2 and disruption of its role in signaling pathways (Labunskyy et al., 2014; McClung et al., 2004). Therefore, imbalance in Gpx1 cellular level can react back negatively on the cell redox status.

b) **Gpx2** is a homotetramer and is expressed in both gastrointestinal epithelium and epithelium-derived tumors. Several studies showed the role of Gpx2 in the development and prevention of cancer in the late stages as a dual function. This was concluded from the observation that Gpx2 gene expression is under the control of the antioxidant response transcription factor Nrf2 which suggests the protective role of Gpx2 against cancer. On the other hand, Gpx2 expression was found to be also regulated by the Wnt pathway, a pathway required for the activation of cellular proliferation (Brigelius-Flohé et al., 2012; Hatfield et al., 2014; Labunskyy et al., 2014).

c) **Gpx3** is a plasma glutathione peroxidase, also called extracellular glutathione peroxidase. Gpx3 is a homotetramer mainly expressed in the kidney but also found in other tissues (An et al., 2018, p. 3; Burk & Hill, 2010). Gpx3 uses both thioredoxin and glutathione as reducing substrates (Björnstedt et al., 1994). Gpx3 expression level decreases in a number of Human cancer which suggests its role as a tumor suppressor (An et al., 2018). Although Gpx3 is an extracellular protein, and the concentration of GSH is lower in the extracellular space than from inside the cells, the GSH concentration near the cell plasma membranes is enough for the activity of the Gpx3 (Burk & Hill., 2010).

d) **Gpx4** family member is called phospholipid hydroperoxide glutathione peroxidase (PHGPX) and is the most essential peroxidase. Gpx4 is a monomer found in three forms, a cytosolic (cGpx4), mitochondrial (mGpx4), and a nuclear form (nGpx4). It functions in reducing fatty acid hydroperoxides and therefore protects lipids against peroxidation (Burk & Hill., 2010). The role of Gpx4 in inhibiting lipid peroxidation was observed in knock-out mice which developed an embryonic lethality, indicative of its role in early embryonic development (Yant et al., 2003).

Gpx4 is highly expressed in the brain where it plays a protective role in the brain against oxidative stress and neurodegeneration (Zhang et al., 2008). Low levels of GSH detected in many neurodegenerative diseases led to the inhibition of Gpx4 activity and therefore increased lipid peroxidation status (Labunskyy et al., 2014). Due to its importance in peroxyliplids reduction, Gpx4 was shown to play a central role in ferroptosis, a non-apoptotic form of

programmed cell death induced by iron-dependent membrane lipid peroxidation (Ursini & Maiorino, 2020; Zheng & Conrad, 2020)

e) **Gpx6** is a homotetramer protein excreted in the Bowman's gland (nasal gland) of the olfactory system (center of smell perception). As mentioned previously, Gpx6 is a mammalian Sec-containing Gpx, but a Cys-containing protein in mice, rats, and some other species (Labunskyy et al., 2014). The exact function of the Gpx6 is unknown, but it is thought to have a function related to odorant perception and control.

2.1.2 Thyroid Hormone Deodinases (DIO)

Thyroid hormones are essential for development and growth. This family is composed of three selenoproteins, Iodothyronine deodinases (DIO1, DIO2, and DIO3). The DIOs are integral membrane selenoproteins characterized by a thioredoxin fold connected to a single transmembrane domain at the N-terminus. The Deodinases family is involved in the regulation of thyroid hormone metabolism, maintaining the levels of thyroid hormone by activation and deactivation (Dentice et al., 2013). They have distinct subcellular localizations and tissue expressions: DIO1 and DIO3 are located in the plasma membrane, as DIO2 is localized in the endoplasmic reticulum (ER). Surprisingly, if the loss of deiodinases is lethal in Humans, it is tolerated in mice (Santesmasses et al., 2019). In addition to mammalian deodinases, homologs are also found in lower eukaryotes and bacteria but their function is unknown yet (Labunskyy et al., 2014).

The catalytic site is conserved among all deodinases with the structural motif SXXU that have similarities with the peroxiredoxins S/TXXC motif. This analogy between deodinases and peroxiredoxins was further evidenced by the analysis of the Dio3 crystal structure, suggesting a common origin and a functional relationship (Schweizer et al., 2014). Recently, the catalytic mechanism of DIOs was unveiled (Rodriguez-Ruiz et al., 2022)

Thyroid hormones are generally produced in an inactive form. The main thyroid prohormone produced in the thyroid gland and found in the circulatory system is thyroxine or tetraiodothyronine (T4). It is the precursor of the active triiodothyronine (T3). The activation of

T4 to T3 in the circulatory system is under the control of DIO1, while within the intracellular space, it is under the control of DIO2 (Bianco et al., 2002.; Dentice et al., 2013). In addition, the concentration of T3 is under the control of DIO3, and under specific conditions under the control of DIO1 that forms inactive T2 and reverse T3 (rT3), respectively.

Vital processes are regulated by thyroid hormones, including growth, development, and metabolic rate. Studies on DIO2 showed its importance in certain physiological processes such as muscle development and regeneration by regulating the activity of T3. DIO2 expression level is maximum during the normal development of skeletal muscle tissues right after birth and drops gradually. In addition, DIO2 activity increased in injured-muscle tissues. DIO2 activity was also linked to enhanced transcription of T3-dependent genes that are required for muscle differentiation and regeneration (Dentice et al., 2010, 2014).

Moreover, DIO2 has a role in regulating brown adipose tissue (BAT)-dependent thermogenesis that maintains the body's temperature and is involved in animal adaptation to cold conditions. DIO2 increases levels of activated T3 which in turn controls the activation of several genes involved in the BAT thermogenesis pathway (Silva, 1995; Silva & Larsen, 1985).

2.1.3 Selenophosphate Synthetase 2 (SEPHS2/SPS2/SepS2)

SPS2 (SelD in prokaryotes) synthesizes selenophosphate from selenide and ATP. It is characterized as the active Se donor for the synthesis of Sec and other Se-utilization traits as mentioned previously (Manta et al., 2022). The Sec-containing SPS2 was found in all vertebrates, whereas Sec residue is replaced with Cys in lower eukaryotes. SPS2 was also proposed to have a role in regulating selenoproteins synthesis (Labunskyy et al., 2014; X.-M. Xu, Carlson, Irons, et al., 2007). Some animals possess SPS1, a paralog of SPS2. However, the Sec in SPS1 catalytic site is replaced with other amino acids. Accordingly, SPS1 proteins do not synthesize selenophosphate and are not required for Sec synthesis. Even though its function is unknown, it is essential in fruit flies and mice. Recent convergent observations indicate a role for SPS1 in the control of gene expression to regulate redox homeostasis and pyridoxal phosphate (PLP) synthesis (Manta et al., 2022).

2.1.4. **Selenoprotein W (SelenoW)**

SelenoW contains a thioredoxin-like fold characterized by the presence of a conserved CXXU motif which suggests its oxidoreductase activity. This structural feature and a conserved tGxFEI sequence located in the C-terminal extremity is a signature for a sub-family of four selenoproteins (including SelenoW, SelenoT, SelenoV, and SelenoH), related to thioredoxins and sometimes classified as redoxins (Aachmann et al., 2007; Dikiy et al., 2007).

SelenoW, the smallest selenoprotein of 9 kDa, is localized in the cytosol and found expressed at high levels in mammalian muscles and brain. SelenoW activity remains elusive, however, it is one of the stress-related selenoproteins involved in the protection of cells from oxidative stress-induced cell death in a glutathione-dependent manner. Its expression is highly sensitive to dietary Se **(Gu et al., 2002)**. In addition, SelenoW was shown to be highly expressed in proliferating myoblasts, and enhances skeletal muscle differentiation by inhibiting TAZ binding to 14-3-3 protein **(Jeon et al., 2014)**. SelenoW interacts with the signaling 14-3-3ß protein, which functions in the signaling pathway, through its CXXU catalytic motif, inhibiting 14-3-3ß interaction with other target proteins, such as Rictor, and leading to increased phosphorylation of Akt. It was reported that SelenoW can protect Trx1-deficient cells from etoposide-induced cell death through its interaction with 14-3-3ß (Kang et al., 2021).

2.2. Selenoproteins specific to Eukaryotes

2.2.1. Thioredoxin Reductases (TrxR/TxnRd/TR)

TrxRs are oxidoreductases found as 55–60 kDa homodimers, each subunit contains a flavin adenine dinucleotide (FAD) domain, an NADPH binding domain, and an interface domain. The active site is constituted of two redox motifs: GCUG at the C-terminus and CVNVGC at the N-terminus. Moreover, TrxR exists in different isoforms possessing different functions. TrxR main substrate is Trx, but it can also reduce protein disulfide isomerase (PDI), Grx2, and dehydroascorbate (Hanschmann et al., 2013).

Three TrxRs have been identified in mammals, and all are Sec-containing proteins: cytosolic/nuclear TrxR1, mitochondrial TrxR2, and testicular TrxR3 also called thioredoxin glutathione reductase (TGR) (Reeves & Hoffmann, 2009).

The TrxR1 reduces cytosolic Trx1 in an NADPH-dependent reaction. It can also reduce hydroperoxides, vitamin C, and selenite which are non-disulfide substrates. There are at least six different TrxR1 isoforms in mammals, resulting from alternative splice sites or transcription initiation start sites. TrxR1 plays a role in many physiological processes including, DNA repair, redox homeostasis, regulation of cell signaling, regulation of transcription factors, and apoptosis (Arnér & Holmgren, 2000).

TrxR2 reduces mitochondrial Trx2 and Grx2 and it also has multiple isoforms. TrxR3 is characterized by the presence of a Grx domain at the NH2-terminal part of the protein, suggesting that the TrxR3 displays Grx in addition to the Trx activity, but still, the Grx domain function is unknown (Labunskyy et al., 2014).

Research groups have used TrxR1 as a model to study the advantages of Sec over the Cys-containing selenoproteins homologs. However, the substitution of Sec with Cys in the Human TrxR1 resulted in incomplete inactivation of the enzyme (Gromer et al., 2003; Labunskyy et al., 2014).

2.2.2. Methionine-R-Sulfoxide Reductases (Msrs).

The two amino acids methionine (Met) and Cys are highly susceptible to oxidation which can cause a significant alteration of their structure and disruption of protein functions. Msrs are antioxidant enzymes that catalyze the reduction of methionine sulfoxide to Met implicated in protein repair and protection against oxidative damages. MsrB1 was previously known as selenoprotein R (Kryukov et al., 1999) and selenoprotein X (Lescure et al., 1999b), the R abbreviation referring to its repair activity of the R enantiomer of oxidized methionine residues in proteins. MsrA (methionine-S-sulfoxide reductase) and MsrB1 (methionine-R-sulfoxide reductase 1) both with either Cys- or Sec-containing forms, have different structures and sequences, but each protein function in a similar way on one of the two stereoisomers (Labunskyy et al., 2014). Based on this property, reversible methionine oxidations were employed in the regulation of cellular functions in a stereospecific manner (B. C. Lee et al., 2013).

The Sec-containing MsrB1 protein is the major MsrB in mammals, localized in the cytosol and nucleus, and with high expression in the liver and kidney. There are two Cys-containing homologs, MsrB2 localized in the mitochondria, and MsrB3 is localized in the ER (Tsuji et al., 2022).

MsrA is found in some unicellular eukaryotes with a Sec-containing active site, that displays higher activity than its Cys homologs, suggesting that Sec provides catalytic advantages in these redox-active enzymes (H.-Y. Kim et al., 2006; Labunskyy et al., 2014).

2.2.3. Selenoprotein P (SelenoP)

Selenoprotein P is very unique since it contains multiple Sec residues. *SELENOP* genes are present across metazoan with highly variable numbers of Sec-TGA codons, ranging from a single TGA in certain insects and up to 132 in bivalve mollusks (Baclaocos et al., 2019). SelenoP is found to contain 10 Sec-UGA codons in Humans and two SECIS elements (Labunskyy et al., 2014; Lobanov et al., 2008). In invertebrates, SelenoP is very diversified but it was lost in most nematodes, most insects, tunicates, and Platyhelminthes (Baclaocos et al., 2019).

The N-terminus of SelenoP consists of one Sec residue located in a UXXC redox motif, while the C-terminus is rich in multiple Sec residues. The N-terminal motif was predicted to provide a reductase activity, while the main function of SelenoP is to transport selenium to peripheral tissues, in particular the endocrine glands, testis, and brain (Burk & Hill, 2005; Labunskyy et al., 2014). In mammals, SelenoP is synthesized primarily in the liver and transported to the plasma as it constitutes the major selenoprotein in plasma. Mouse models indicate that SelenoP is not essential for life, as supplemental Se supply was capable of preventing the development of severe symptoms, particularly in the brain. However, knockout mice died under limiting Se supply, arguing for an essential role of SelenoP in Se deficiency and its role as a Se reservoir utilized in Se deprivation time (Schomburg, 2022; Schweizer et al., 2022). In line with the importance of SelenoP in Se transport and selenoproteins synthesis, a recent report highlighted the central role of SelenoP to restore neurogenesis and reverse cognitive decline in aging in a model of hippocampal injury. Indeed, using *Selenop* -/- mice, it was shown that SelenoP is required for the exercise-induced activation of quiescent hippocampal neural precursor cells (NPC) and their subsequent recruitment into the neurogenic differentiation in adult (Leiter et al., 2022).

2.3. Endoplasmic reticulum resident selenoproteins

Seven selenoproteins are localized in the endoplasmic reticulum (ER), including SelenoS, SelenoK, SelenoM, SelenoN, SelenoF, SelenoT, and DIO2. An illustration of ER-resident selenoprotein is displayed in (Figure 11).

2.3.1. Selenoprotein T (SelenoT).

SelenoT is another small selenoprotein, a member of the redoxin sub-family, containing a thioredoxin-like fold and a characteristic conserved CXXU motif, conferring an oxidoreductase activity (Dikiy et al., 2007). SelenoT also contains two hydrophobic domains that allow its insertion in the lipid membrane in a hairpin-like structure, the catalytic motif residing within the luminal part of the ER. SelenoT is localized in the ER and Golgi, where it interacts with a subunit of the oligosaccharyltransferase (OST) complex and controls UPR signaling (Hamieh et al., 2017).

SelenoT is most conserved among selenoproteins, with Cys- or Sec-homologs identified in numerous eukaryotes, including plants, metazoan and unicellular (Pothion et al., 2020). In addition, together with Gpx4 and TrxR1, it is one of the most preserved selenoproteins in case of selenium restriction, demonstrating its high priority in the hierarchy of selenoprotein production. In agreement with its essential nature, SelenoT, similarly to TrxR, was shown not to tolerate loss-of-function mutations both in Humans and mice (Tanguy et al., 2011).

SelenoT is highly expressed during embryonic development in rats, fish, chicken, and *C elegans*. In the nematode and chicken, SelenoT persists in adult tissues. In rats, SelenoT is abundantly expressed during central nervous system development in specific neuronal progenitors, and its expression decreases after birth (Tanguy et al., 2011). In adult rodents, SelenoT only remains abundantly expressed in endocrines and neuroendocrine tissues such as the pituitary gland, pancreas, testis, thyroid, brain, thymus, and kidney (Dikiy et al., 2007; Prevost et al., 2013; Tanguy et al., 2011). Two *SELENOT* knockout mice were generated: In the first one, the Cre-Lox invalidation of the gene led to early embryonic lethality (Boukhzar et al., 2016). In the

second (K. Li et al., 2021), the partial deletion of exon 2 preserving the redox center, was viable although leading to reduced size, infertility, and decreased glycemia, in agreement with previous results from (Prevost et al., 2013) demonstrating that SelenoT is implicated in the regulation of pancreatic β-cell function and glucose homeostasis. Moreover, conditional invalidation of the *SELENOT* gene in mice brains led to reducing cerebral structures linked to increased ROS levels and apoptosis in neuroblasts and hyperactivity in adults (Castex et al., 2016). Accordingly, SelenoT was shown to be overexpressed in a calcium-dependent manner during neuroendocrine differentiation of the PC12 cell model (Abid et al., 2019) (Grumalo et al., 2008, 2013). SelenoT expression is stimulated by the neuropeptide PACAP and cyclic-AMP, inducing increased Ca^{2+} concentration in the cell. SelenoT expression is also stimulated by treatments perturbing the ER homeostasis and ROS level, such as tunicamycin or thapsigargin (Hamieh et al., 2017; Pothion et al., 2020).

2.3.2. Selenoprotein (SelenoF, 15-kDa, or Sep15) and Selenoprotein M (SelenoM)

SelenoM and its distant homolog SelenoF are two ER resident proteins. SelenoM was identified by bioinformatics approaches, while SelenoF was identified by experimental means. Both share 31% sequence identity and have a common thioredoxin-like fold and an N-terminal signal peptide. The SelenoM C-terminus encodes a retention signal in the ER lumen, while SelenoF doesn't. The N-terminal part of SelenoF contains an extension of a cysteine-rich domain, six cysteine residues, identified as a protein-protein interaction domain necessary for the binding of SelenoF with its partner, the UDP-glucose:glycoprotein glucosyltransferase (UGGT). UGGT is an ER-resident chaperone involved in the regulation of calnexin (CNX), a calcium-binding protein involved in the quality control pathway within the ER. UGGT also acts as a sensor for misfolded glycoproteins (Labunskyy et al., 2005, 2014). The SelenoF active site is composed of CXU, while SelenoM active site is composed of CXXU. NMR spectroscopy study showed that both proteins contain the α/β-fold specific for all thioredoxin-like oxidoreductases where they gained the oxidoreductase function and both have disulfide exchange activity (Ferguson et al., 2006).

The expression of both proteins is well distributed among mammalian tissues. SelenoF expression is sensitive to dietary Se and is highly expressed in the prostate, liver, kidney, and

testis, whereas SelenoM is predominantly expressed in the brain, which suggested its role in neuroprotection. SelenoM also showed a role in Ca^{2+} regulation release from the ER in neuronal cultures treated with H_2O_2 (Reeves et al., 2010). Several studies on different cellular and animal models have been conducted to investigate the neuroprotective role of SelenoM in the physiopathology of Alzeimer's disease. It was reported that differential SelenoM expression was linked to the inhibition of Amyloid-beta (Aß) protein aggregation and cytotoxicity (Gong et al., 2021; Pitts et al., 2013; Reeves et al., 2010). This activity still needs to be validated *in vivo*, and the mechanism linked to redox-regulated Ca^{2+} homeostasis remains to be characterized. Moreover, SelenoM activity was also shown to play an important role in energy metabolism controlled by leptin, also linked to calcium signaling in the cell (Chen et al., 2013; Du et al., 2013; Yim et al., 2009).

2.3.3. Selenoproteins K and S (SelenoK, SelenoS)

SelenoK and SelenoS are both ER-resident transmembrane proteins, widely spread in eukaryotes. Even though they have no sequence similarity, both proteins belong to one sub-family characterized by a single N-terminal transmembrane domain and a glycine/proline-rich region at the C-terminal. The catalytic site containing the Sec is located in the C-terminal end pointing into the cytosol. Cys-containing SelenoK/SelenoS-like proteins were identified in other species including fungi, insects, and plants (Shchedrina et al., 2011). SelenoS contains an additional coiled-coil cytosolic domain that is not present in SelenoK, this domain is suggested to be responsible for the interaction of SelenoS with other proteins or to form oligomers (Ye et al., 2004).

As discussed above, SelenoF and SelenoM function in protein folding in the ER, while, SelenoK and SelenoS have been predicted to have a role in the elimination of misfolded protein in the Derlin-dependent ERAD-associated protein degradation (ERAD) pathway. Both proteins were found associated with the retrotranslocation-protein complex in the ER, but with different affinities, indicative of substrate specificity for each protein. In addition, the expression of both proteins is upregulated by the accumulation of misfolded proteins in the ER (Shchedrina et al., 2011).

Moreover, SelenoK expression is sensitive to Se dietary intake. SelK was shown to be implicated in the protein's palmitoylation reaction of CD36 and inositol-1,4,5-triphosphate receptor (IP3R), a Ca^{2+} channel protein in the ER membrane. This reaction occurs through the binding of SelenoK to the acyltransferase, DHHC6, in which SelenoK acts as a cofactor in the reaction (Fredericks et al., 2017; Fredericks & Hoffmann, 2015).

SelenoS is not yet proven to have a role in Ca^{2+} regulation, but it may indirectly influence Ca^{2+} signaling by maintaining ER homeostasis through its role in misfolded protein degradation (Pitts & Hoffmann, 2018).

2.3.4 Selenoprotein N (SelenoN)

SelenoN is very unique among all ER resident proteins in that it is the only selenoprotein that has been identified with a calcium EF-hand binding domain. SelenoN plays an important role in regulating Ca^{2+} homeostasis in the ER/SR (Sarcoplasmic Reticulum) and in defending the cell from ER stress. In addition, SelenoN deficiency in mice and Human were linked to congenital myopathy under a group of disorder known as SEPN1-related myopathies. More details about SelenoN are presented in the next section.

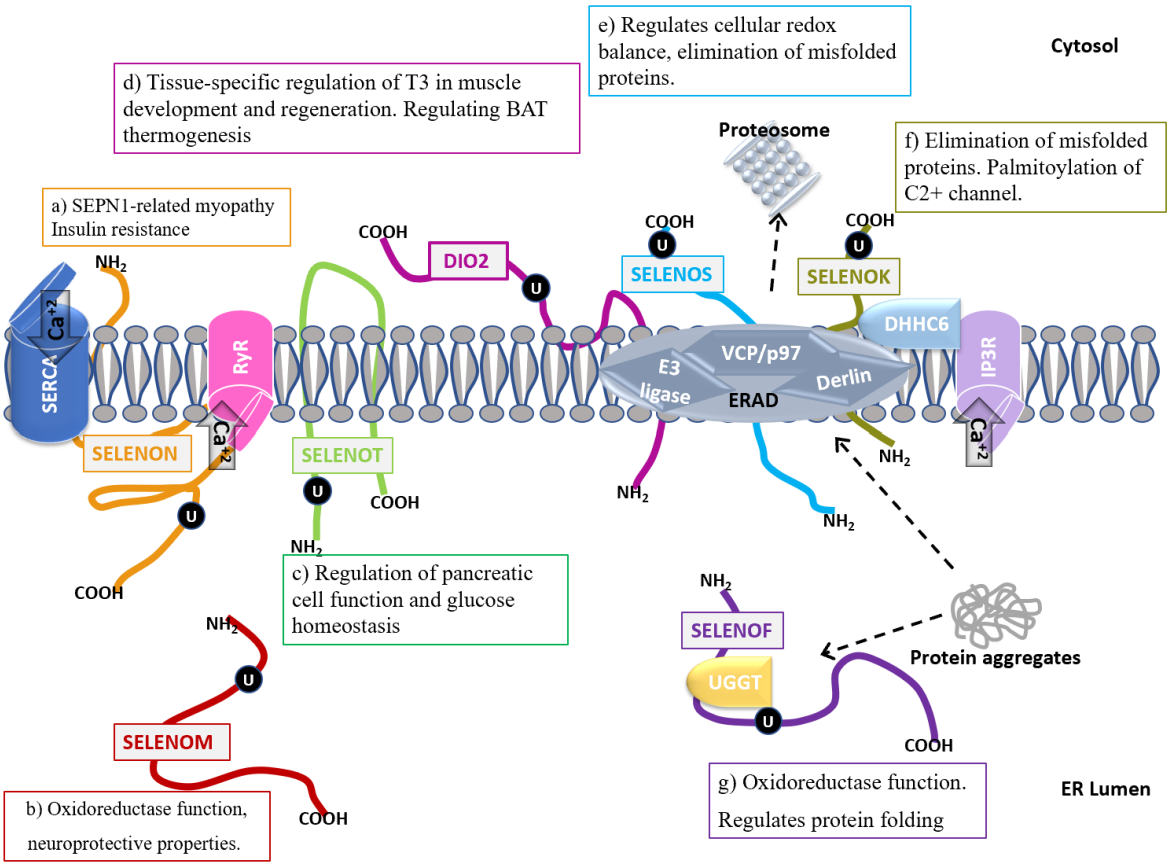


Figure 11. Schematic representation of Endoplasmic Reticulum-resident selenoproteins. The Sec residues along with transmembrane domains are shown for each selenoprotein. Proteins involved in calcium homeostasis, ERAD pathway, and protein folding are shown as well. A summary of functions is indicated for each ER resident selenoprotein.

A) upon Ca^{2+} depletion in the ER, SelenoN is activated by interaction with the SERCA pump. SelenoN deficiency results in SEPN1-related myopathy accompanied by insulin resistance. **B)** SelenoM showed to play a protective role in preventing neurodegeneration in the brain. Deletion of *SELENOM* leads to obesity. **C)** SelenoT deletion was found associated with elevated levels of ROS in the brain and leads to growth retardation and neurodegeneration. SelenoT regulates pancreatic cell function and glucose homeostasis and its inhibition cause ER stress in endocrine cells. **D)** DIO2 regulates T3 production in muscles and is very important in muscle development, differentiation, and regeneration. DIO2 is also found to regulate brown adipose tissue (BAT) thermogenesis. **E)** SelenoS plays a role in the elimination of misfolded protein in the Derlin-dependent ERAD-associated protein degradation (ERAD) pathway. **F)** SelenoK binds to DHHC6 and is implicated in the palmitoylation reaction of CD36 and IP3R. SelenoK also plays a role in the removal of misfolded proteins. **G)** SelenoF binds to UGGT, a sensor for misfolded glycoproteins, and regulates protein folding. Created by Microsoft PowerPoint, and adapted from (Pitts & Hoffmann, 2018) (Jehan et al., 2022). Abbreviations; ER, endoplasmic reticulum; U, Sec residues; IP3R, inositol-1,4,5-triphosphate receptor; UGGT, UDP-glucose-glycoprotein glucosyltransferase

2.4. Selenoproteins in Prokaryotes.

Most of the prokaryotic selenoproteins were identified through bioinformatics analyses searching for in-frame UGA codons and the SECIS element. However, few of them were experimentally validated (Zhang et al., 2022). Based on the computational analysis 25 selenoproteins were identified in prokaryotes. The size of prokaryotic selenoproteome varies among different species, ranging from single selenoprotein in *Mycobacterium* species, formate dehydrogenase, to 39 selenoproteins in *Syntrophobacter fumaroxidans* (Hatfield et al., 2016). Here we will go through two of the most popular selenoproteins in prokaryotes.

2.4.1. Formate Dehydrogenase

$HCOO^- + NAD^+ \rightarrow CO2 \uparrow + NADH.$

Formate dehydrogenase (FDH) is by far the most represented among prokaryotic selenoproteins, found as a Sec-containing protein in about 75% of the Sec-utilizing bacteria, whereas SelD, the selenophosphate synthetase, an important actor of the selenium insertion machinery, was identified as a selenoprotein in only 35% of the bacteria of this group (Zhang et al., 2022).

FDHs are a group of enzymes that belong to the oxidoreductases family, catalyze the reversible oxidation of formate to carbon dioxide (CO_2), and function in formate metabolism. FDH donates the electrons to a second substrate, such as NAD+/ NADP+, cytochrome, quinones, or ferredoxin. There are two main classes of FDHs, metal-independent FDHs, and metal-containing FDHs. The molybdenum or tungsten-containing enzyme families constitute the metal-containing FDH class found mainly in bacteria and archaea. FDHs contain Fe/S clusters and molybdopterin incorporates Sec through the utilization of the Sec-cofactor trait (Hatfield et al., 2016; Yang et al., 2022). On the other hand, metal-independent FDH class enzymes are mainly found in aerobic bacteria, yeasts, fungi, and plants. The metal-independent FDH mediates direct electron transfer to the electron acceptor, while in the metal-containing class the metal center in the active site act as a mediator for electron transfer (Maia et al., 2015). FDHs are utilized in industry for the regeneration of NADH/NADPH since they are very demanding in the industrial application as a cofactor in many oxidoreductase reactions. Even though the regeneration of both cofactors can be done through other enzymes like glucose dehydrogenase (GDH) and alcohol dehydrogenase (ADH), these enzymes result in the accumulation of byproducts, that are not costly favorable, and can be avoided by utilizing FDHs instead (Alpdagtas & Binay, 2021; Yang et al., 2022).

2.4.2. Hydrogenase

Microbial hydrogen metabolism is a very ancient process before oxygen evolved in the atmosphere. H_2 is one of the trace gases present in the atmosphere. Many organisms such as fermentative organisms, photosynthetic prokaryotes, aerobes, anaerobes, autotrophs, heterotrophs, and some lower eukaryotes use H_2 as an energy source, and the reaction is catalyzed by hydrogenase enzymes (Morra, 2022). Hydrogenases are metalloenzymes that catalyze the reversible oxidoreductase reaction of H_2 [H2↔2H++2e-] and are essential to microbial hydrogen metabolism. Most of the identified hydrogenases are iron-sulfur proteins and they are classified based on the metal atom bound at the active site: (**i**) two iron[Fe-Fe]-hydrogenases, (**ii**) Nickel-iron [Ni-Fe]-hydrogenases, (**iii**) and [Fe]-hydrogenases (known as Hmd) (Vignais & Billoud, 2007). The oxidoreductase reaction of H_2 requires the presence of electron donors and acceptors, NAD, cytochromes, coenzyme F420, or ferredoxins

[Ni-Fe]-hydrogenase is a very common selenoprotein in prokaryotes (Hatfield et al., 2016). This type of hydrogenase contains three Fe-S clusters in the small subunit and a Ni-Fe-containing active site in the large subunit that includes a Sec ligand. The Ni-Fe ions are coordinated by either four Cys residues or three Cys along with a Sec residue, where two Cys bind to the Fe ion (Garcin et al., 1999; Hatfield et al., 2016).

[Fe-Fe]-hydrogenases are found in the genome of anaerobic bacteria. It is also the only type of hydrogenase to have been found in eukaryotes such as in chloroplasts of green algae, and hydrogenosome organelles of parasitic protozoans or anaerobic fungi. However, archaea are free of this type of hydrogenase. The enzymatic ability in [Fe-Fe]-hydrogenases appeared in the formation of the Fe-S cluster between the [Fe-Fe]subcluster and the [4Fe-4S]-subcluster via a conserved Cys residue. [Fe-Fe]-hydrogenases play a role in physiological functions such as H_2 uptake (oxidation), H_2 sensing, and CO_2 fixation. [Fe-Fe]-hydrogenases catalyze the reversible oxidation of H_2 very efficiently (Morra, 2022).

The [Fe]-hydrogenases function as an H2-forming methylene tetrahydromethanopterin dehydrogenase (Hmd), in methanogenic archaea (Vignais & Billoud, 2007).

III. Selenoprotein N (SelenoN)

1. Identification

The *SELENON* gene (also called SELN or SEPN1) was first identified using computational approaches based on the identification of the SECIS structure and next validated *in vitro* (Lescure et al., 1999a). In Humans, the *SELENON* gene is located on chromosome 1 position p35-p36 and is composed of 13 exons (Moghadaszadeh et al., 2001) and reviewed in (Castets et al., 2012). In all animals, the *SELENON* gene is transcribed as one transcript, except in primates, including Humans, in which two *SELENON* isoforms were characterized. Isoform 1 corresponding to the longer transcript contains two UGA-Sec codons, while in the shorter isoform 2 exon 3, which corresponds to a primate-specific Alu sequence, is spliced-out and matches the non-primate transcript, encoding only one in-frame Sec residue. Both transcripts were found expressed in skeletal muscle, brain, lung, and placenta, with isoform 2 being the predominant transcript and the only one to be translated (Petit, 2003). A recent study showed that the primate-specific long non-coding transcript regulates a sophisticated mechanism controlling *SELENON* expression both at the post-transcriptional and translational levels during muscle differentiation (Chernorudskiy et al., 2020; Noda et al., 2022).

SelenoN has a calculated mass of 65 kDa in size but migrated as a 70 kDa protein on SDS-PAGE due to its glycosylation (Rederstorff et al., 2011). It is a ubiquitous endoplasmic/sarcoplasmic reticulum (ER/SR) resident glycoprotein that contains, a type II transmembrane domain at the N-terminus with a stretch of hydrophobic residues, and about ten amino acids exposed to the cytoplasm. SelenoN sequence inspection identified a calcium-binding EF-hand domain and a thioredoxin-like domain, both oriented inside the ER lumen (Figure 12). A structural functional study of SelenoN topology in the membrane and of the calcium-binding domain demonstrated that despite its low similarity with the EF-hand consensus sequence, this domain still binds calcium with an affinity constant in the range of the calcium concentration in the ER, making this protein a potent sensor of calcium in this compartment. Importantly, it has been shown that calcium sensing regulates SelenoN redox reactivity and its interaction with different cellular substrates (Chernorudskiy et al., 2020).

The thioredoxin-like domain contains the catalytic site for SelenoN SCUG, predicted based on the presence of the Sec residue. Its resemblance with the redox active motif of TrxR (GCUG) implicates an oxidoreductase activity for SelenoN (Lescure et al., 2018; Pitts & Hoffmann, 2018).

Aside from the EF-hand and thioredoxin-like domains, SelenoN presents no homology to any protein of known function and no enzymatic activity has been attributed to SelenoN yet.

In the next section, a summary is given on SelenoN relation to muscle development, calcium homeostasis, and ER oxidative stress.

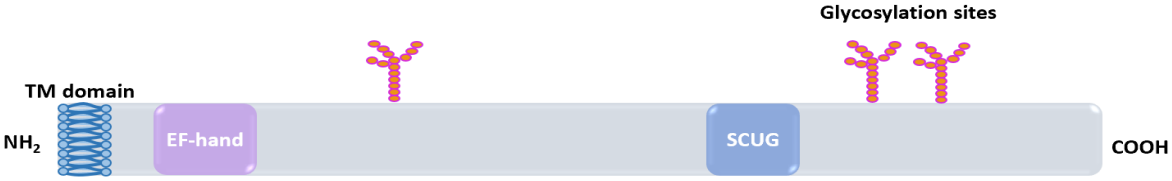


Figure 12. Organization of Human SelenoN. A functional transmembrane domain (TM) in blue. EF-hand Ca^{2+}-binding domain in purple, and three glycosylation sites in pink. The SCUG sequence corresponds to the catalytic site of the protein. Created by Microsoft PowerPoint and adapted from (Chernorudskiy et al., 2020)

2. Selenoprotein N Related Myopathies.

Mutations in the Human *SELENON* gene were linked to a group of recessive congenital muscular dystrophies (CMD) collectively known as *SELENON* or *SEPN1*-related myopathy (SELENON-RM*)*, which consists of four autosomal recessive inherited muscle disorders; rigid spine muscular dystrophy (RSMD1), the classical form of multiminicore disease (MmD) (Ferreiro et al., 2002), desmin related myopathy with Mallory-body like inclusions (MB-DRM) (Ferreiro et al., 2004), and congenital fiber-type disproportion (CFTD) (Clarke et al., 2006; Moghadaszadeh et al., 2001). All of them are rare neuromuscular disorders characterized by generalized muscle weakness, spinal rigidity, spine deformity and neck weakness, and respiratory failure developing in severe cases or late adulthood (Silwal et al., 2020; Villar-Quiles et al., 2020). Interestingly, previous observations described a relation between selenium deficiency and the development of muscular dystrophy in livestock, called nutritional muscular dystrophy (Lescure et al., 2016).

Many mutations scattered all over the *SELENON* sequence have been identified, with two major hot spots at the beginning of the coding sequence and flanking the codon coding for the Sec residue, reviewed in (Castets et al., 2012) and most recent reviews (Fan et al., 2022; Villar-Quiles et al., 2020). In addition, a single homozygous point mutation was found in the 3' UTR SECIS sequence in patients with RSMD1; this mutation interferes with the Sec insertion due to the inability of SBP2 to bind the SECIS (Allamand et al., 2006). Moreover, a muscle biopsy from patients with a missense point mutation in the *SELENON* Sec Redefinition Element (SRE) showed very neglectable levels of SelenoN expression. This was interpreted by a reduction in the efficacy of Sec insertion during mRNA translation (Maiti et al., 2009).

Intriguingly, SelenoN is not a muscle-specific gene but is expressed ubiquitously in all tissues analyzed so far, also at different levels (Castets et al., 2009; Lescure et al., 1999a).

3. Animal models for SELENON-Related Myopathies.

Since SelenoN loss-of-function in Humans causes SELENON-RM, most studies conducted focused on the characterization of SelenoN's role in muscle tissue physiology in different animal models. In studies using the zebrafish model, SelenoN was found to be highly expressed in somites, differentiated structures in embryos, during early embryogenesis then decreasing and remaining low in adult muscle structures (Deniziak et al., 2007; Jurynec et al., 2008). Accordingly, SelenoN expression was shown to be regulated during Human myoblast differentiation: its expression picks at an early stage, then decreases with further myotubes differentiation (Petit, 2003).

In another study, a zebrafish model was depleted for SelenoN expression during embryogenesis by introducing antisense morpholinos. The phenotype of the zebrafish embryos showed a disorganization in the somites, leading to an alteration in muscle architecture and malformation of the tail (Deniziak et al., 2007; Jurynec et al., 2008). These studies imply the importance of SelenoN in early muscle development and differentiation in zebrafish (Figure 13a).

In mouse embryos as well, SelenoN appeared to be abundantly expressed in the muscle precursors during development (Castets et al., 2009). However, in *SelenoN-/-* mice in opposite to what was observed in zebrafish, no change in phenotype or obvious muscle defects were observed, and the knock-out mice looked similar to the wildtype (Figure 13b). Regarding functional analysis, both knock-out and wildtype were exposed to regular exercise, and at the same time, a contractility test was performed, but no remarkable observation of functional change (Rederstorff et al., 2011). On the other hand, under intensive exercise that is depicted in repeated forced swimming tests (FST) the *SelenoN-/-* mice developed spin rigidity and reduced mobility in comparison to wildtype. Persistent FST resulted in the development of kyphosis, a severe spine deformity (Rederstorff et al., 2011) (Figure 13c and d). Moreover, muscle tissues of *SelenoN-/-* mice were treated with cardiotoxins to test their ability to regenerate, and the results obtained showed the inability of muscle regeneration after repeated injuries. This incapacity was shown to be due to the loss of muscle progenitor satellite cells responsible for muscle regeneration (Castets et al., 2011). Thus the obtained results imply the role of SelenoN in muscle tissue regeneration and satellite cell maintenance. Another study using SelenoN-depleted mice reported abnormal lung development in addition to the muscle phenotype, indicative of a

possible contribution to the restricted respiration phenotype described in patients (Moghadaszadeh et al., 2013). A recent study comparing wildtype and *SelelnoN-/-* mice showed modified glucose metabolism and clear glycogen-contain depletion in liver and muscle tissues (Figure 14) (Filipe et al., 2021).

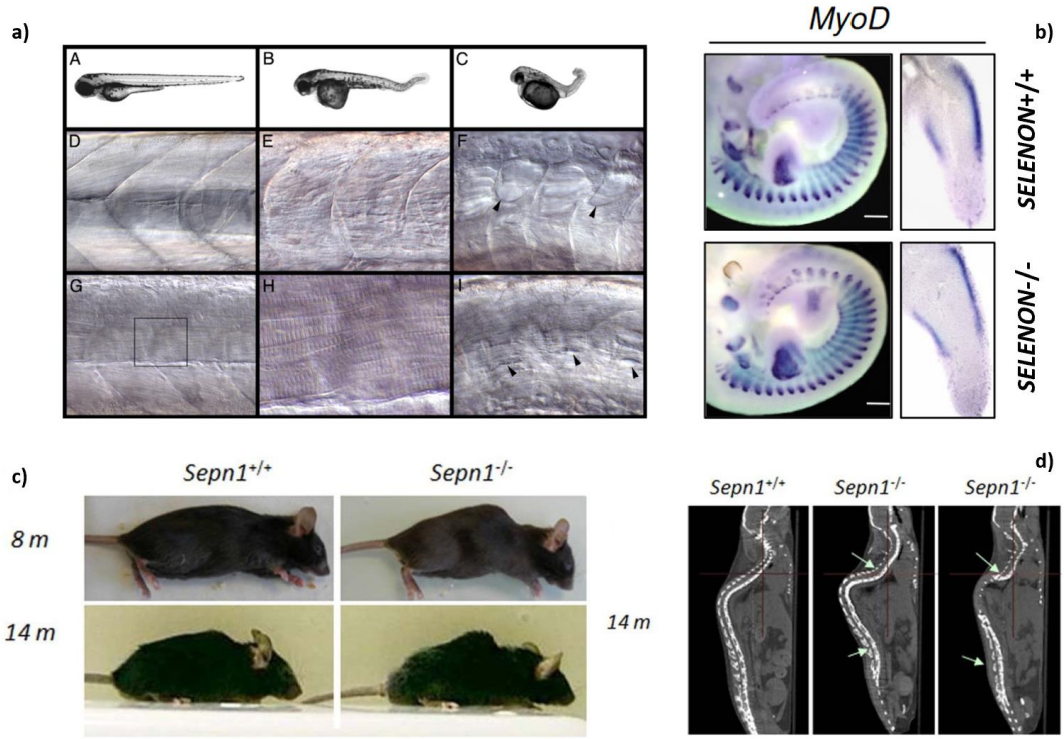


Figure 13. a) Muscles Phenotype of SelenoN knockdown in zebrafish. SelenoN depleted zebrafish displayed strong morphological alterations, tail malformation, and alteration of muscle architecture. Lateral views of 48 hpf wild-type control (A), embryo injected with 250 µM (B) and 500 µM (C) of Morpholino (Deniziak et al., 2007). **b) Somitogenesis in wildtype and *SELENON*−/− in mice embryos.** Vibratome sections whole mount in situ hybridization against *MyoD* showed no difference in somite organization and size between wildtype and knock-out models (Castets et al., 2009). **c) and d) *SelenoN*−/− mice**. Under normal housing, no muscular defects were observed. Following an extensive FST, *SelenoN*−/− adult and young mice developed kyphosis, a spine deformity (Rederstorff et al., 2011).

Abbreviation: *MyoD*, myogenic determination factors; hpf, hours post-fertilization; FST, forced swimming test; 8m, 8 months; 14m, 14 months; Sepn, SelenoN.

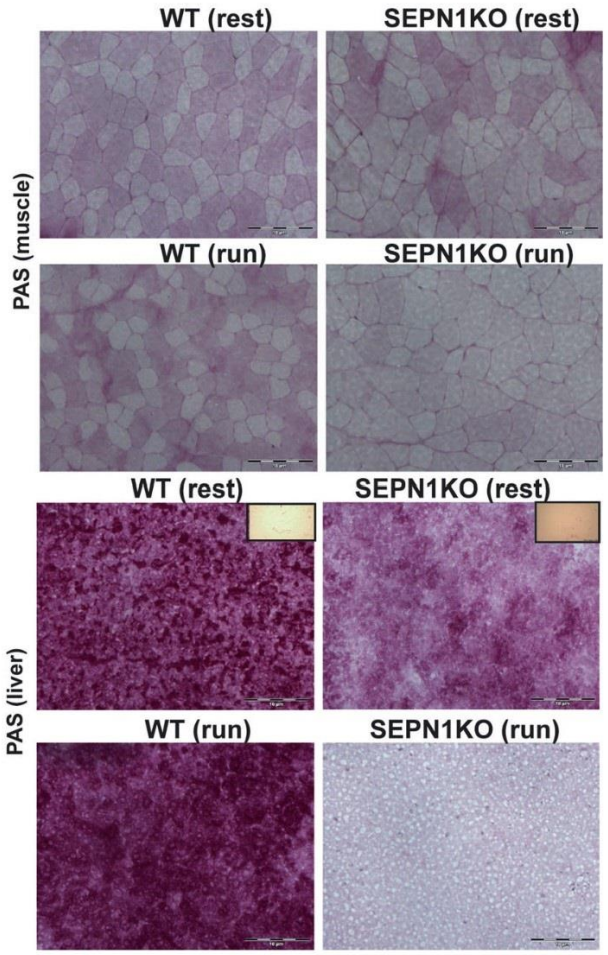


Figure 14. *SelenoN–/– mice* muscle and liver histology. Muscle tissues from the liver and gastrocnemius sections are stained with periodic acid-Schiff (PAS) for glycogen staining. The structure of tissues from the liver and muscles are shown at rest and after running for both knock-out and wildtype, showing a clear depletion of glycogen in exercised knock-out animals in both tissues (Filipe et al., 2021).

4. SelenoN activity in the Endoplasmic Reticulum compartment and the cell

4.1. Endoplasmic reticulum stress and calcium homeostasis

SelenoN is a resident protein of the endoplasmic reticulum (ER) and sarcoplasmic reticulum (SR), compartments that contribute to a variety of critical cellular functions, including protein folding, assembly, and transport, calcium storage and signaling, lipid synthesis and storage, and carbohydrate metabolism. Alteration of ER homeostasis lead to ER stress and that can be due to multiple factors, including; accumulation of misfolded proteins, oxidative stress, and disturbance in Ca^{2+} homeostasis. Therefore, ER role is very important for many biochemical processes required for cellular functions and cell survival. Here we will have a brief summary about ER stress causative factors and response to these factors before we talk about the role of SelenoN as an ER resident protein.

4.1.1. Calcium homeostasis

The ER is the primary reservoir of Ca^{2+} ions in micromolar concentrations. The intracellular Ca^{2+} levels are regulated by a variety of elements including channels, pumps, and Ca^{2+} binding proteins to maintain resting cytosolic concentrations in nanomolar levels (Jousset et al., 2007; Pitts & Hoffmann, 2018). As one key regulator of cellular Ca^{2+} homeostasis the sarcoendoplasmic reticulum calcium (SERCA) pump strictly controls the pumping of Ca^{2+} from the cytosol back to the SR/ER. The SERCA pump belongs to the family of P-type ATPases. There are three differentially expressed genes SERCA 1, SERCA 2, and SERCA 3 encoding seven isoforms of the SERCA pump (SERCA 1a/1b, SERCA 2a/2b, and SERCA 3a/3b/3c). SERCA1a and 1b are skeletal muscles isoforms in adults and infants and are expressed in fast-twitch muscle fibers; SERCA 2a is primarily expressed in slow-twitch skeletal muscle and cardiac muscle, SERCA 2b is a ubiquitous isoform expressed in all cell types at a low abundance; lastly, SERCA 3 is very rare in muscle tissues, but expressed at low levels in non-muscle tissues (Marino et al., 2015, p. 1; H. Xu & Van Remmen, 2021).

In addition, two classes of intracellular Ca^{2+} channels are responsible for Ca^{2+} release from the ER and SR to the cytoplasm upon activation of certain stimuli: the inositol 1,4,5-trisphosphate receptors (IP3Rs) and the ryanodine receptors (RyRs). For example, inositol 1,4,5-trisphosphate (IP3) produced by immune cells binds to the IP3Rs inducing Ca^{2+} ions release from the ER (Pitts

& Hoffmann, 2018). In mammals, there are three differentially expressed RyRs proteins. RyR1 and RyR2 are expressed predominantly in skeletal and cardiac muscle respectively, whereas RyR3 is expressed at low levels in other tissues. The primary role of the RyR channels is shown in the excitation-contraction coupling (E-C coupling) in both skeletal and cardiac muscle (Jurynec et al., 2008; H. H. Wu et al., 2011). In Humans, mutations in the *ryr1* and *ryr2* genes cause skeletal myopathies and cardiac disease respectively (Betzenhauser & Marks, 2010).

Furthermore, ER Ca^{2+} levels are monitored by the sensor stromal interaction molecule 1 (STIM1). An important process known as Store Operated Ca^{2+} Entry (SOCE) is a process in which Ca^{2+} ions are transported back to the ER in response to ER-Ca^{2+} depletion, this is achieved via a complex formed between STIM1 and the plasma membrane channels, Ca^{2+} channel protein (Orai1) or transient receptor potential canonical 1 (TRPC1). SOCE is crucial for cell proliferation, migration, and other functions (Ambudkar et al., 2017). Furthermore, an alteration in Ca^{2+} homeostasis was described in tumor cells and referred to as Ca^{2+} remodeling (Villalobos et al., 2016). Several selenoproteins, including SelenoN, SelenoT, and SelenoW, are believed to contribute to SOCE and some may also be involved in Ca^{2+} remodeling in cancer. A simple illustration of the ER with the different calcium pumps and channels is illustrated in (Figure 15).

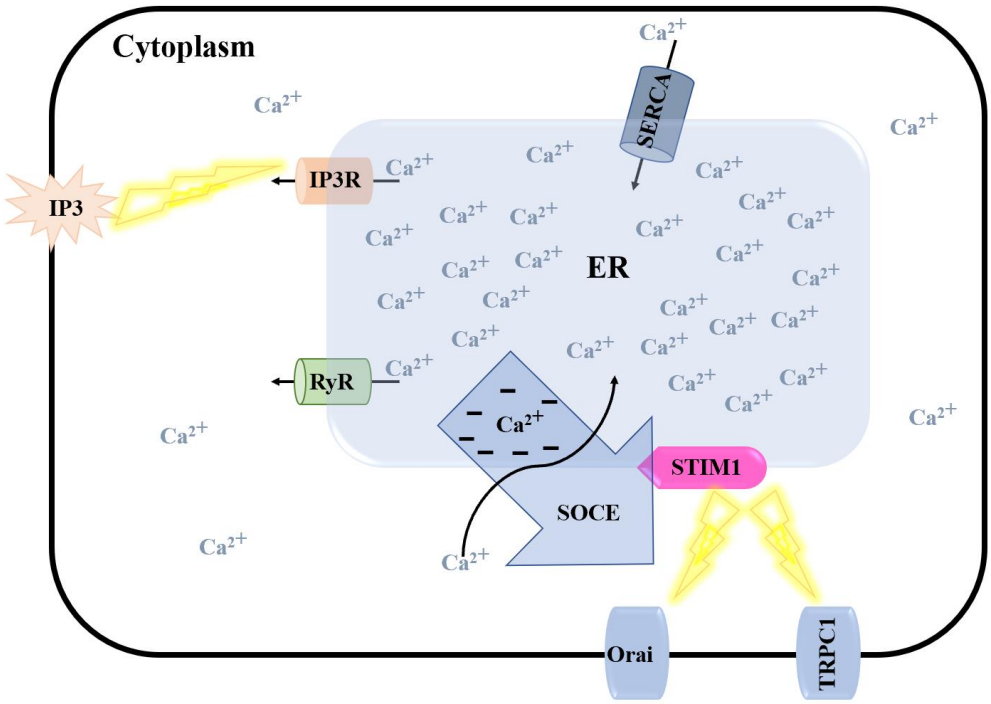


Figure 15. Endoplasmic reticulum calcium channels. ER pumps and channels that control the flux of calcium ions into and out of the ER are illustrated in the scheme. The Store operated Ca^{2+} entry (SOCE) is illustrated as well. Created by Microsoft PowerPoint.

4.1.2. ER stress and UPR signaling

The ER is responsible for the folding and assembly of nascent proteins underway for secretion or anchoring to the cell membrane. These proteins are tagged by a signal peptide at the N-terminal to pass through the Sec61 translocon and enter ER. Once in the ER, different chaperones and enzymes are in charge of maintaining proteins folding and integrity. These include glucose-regulated protein (GRP) chaperones, glucosidases, transferases, oxidizing enzymes, protein disulfide isomerases (PDIs), and the Ca^{2+}-regulated chaperones, calnexin, and calreticulin, reviewed in (Pitts & Hoffmann, 2018).

Under ER stress, such as the accumulation of unfolded protein or Ca^{2+} depletion, the unfolded-protein response (UPR) pathway is triggered. The UPR upregulated the expression of chaperones involved in protein folding, inhibition of global translation, and elimination of misfolded proteins. ER stress is sensed by three enzymes: protein kinase RNA (PKR)-like ER kinase (PERK), inositol-requiring kinase 1-alpha (IRE1α), and activation of transcription factor 6 (ATF6). All these enzymes are inactive under normal conditions while they remain bound to the chaperone GRP78, also known as BiP. Even more, The GRP78 is also found bound to Sec61 translocon under resting conditions and preventing Ca^{2+} leakage from the ER. However, the accumulation of misfolded proteins competes with GRP78 and later dissociates from the sensors enzymes and the Sec61 translocon, resulting in activation of the UPR and Ca^{2+} loss from the ER (Hammadi et al., 2013; Yeganeh et al., 2015) (Almanza et al., 2019).

The UPR can be activated by either one of two pathways; adaptive UPR which can restore ER equilibrium through the ER-associated degradation (ERAD) pathway, or apoptotic UPR if the damage is too severe through the activation of apoptotic cell signaling pathways (Walter & Ron, 2011; Yeganeh et al., 2015). Moreover, upon severe ER stress, mitochondria physiology is severely affected due to Ca^{2+} ER to mitochondria influx through the mitochondria-associated membranes (MAMs), causing the mitochondria to release cytochrome c, an important signal in cellular apoptosis, reviewed in (Pitts & Hoffmann, 2018). Many chronic diseases, including type 2 diabetes, cardiovascular disease, and multiple neurodegenerative disorders have been linked to chronic ER stress. A simple illustration of ER stress and UPR signaling pathway is displayed in (Figure 16).

UPR also activates the transcription of ER-oxidoreductin 1 (ERO1), an ER oxidoreductase enzyme. ERO1 oxidative activity leads to the production of H_2O_2, which results in increased levels of ROS. Accumulation of H_2O_2 in the ER lumen leads to the oxidation of free thiols -SH through H_2O_2-mediated oxidation, a process known as ER oxidative poised.

In fact, Ca^{2+} homeostasis, protein folding, and oxidative stress are related and intricated incidences. The high Ca^{2+} levels could be responsible for the oxidizing state of the ER compared to the reduced state of the cytosol. A decline of Ca^{2+} levels in the ER leads to a more reducing state, this reduction results in the activation of the SERCA2B pump by reducing sensitive disulfide bonds and eventually pumping of Ca^{2+} into the ER, while oxidation of cysteine residues on RyRs promotes ER Ca^{2+} leakage (Andersson et al., 2011; Y. Li & Camacho, 2004).

4.1.3. Endoplasmic reticulum-associated protein degradation (ERAD)

The proper folding of proteins is very critical to ER homeostasis. Accumulation of misfolded proteins in the ER leads to irreversible consequences such as ER stress, and loss of proteins functions, which has been linked to several diseases. The ERAD pathway is a quality control mechanism in the ER for the integrity of protein folding. ERAD is a complex mechanism composed of multiple steps from the recognition of misfolded proteins, ubiquitination, retrotranslocation, and finally cytoplasmic degradation by the proteosome. There are a variety of key factors involved in ERAD (Cyr & Hebert, 2009; Tepedelen et al., 2019; Walter & Ron, 2011). The ERAD pathway can be divided into three main steps:

- **Recognition of misfolded protein**

Different factors can result in improper protein folding including, genetic mutations, errors in transcription and translation, toxic compounds, hypoxia, and ROS. The ER can recognize misfolded proteins through either their unpaired Cys residues or exposed hydrophobic residues. Approximately 30% of newly synthesized proteins enter the ER for assembly through the Sec61 translocon. The ER oxidizing environment and high calcium levels are a very convenient medium for protein folding and co/post-translational modifications such as N-glycosylation, disulfide bond formation, or glycosylphosphatidylinositol (GPI) anchoring at the C-terminus of

proteins. Chaperones, such as calnexin/calreticulin (CNX/CRT), are responsible for restoring the native conformation of the immature glycoproteins through deglycosylation/reglycosylation cycles by UDP-glucose-glycoprotein glucosyltransferase (UGGT). Glycoproteins targeted for degradation are recognized by ER mannosidase I which can remove mannose residue from glycoproteins, to be ready for degradation by the EDEM (ER degradation-enhancing α-mannosidase-like protein) family. Other chaperones such as HSP70-like chaperone, and enzymes like protein disulfide isomerases (PDI) are also dedicated to aiding in protein folding through a disulfide-bond exchange.

- **Ubiquitination**

Ubiquitination is a very important step in the ERAD pathways. The misfolded substrates are marked with ubiquitin for proteasomal degradation. Three enzymes are involved in the ubiquitination process: E1ubiquitin activating enzyme, E2-ubiquitin conjugating enzyme, and E3-ubiquitin ligase. Hsp70/CHIP and RMA1/Derlin1 are types of E3 ubiquitin ligases.

- **Retrotranslocation to the cytosol.**

Proteins tagged for degradation must be translocated to the cytosol via retrotranslocation. The reduction of disulfide bonds of misfolded proteins is necessary for the retrotranslocation step of ERAD. The cytoplasmic AAA-ATPase p97 is the major retrotranslocation protein required for the clearance of misfolded proteins from the ER.

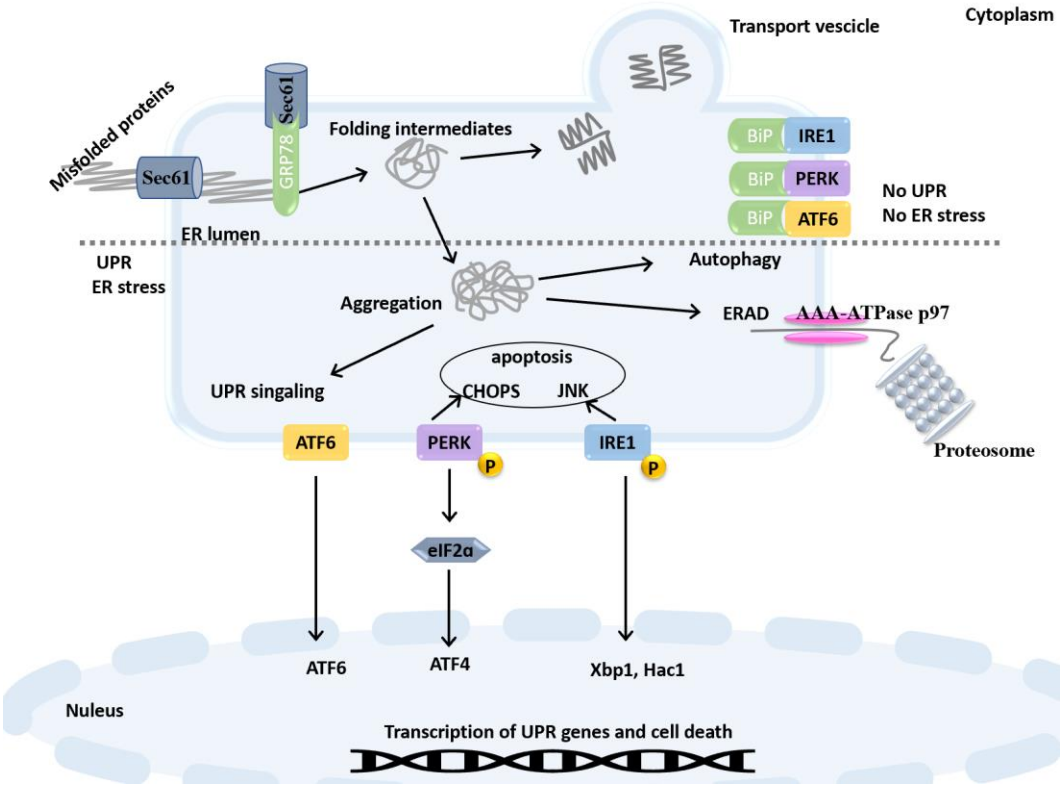


Figure 16. Simple illustration of UPR signaling pathway

Misfolded proteins enter the ER lumen through translocon Sec61. Different intermediate chaperons try to rescue protein folding into their native conformation as properly folding proteins are excreted from the ER. Accumulation of misfolded proteins either destined for the UPR pathway, autophagy by lysosomes, or enters the ERAD pathway for retrotranslocation through AAA-ATPase p97 TLC translocon to the cytoplasm, ubiquitination by E3-ubiquitin ligase, and then degradation by the proteosomal complex. Under resting conditions, the GRP78 (Bip/Hsp70) chaperon binds to the UPR transducers keeping them in an inactive state. Also, GRP78 blocks the opening of the TLC and calcium leakage from the ER. Under ER stress, UPR is triggered, and GRP78 disassociates from the UPR transducers resulting in PERK phosphorylation and activation of eIF2α which leads to the suppression of protein synthesis and the activation of the transcription factor ATF4. At the same time, IRE1 is activated and enhances the expression of Xbp1/Hac1, which allows the transcription of different chaperones and quality control factors to limit the accumulation of misfolded proteins. GRP78 release from the TLC causes calcium leakage from the ER to the cytosol. Under chronic stress UPR signals PERK/ATF4-dependent expression of the transcription factor CHOP, and IRE1/JNK signaling pathway leading to apoptosis.
Abbreviation; ER, Endoplasmic reticulum; E3, enzyme 3 ubiquitin ligase; ERAD, Endoplasmic reticulum-associated degradation; UPR, unfolded protein response; TLC, translocon; eIF2α, eukaryotic initiation factor; ATF4/6, activating transcription factor 4/6; BiP, immunoglobulin heavy-chain-binding protein; Hac1, homologous to ATF6/CrebI; Hsp70, heat-shock protein 70; IRE1, inositol-requiring enzyme 1; PERK, protein kinase RNA (PKR)-like ER kinase; Xbp1, X-box binding protein 1.

4.2. Importance of Selenoprotein N in ER stress and Ca^{2+} mobilization.

SelenoN plays an important role in regulating the redox homeostasis and cell protection against oxidative stress (Arbogast et al., 2009; Arbogast & Ferreiro, 2010) as it was proposed to be involved in ER redox and calcium homeostasis control (Zito & Ferreiro, 2021). SelenoN and ERO1 expressions were found to be co-regulated in the ER. As mentioned previously ERO1 oxidase activity leads to elevated levels of H_2O_2, therefore SelenoN was proposed to protect the ER from the ERO1-generated peroxides. Moreover, the Ca^{2+} transporter SERCA2 was identified as a redox partner of SelenoN, where SelenoN acts as a calcium sensor through the EF-hand domain. An illustration of the SelenoN model of EF-hand domain sensing for Ca^{2+} is illustrated in (Figure 17) (Chernorudskiy et al., 2020). The working hypothesis is that ERO1-generated peroxides oxidize SERCA2, whereas SelenoN reduces SERCA2 and enhances its activity (Marino et al., 2015).

RyRs were also suggested to be a binding partner of SelenoN since deficiency in SelenoN had led to a low binding affinity between RyRs and ryanodine and caused a defect in RyRs response to oxidative stress. This interaction of both proteins makes SelenoN an important cofactor in RyRs activity and an important factor in calcium level regulation within the ER. In addition, a study on zebrafish embryos described SelenoN interaction with RyR1 (Jurynec et al., 2008). The presence of EF-hand motifs suggests that SelenoN binds to Ca^{2+} and upon binding a conformational change occurs on SelenoN, as with all other proteins-containing EF-hand motifs. Also, mutations in either *SELENON1* or *RYR1* in zebrafish embryos led to defects in Ca^{2+} mobilization. This was concluded by measuring the free Ca^{2+} levels around the Kupffer's Vesicle, a ciliated organ in zebrafish embryos used for Ca^{2+} flux measurement (Jurynec et al., 2008). This study supported the notion that SelenoN acts as a cofactor for RyR1 and regulates Ca^{2+} flux into and out of the ER in a process that is crucial for muscle development and contraction. The defect in Ca^{2+} homeostasis and mobilization could be an important factor leading to the development of congenital myopathy disorders. However, no clear molecular or enzymatic mechanism has been elucidated for SelenoN yet and the hypotheses proposed based on the presence of the SCUG and the EF-hand motifs, still await final demonstration.

Also, SelenoN deficiency in muscles myotubes was found to be a causative factor for the susceptibility of myotubes to oxidative stress induced by accumulation of H_2O_2, in addition to

elevated Ca^{2+} in the cytosol compared to the ER that leads to Ca^{2+} homeostasis imbalance (Marino et al., 2015; Pitts & Hoffmann, 2018).

Altogether, SelenoN loss-of-function was shown to result in impaired Ca^{2+} homeostasis within the ER/SR, and in addition, SelenoN acting as an ER Ca^{2+} redox sensor was proposed to regulate ER/SR calcium release by RyRs and reuptake by SERCA, through redox control of their activities.

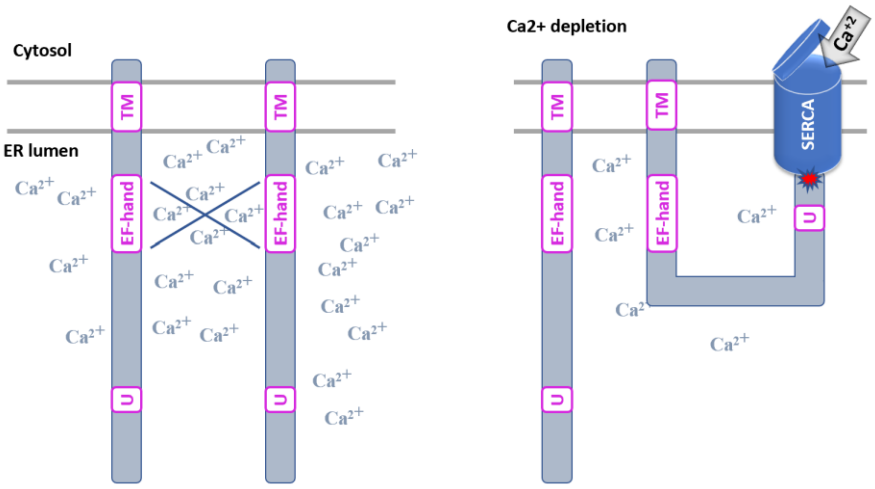


Figure 17. SelenoN model within the Endoplasmic/Sarcoplasmic Reticulum. SelenoN senses calcium through its EF-hand domain. Lower calcium concentration in the ER (below 100–300 µM) exerts conformational changes on SelenoN to the redox active monomer form that can activate SERCA pump and allows the influx of calcium ions in the ER lumen. Created by Microsoft PowerPoint and adapted from (Chernorudskiy et al., 2020).

5. Selenoprotein N effects on mitochondria.

As we know, mitochondria are the powerhouse of the cell, while ER is the factory of protein synthesis, folding and assembly, and carbohydrate and lipid metabolisms. Both organelles communicate through contacts between the two compartments called mitochondria-associated membranes (MAMs). Alterations in MAMs structure caused by ER stress disrupt the communication between ER and mitochondria. This miscommunication was associated with several metabolic dysfunctions and neurological disorders. However, SelenoN was found to accumulate near MAMs and to influence the capacity and function of mitochondria. Moreover, its depletion affects ER-mitochondria communication and calcium exchange (Perrone et al., 2020).

In a recent study, small areas of mitochondria depletion described as minicores and sarcomere disorganization were observed in muscle biopsies taken from *SELENON*-RM patients (Filipe et al., 2021). Even more, several physiological complications were observed in *SELENON*-RM patients such as insulin resistance, cachexia (loss of muscle), and alteration in fat distribution, but no clear explanation was obtained yet. To investigate the underline cause of these physiological modifications, metabolic studies on wildtype and *SelenoN*$^{-/-}$ mice were performed. SelenoN was found to regulate body mass, fat distribution, energy consumption, and metabolic efficiency of the cell, providing a possible explanation for the low body mass index (BMI) and abnormal glucose metabolism in *SELENON*-RM patients (Filipe et al., 2021).

Moreover, Ca^{2+} levels regulate mitochondrial enzymes and the speed of the oxidative phosphorylation pathway. Hence, *SelenoN*$^{-/-}$ mice experienced impaired mitochondrial function, reduced efficiency of oxidative phosphorylation (OXPHOS) due to impaired functions of protein complexes (I-IV) of the electron transfer system (ETS), decreased mitochondrial membrane potential, reduced ATP production, and increased glucose metabolism. This study concluded that SelenoN controls metabolism, acting as an intermediate between mitochondria bioenergetic and ER/SR homeostasis (Filipe et al., 2021). Furthermore, as a result of impaired mitochondrial function, glycogen metabolism increased in the liver and muscles of *SelenoN*$^{-/-}$ mice indicating switching to anaerobic glycolysis, which is a poor source of energy compared to OXPHOS. This loss of mitochondrial capacity in managing the ETS makes anaerobic metabolism a more favorable source of energy than aerobic one (Wesolowski et al., 2022).

INTRODUCTION II

The clinical phenotype associated with the SelenoN loss of function has labeled this protein as a protein with muscle-specific activity and focused research on muscles using standard animal models. Yet animal models display major differences with the patients. For example, the *SELENON*-/- mice did not develop the expected muscle symptoms spontaneously, and their study provided very limited information on SelenoN molecular activity. Furthermore, SelenoN expression is not restricted to the muscle tissue, since the *SELENON* gene is evolutionarily conserved in the animal kingdom, including lower eukaryotes with no mesodermal structure. These observations suggest that SelenoN has a more ancient and broader spectrum of functions.

Unexpectedly, a survey on the genomic databases identified a bacterial gene homologous to *SELENON* in a unique group of unclassified and yet-to-be-cultured bacteria: the candidate phylum *Poribacteria*. Moreover, a *SELENON* gene in *Poribacteria* genomes is detected in a subgroup living in symbiosis with sponges, such as *Aplysina aerophoba*, but not in their free-living counterparts. This observation suggests that SelenoN activity contributes to bacteria-sponge symbiotic interactions. Therefore, we got interested in studying the bacterial SelenoN and its association with bacterial-host symbiotic interactions.

The next section will appear to be not related to Selenoproteins but it is necessary to highlight the environment hosting the bacterial Selenoprotein N to understand the physiological function of SelenoN in the ecological model.

I. Marine sponges (*Porifera*)

1. Anatomy and cellular functions

1.1. Sponges taxonomy:

Sponges are very ancient metazoans, also named parazoans, that evolved back over 580 million years ago. Sponges are members of the Animalia kingdom and *Porifera* phylum. They are sessile animals "immobile", filter-feeders (using a water-filter feeding system). They have asymmetrical body shapes.

Sponges are diploblasts, meaning that the body of a sponge is made up of two cell layers, the ectoderm, and the endoderm, separated by a jelly-like extracellular matrix or mesenchyme, called the mesohyl. Despite their limited architectural organization, sponges are constituted of specialized cells that carry out dedicated functions but do not differentiate into tissues or true organs. This kind of cell specialization represents the first step of evolution into tissues and organs.

Sponges are classified into four monophyletic classes based on their skeletal spicules components (Alvarez et al., 2017):

- *Calcerea* (calcareous sponges) spicules are made up of calcium carbonate;
- *Hexactinellida*, also called glass sponges, in which the skeleton is made up of silica;
- *Desmospongiae*, spicules are made up of silica and fibers constituted of the collagen-like protein spongin;
- Homoscleromorpha, a group containing both speculate and a speculate species.

Demospongiae is the most studied class, and the most abundant and diverse class making up 76% of all sponge species. *Aplysina aerophoba,* one of the most studied models, is a sponge of the class Demospongiae, order *Verongidae*, and family *Aplysinidae*.

1.2. Structure of sponges

A sponge's body is constructed from pores, canals, and passageways. The body structure of sponges is divided into three main types based on the complexity of the canal system as depicted in (Figure 18): asconoid, synconoid, and leuconoid. Spongocoel is the sponge's central cavity, also called the central water channel. Osculum is the opening for expelling filtered water out of the sponge cavity in which water flows into and out of a sponge in a unidirectional path (Godefroy et al., 2019).

Desmospongiae, including *A. aerophoba*, are mainly leuconoid sponges with a hard skeleton made up of siliceous spicules.

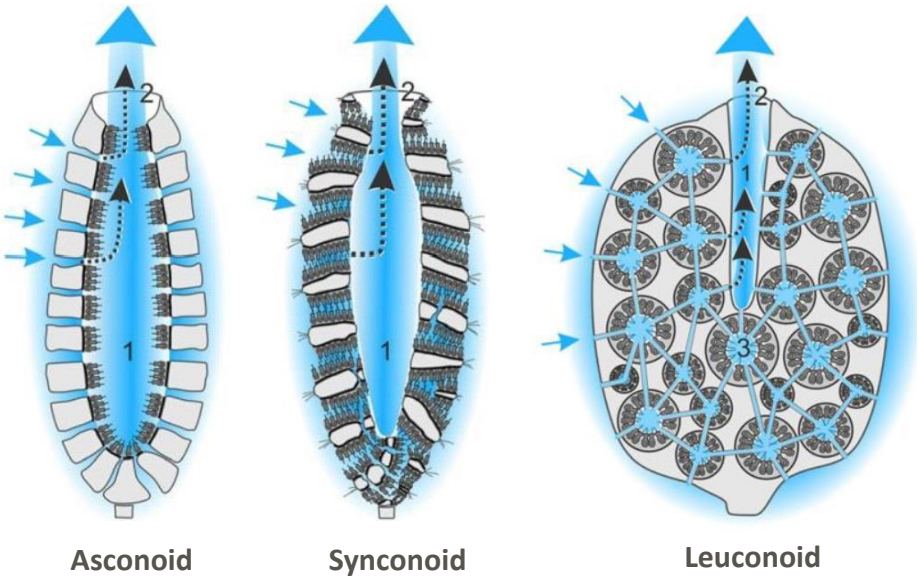


Figure 18. The main three body plans of *Porifera*. Asconoid is the basic body plan**.** The canal system is very simple and water pumps directly from the pores to the central canal. **Synconoid** is a second in hierarchical complexity. A network of small canals surrounds the central canal, and digestion takes place in the canal network only since no choanocytes line up the central canal. **Leuconoid** is a complex body plan. The central canal is very narrow and surrounded by a branched network of canals lined up with choanocytes. Arrows represent the seawater flow direction. All three types contain a mounting base to help sponges anchored to a solid surface very tightly. Picture source (Godefroy et al., 2019).

1.3. Specialized sponge cells and their functions.

Marine sponges are constituted of few, but highly specialized cells, a description of the different sponge cells and their associated functions are as follows;

- The epidermal cells or pinacocytes, a thin flat layer that lines the external body of the sponge ectoderm. They act as a protective layer and also digest organic materials.
- Collar cells, or choanocytes, are flagellated cells with a collar of microvilli. They line up the water canals forming the endoderm. With their attached flagella, choanocytes are capable to create currents to draw water into the central cavity through pores and out through the osculum. They also strain and trap food particles in the mucus coat of the collar, then the food particles are engulfed by phagocytosis and form a vacuole. The food vacuole is either digested by lysozymes or transferred to amoebocytes. In short, choanocytes have a very sophisticated function in sponge nutrition, since they absorb and digest food particles. In addition, they are responsible for producing and transporting sperm.
- Amoebocytes, also known as archeocytes, are scattered cells in the mesohyl. Amebocytes transport nutrients, and can differentiate into other different cell types, such as oocytes.
- Spongocytes secret the spongin fibers.
- Collenocytes, also called lophocytes, secrete the collagen-like protein that supports the mesohyl.
- Sclerocytes excrete spicules. They localize in the mesohyl and maintain the structure of sponges and deter predators.
- Porocytes form the Ostia to regulate water circulation. Water flow is very important to sponges because it allows food particles to go through. In addition to feeding and removal of waste, water circulation provides sponges with the required amount of oxygen through diffusion.
- Myocytes are contractile muscle-like cells surrounding canal openings and porocytes. Myocytes can contract to regulate water flow in the sponges.

Sponge anatomy and organization of different cell types within the sponge body are depicted in (Figure 19)

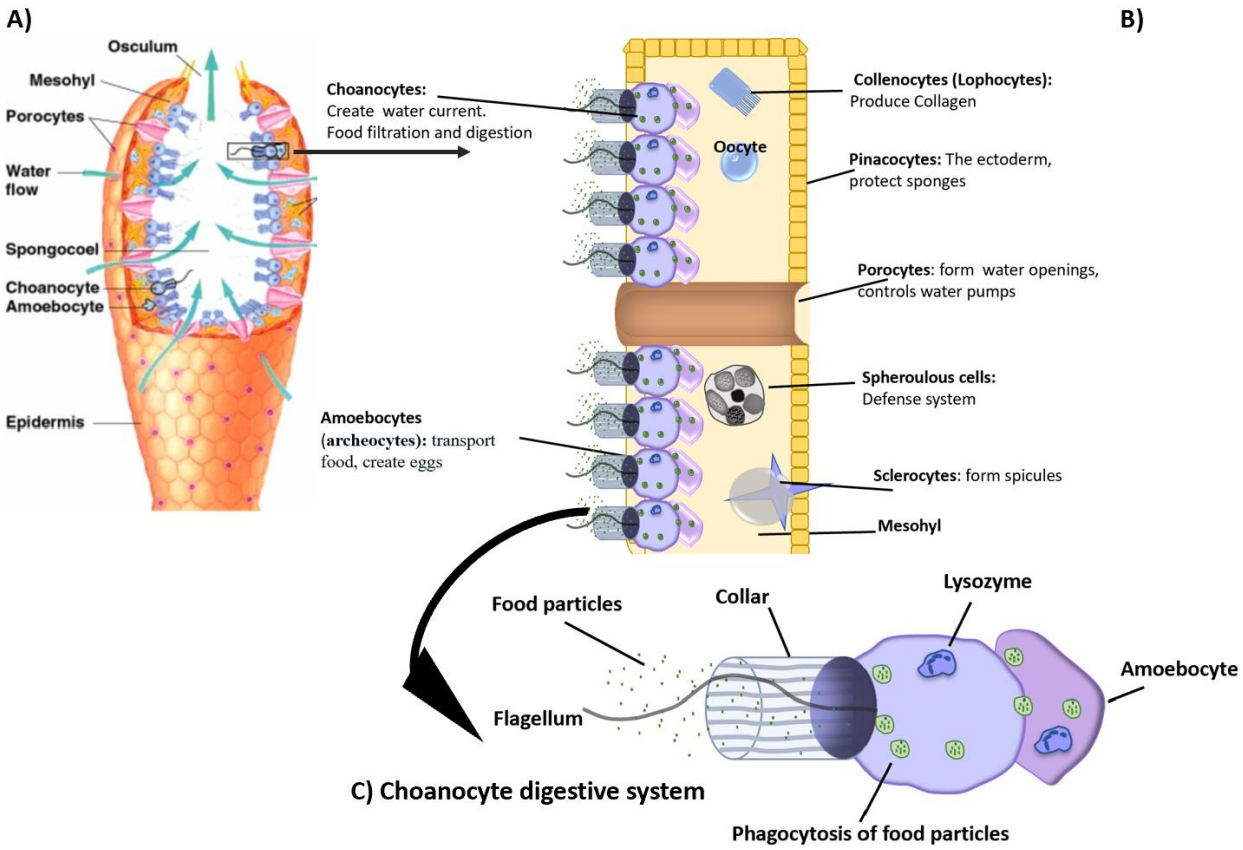


Figure 19. *Porifera* body structure. A) An overall depiction of *Porifera* anatomy. B) A magnification of the two layers, endoderm, and ectoderm with different sponge-specialized cells. C) AN illustration of the digestive system in *Porifera*. A) is original, B) and C) are created by Microsoft power point and adapted from the University of Maryland geology website *https://www.geol.umd.edu/~tholtz/G331/lectures/331porifera.html*

2. Symbiosis in marine sponges

Holobiont is a term that describes a sponge host with its associated consortium of bacteria, archaea, unicellular algae, fungi, and viruses. The microbial consortia make up 35-50% of the sponge's biomass and contributed to different aspects of the sponge's physiology and ecological importance. The most studied sponge symbionts are bacterial symbionts (Hentschel et al., 2006; Webster & Thomas, 2016), which are investigated through metagenomic analyses. The metagenome of the sponge holobiont is named a hologenome.

2.1. Sponge Microbes: Abundancy and Diversity

Marine sponges can be divided into species containing either a high microbial abundance (HMA) or low microbial abundance (LMA), as shown in an electron microscopy study on 56 sponge species (Gloeckner et al., 2014). Sponge-associated microorganisms are scattered in the mesohyl, and some of them are concentrated around the choanocytes chambers. According to 16S rRNA gene sequence and several phylogenetic analyses, the most abundant sponge microbial consortia are represented in the taxa Gamma- and Alpha *Proteobacteria*, *Actinobacteria*, *Chloroflexi*, *Nitrospirae, Cyanobacteria, Planctomycetes, Acidobacteria, Verrucomicrobia,* the candidate phylum *Poribacteria*, and the archaea *Thaumarchaea*. One of the most consistent symbionts are the photoautotrophic *Cyanobacteria* and other microalgae that are required for photosynthesis and are mostly localized below the outer layer, pinacoderm, close to the source of light (Brinkmann et al., 2017; Siegl et al., 2008)

Sponge-associated microorganisms could be specific to the host, but also stable across different biogeographical locations and environmental conditions, while some microorganisms found to share between phylogenetically unrelated sponges from different geographical locations (Pita et al., 2018; Webster & Thomas, 2016). Many phylogenetic trees have been constructed showing the abundance and diversity of the microbiome of different sponge species. A recent phylogenetic tree shown in (Figure 20) was constructed based on the analysis of 1188 metagenome-assembled genomes (MAGs) from seven marine sponges of the class *Demospongiae,* demonstrating the distribution of 25 microbial phyla (Robbins et al., 2021). These observations highlight the very dynamic interaction between the bacteria community and its sponge host.

The forces that influence the diversity and distribution of microbial symbionts within their sponge hosts are not well understood and set a question mark on the mechanism that the sponge holobiont follows to control the size of the symbiont population under a tolerance threshold (Webster & Thomas, 2016).

2.2. Modes of Symbiont Transmission.

Sponge's bacterial symbionts, including the hologenome, can be transferred to newborn sponges either by vertical transmission from parents, or horizontal transmission through acquisition from the surrounding environment. Either of these modes of transmission has certain benefits and costs. In addition, numerous studies supported the hypothesis of a mixed mode of symbionts transmission. In vertical transmission, the loss or reduction of symbiont population/genome among successive generations is much possible especially if larvae were disposed of in distant habitats not convenient or not optimal for the symbiont. While in horizontal transmission the symbiont acquisition is limited by the availability of microorganisms in the surrounding environment (Carrier et al., 2022).

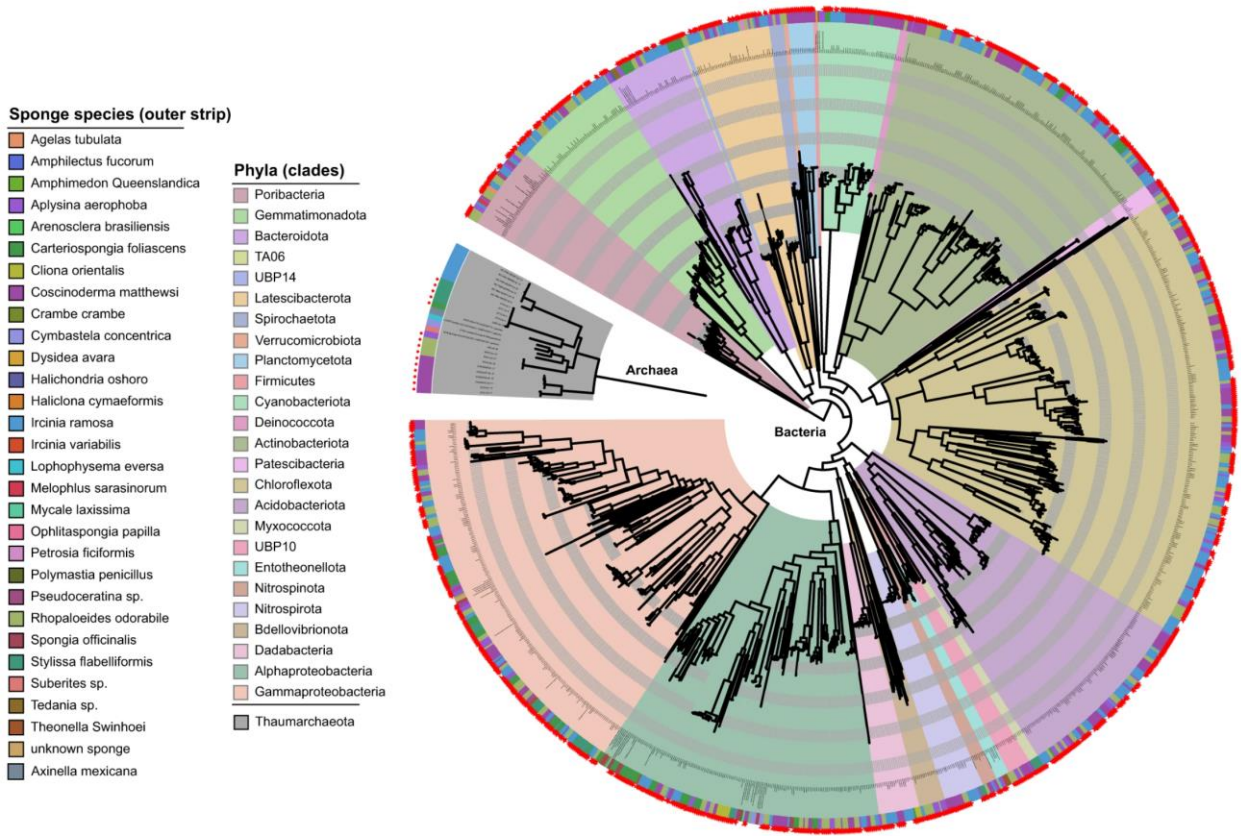


Figure 20. Phylogenetic tree of marine sponges associated-microbiome. A total of 30 sponge species MAGs were used to construct this phylogenetic tree. Seven marine sponges of the class demospongiae in this study, in addition to previous bacterial and archaeal MAGs from previous studies with >50% completeness and <10% contamination. Microbial affiliation is denoted by the inner clade colour, while sponge species are denoted by the outer coloured strip. The red stars denote the MAGs in the study (Robbins et al., 2021).

2.3. Symbionts' physiological function within the sponge holobiont.

Symbiotic bacteria contribute to diverse functions within the host, including the removal of heavy metal and metabolic waste, recycling of organic matter, nutrient supplementation, and vitamin synthesis. Another important role that the symbiont bacteria provide for their host is the defense against infections and predators.

Not all sponge-associated microorganisms are symbionts, some are food sources, and sponges have the capability to distinguish between both. A feeding experiment study was performed on *A. aerophoba* to investigate the uptake and processing of microorganisms by the sponge (Wehrl et al., 2007). The *Vibrio* strain MMW1 was labeled with a green fluorescent protein (GFP), incubated with the sponge, and cryosections of sponge tissues detected the GFP-labeled *Vibrio* bacteria inside the choanocytes and not in the mesohyl, which was proof for the recognition of *Vibrio* strain as a food particle by the sponge (Brinkmann et al., 2017).

Due to the complexity of the sponge's microbiome, most sponge symbionts are resistant to cultivation outside of their hosts, and this is the main obstacle to understanding the functional link between symbionts and their hosts. However, genomic approaches such as single-cell genome sequencing methods and physiological experiments have in some way overcome this obstacle and provided some knowledge into the metabolic pathways in sponge holobiont. Some of these metabolic pathways are carbon metabolism, nitrogen fixation, vitamin synthesis, sulfur oxidation, and waste product removal (Hentschel et al., 2012; Webster & Thomas, 2016). Below are some of the functions provided by sponge-associated bacteria.

A) Carbon metabolism

Sponge holobiont contains both heterotrophic and autotrophic bacterial symbionts. In addition, genes encoding proteins involved in glycolysis and pentose phosphate pathways (PPP), the tricarboxylic acid (TCA) cycle, and oxidative phosphorylation have been identified in the genomic sequencing of some sponge symbionts. As an example, genomic analysis of a member of *Poribacteria* revealed the presence of ATP-citrate lyase genes in addition to other genes involved in the reductive TCA cycle, therefore, *Poribacteria* is capable of autotrophic carbon fixation (Hentschel et al., 2012).

B) Nitrogen metabolism

Nitrogen metabolism is very important to sponge host and their symbionts. Sponge-associated microorganisms uptake ammonia and nitrite which are found as metabolic waste products in sponges. Moreover, urea was detected as an alternative source of nitrogen. Urease-encoding gene clusters, like urea transporters and other accessory genes, have been reported in two symbiont genomes (Siegl et al., 2011). Both archaea and bacteria monooxygenases are responsible for ammonia oxidation ‘‘Nitrification’’. Ammonia-oxidizing archaea (AOAs) and ammonia-oxidizing bacteria (AOBs) are found widespread in sponges from different geographical locations. In addition to nitrification, several genes encoding enzymes that are necessary for denitrification have been identified by a single-cell genomics analysis of a member of *Poribacteria*. (Bayer et al., 2008; Hentschel et al., 2012).

C) Sulfur

Both sulfur-oxidizing bacteria (SOB) and sulfate-reducing bacteria (SRB) were found as symbionts in marine sponges. An isotopic study on *Geodia barretti* sponge species detected high sulfur reduction rates (Tian et al., 2014).

D) Phosphate

Phosphate was found stored in polyphosphate granules (poly-P) in different sponge species, and symbiont bacteria such as *Cyanobacteria*. In addition, poly-P kinase (ppk) genes encoding the kinase of poly-P synthesis, were amplified from the sponge genome. The poly-P backbone is constituted from the high-energy phosphoanhydride bonds that could be an energy reservoir for sponges in times of energy deprivation. The occurrence of poly-P genes in the sponges’ genomes, and the accumulation of poly-P granules in sponges and symbiotic bacteria suggest the uptake and recycling of phosphate by the symbiotic bacteria (Zhang et al., 2015).

E) Vitamins synthesis

Essential vitamins have to be provided to sponges as they are unable to synthesize them. Therefore, symbionts play an important role in the maintenance and stability of sponges by providing them with essential vitamins. Several studies on genomic and metagenomic analyses have identified genes encoding enzymes involved in vitamin biosyntheses such as cobalamin

(vitamin B12), riboflavin (vitamin B2), biotin (vitamin B7), thiamine (vitamin B1) and pyridoxine (a form of vitamin B6) (Hentschel et al., 2012; Webster & Thomas, 2016).

However, there are still more undiscovered metabolic interactions of sponge symbionts due to the high complexity of the sponge microbiome and more studies will add more value to understanding the functions of sponge holobiont and the force driving the evolution of these functions.

2.4. Sponge-Derived Natural Products.

Sponges recruit their associated microbiome as a defense army against predators and pathogens. A wide range of natural products or bioactive compounds were identified in sponge extracts, including polyketides, macrolides, alkaloids, sterols, peptides, porphyrins, and terpenes. Interestingly, many of these compounds carry out medicinal properties, such as anti-viral, anti-inflammatory, antimicrobial, antifungal, and anti-cancer. Bioactive compounds were suggested to be produced either by the host or the symbionts and sometimes by the collaboration of both under certain conditions (Brinkmann et al., 2017).

Nevertheless, some studies have proven the role of symbionts in synthesizing those natural products. In an old study by (Unson et al., 1994) *Cyanobacteria* along with heterotrophic bacteria were isolated from a marine sponge. Coupled gas chromatography-mass spectrometry and proton nuclear magnetic resonance spectroscopy confirmed the presence of polybrominated biphenyl only in the *Cyanobacteria* isolate and not the other isolated symbiont.

Moreover, another study by (Nicacio et al., 2017) showed that bromotyrosine-derived alkaloids that were previously isolated from *Verongida* sponges have also been extracted from the *Pseudovibrio denitrificans* isolated bacterium from the *Arenosclera brasiliensis* sponge.

In a recent study analyzing the chemical composition of *A. aerophoba*, halogenated compounds were isolated from the ethanolic extract of the sponge. Bromotyrosines are the main bioactive compound in this sponge species which is characterized by their antibacterial and antitumor properties. Nonetheless, in this study it was not known if these compounds originated from the sponge cells or their associated microorganisms (Orfanoudaki et al., 2021).

3. The candidate phylum *Poribacteria*

3.1. Identification of *Poribacteria.*

The *Poribacteria* lineage was coincidently discovered from the 16S rRNA library construction of microbial consortia from different sponge species. Bacterial cells with cell compartmentation of different morphotypes were previously visualized by electron microscope in two marine sponges using anti-ds/ssDNA in immunogold labeling. This compartment was suggested to be a membrane-bounded nucleoid (Fuerst et al., 1998; Lindsay et al., 1997). Based on that observation, a molecular study on different marine sponges and their associated microorganisms was performed to provide insights into their phylogeny and distribution among sponge species. Since the phylum *Planctomycetes* was the only phylum at that time that was recognized with cell compartmentation, fluorescence *in situ* hybridization **(**FISH) analysis was performed using the *Planctomycetes*-specific probes for FISH and 16S rRNA analysis from different sponge species. 16S rRNA library was constructed for the associated-microbial communities. The phylogenetic analysis of the library revealed the presence of a novel lineage that showed an average of less than 75% sequence similarity to any of the known bacterial phyla, named later as candidate phylum *Poribacteria*. The *Poribacteria* lineage was found to be moderately related to the *Planctomycetes*, *Chlamydia*, and *Verrucomicrobia* superphylum (PVC). After that, *Poribacteria*-specific primers were designed from the 16S rRNA gene library obtained above to confirm the affiliation of *Poribacteria* in marine sponges, in particular sponges from the order *Verongida* such as *Aplysina aerophoba*. FISH analysis of *Poribacteria* showed ring-like structure signals that suggested cell compartmentalization. Furthermore, in the same study, transmission electron microscopy (TEM) on an *A. aerophoba* sample showed six different morphotypes of cell compartmentalization (Fieseler et al., 2004). The poribacterial genome varies in size from 4.25 to 6.27 Mb, rich in guanine and cytosine content by 47% to 50% in some strains. *Poribacteria* is divided into subgroups or sub-phylotypes according to 16s rRNA sequence relatedness (Kamke et al., 2013).

A recent comparative pangenomic study between the sponge-associated *Poribacteria* (Entoporibacteria) and identified free-living *Poribacteria* (Pelagiporibacteria) was performed to study the phylogenetic relationship and to highlight common functional properties between both lineages. Commonly shared features are found in both lineages, like facultative anaerobic

metabolism, denitrification, and organosulfur utilization. The most surprising shared characteristic is the identification of genes encoding for eukaryotic-like cell adhesion molecules, which is not expected to be found in free-living bacteria. In contrast, the main different features that distinguished both lineages, were the presence of flagella and chemotaxis genes in *Pelagiporibacteria*, while *Entoporibacteria* were identified with the presence of widely broaden restriction endonucleases, transposases, CRISPR repeats, and toxin–antitoxin coupled genes (Podell et al., 2019). A representation of the Shared and unique protein family between both subgroups is displayed in (Figure 21).

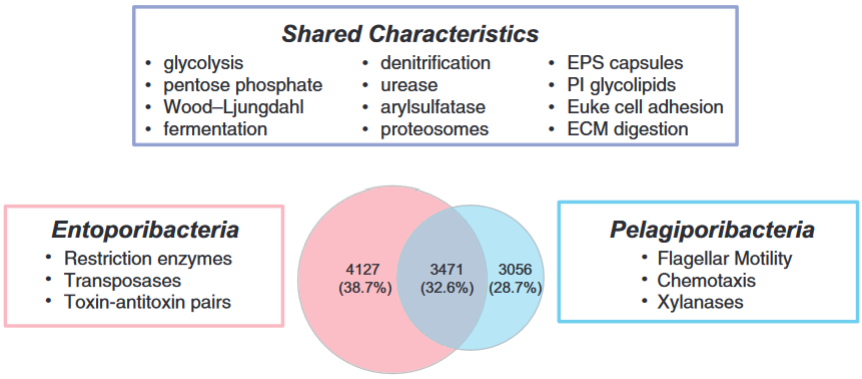


Figure 21. Shared and unique protein family percentages within *Entoporibacteria* and *Pelagiporibacteria* subgroups. Specific functional characteristics attributed to both subgroups were found in all members of their respective groups. Abbreviations; EPS, exopolysaccharide; PI, phosphatidyl inositol; Euke, eukaryotic, ECM extracellular matrix. (Podell et al., 2019).

3.2. Characterization and genomic studies of *Poribacteria.*

The members of candidate phylum *Poribacteria* are gram-negative bacteria, prevalent among different sponge species, specifically *A. aerophoba*. Phylogenetic identity, functional properties, and metabolic potential of *Poribacteria* were studied by single-cell genomics in several studies. Some of the characteristics and potential functions of *Poribacteria* are listed below:

3.2.1. Carbohydrate degradation

Based on the genes identified in their genomes, *Poribacteria* has been implicated in carbohydrate degradation. Six single amplified genomes (SAGs) of poribacterial cells were screened for the presence of carbohydrate degradation genes. Carbohydrate-active enzymes (CAZymes), such as glycoside hydrolase (GH) and glycosyl transferase enzymes, were found to be highly abundant in different poribacterial subgroups. In addition, genes coding for enzymes involved in the degradation of carbohydrates, such as galactoside and fructoside polymers, were encoded in poribacterial genomes. Furthermore, genes of vital pathways such as glycolysis, PPP, TCA cycle, and oxidative phosphorylation were identified.

Genomes encoding for enzymes employed in the inositol degradation pathway were identified in *Poribacteria*. Inositol phosphate, a phospholipid, component of the cell membrane in eukaryotes, participates in signal transduction in sponges. It has been hypothesized that *Poribacteria* uses Inositol phosphate as a carbon source for regulating metabolic functions involved in the sponge–microbe symbiosis (Kamke et al., 2013).

In addition, poribacterial genomes were found to possess genes related to uronic acid degradation. Uronic acids are derived from the oxidation of aldose monosaccharides such as glucose and galactose. Uronic acids are major components of glycosaminoglycans (GAGs), the main building blocks of the extracellular matrix of sponges. Therefore, the presence of such major metabolic pathways suggests the involvement of *Poribacteria* in carbohydrate metabolism in sponges and their vital role in the sponge–microbe symbiosis (Kamke et al., 2013).

3.2.2. Microcompartments in *Poribacteria*

As shown previously *Poribacteria* has been characterized by cell compartmentalization, which was originally suggested to be a nucleoid-like structure. However, SAGs analysis of four poribacterial genomes revealed the presence of genes encoding bacterial microcompartments (BMCs) shell proteins (Kamke et al., 2014). BMCs are cell organelles found in prokaryotes, composed of a set of conserved families of proteins. BMCs provide a convenient environment for the optimization of chemical reactions, and the isolation of toxic compounds (Kirst & Kerfeld, 2019). The enzymatic reactions regulated by BMCs shell proteins are in most cases dependent on cofactors. This explains the identification of riboflavin (vitamin B2) biosynthesis genes near the BMCs (Kamke et al., 2014).

3.2.3. phyH-domain containing Proteins

Genes encoding for eukaryotic phytanoyl-CoA dioxygenase (*phy-H)* were found abundant in poribacterial genomes. Phytanic acid is a fatty acid metabolite of chlorophyll, and phytanoyl-CoA is formed from phytanic acid attached to Co-A. The phy-H is responsible for the α-oxidation of phytanoyl-CoA via a Fe(II) and 2oxoglutarate-dependent reaction. In addition, phyH-domain-containing proteins were found involved in quorum sensing (Kamke et al., 2014) and biosynthesis of mitomycin antibiotics/polyketide fumonisin (Jahn et al., 2016). Fumonisin belongs to the mycotoxin secondary metabolites, therefore, this might lead to the role of *Poribacteria* in protecting sponges against predators and bacterial infections.

3.2.4. Eukaryote-like repeat domain-containing proteins

Eukaryote-like repeat domain-containing proteins are abundant in symbionts of marine sponges. They were suggested to mediate host-microbial interactions. SAGs showed a high frequency of these domains in *Poribacteria* genomes, higher than the average frequency in non-symbionts free-living bacteria. Examples of these domains include ankyrin (ANK), tetratricopeptide repeats (TPR), WD40 domains, leucine-rich repeats (LRR), and low-density lipoprotein (LDL) receptor repeat (Kamke et al., 2014).

ANK domain-containing proteins of sponge symbionts known from a previous study on *E. coli* to prevent phagocytosis by amoebae (Nguyen et al., 2014), hence the function was assigned to *Poribacteria* is to resist phagocytosis by sponges.

All LDL receptor genes were identified with transmembrane helices (TMHs) and signal peptides (SPs), which suggests their function on the cell surface of *Poribacteria* (Kamke et al., 2014). The presence of those eukaryotic-like repeat domains in Poribacterial genomes supports the host-bacterial interaction functions.

Fluorescence *in situ* hybridization correlative light and electron microscopy (FISH-CLEM) combined with the immunohistochemistry (IHC) method was used to identify the *Poribacteria* in *A. aerophoba* ultrathin tissue section through Poribacteria-specific 16S rRNA probe (POR11300) (Figure 22). In addition, FISH-IHC-CLEM method was applied also to detect the expression level and localization of different cell compartmentation-related proteins including, BMC shell marker, gas vesicle protein (GvP), and ExbD. The ExbD proteins are membrane-bound transporters essential for ferric ion uptake in bacteria. While, GvP is a nano compartment that is made from proteins only, and plays a role in cell buoyancy, the ability to float in the surrounding environment. The GvP signal showed a high transcription level throughout the cytoplasm, while ExbD and BMC-shell marker protein signals showed membrane-bound expression of the proteins (Figure 23, 24) (Jahn et al., 2016).

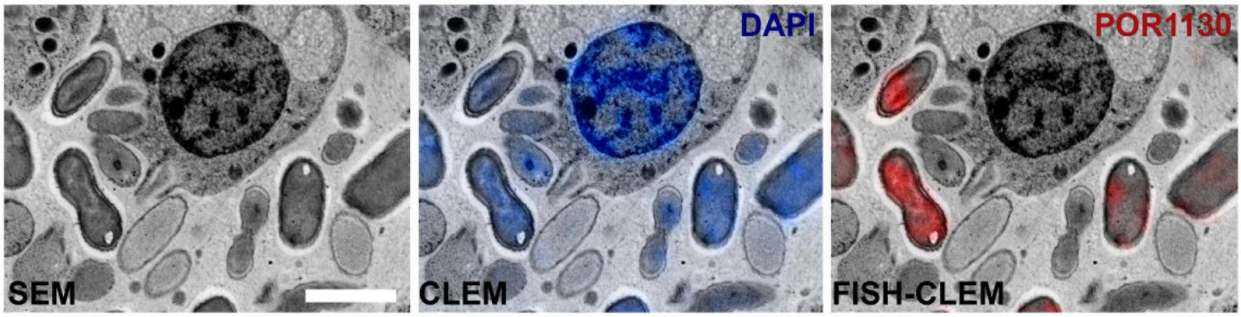


Figure 22. Identification of *Poribacteria* ultracellular structure using FISH-CLEM. (A) Scanning electron microscopy images (SEM) combined with fluorescence signal of the nucleotide stain DAPI (blue; CLEM) and the Poribacteria specific 16S rRNA probe POR1130 (red; Alexa546; FISH-CLEM). Scale bars, 2 µm. (Jahn et al., 2016)

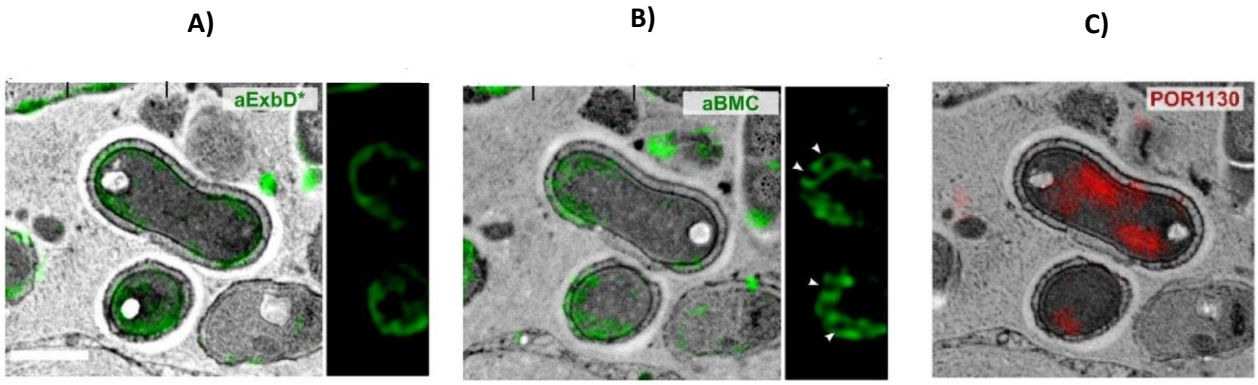


Figure 23. Cell compartmentation-related proteins. Ultrastructural resolution in *Aplysina aerophoba* tissue. **A)** ExbD proteins (green; FITC), **B)** BMC-shell marker (green; FITC) were localized within the *Poribacteria* membrane. **C)** *Poribacteria* cells identified by FISH-CLEM using the Poribacteria-specific 16S rRNA probe POR1130 (red; Alexa546 double 5'3' labeled). Micrographs represent the same cells on 3 consecutive sections of 100 nm distance as illustrated in the scheme. (Jahn et al., 2016)

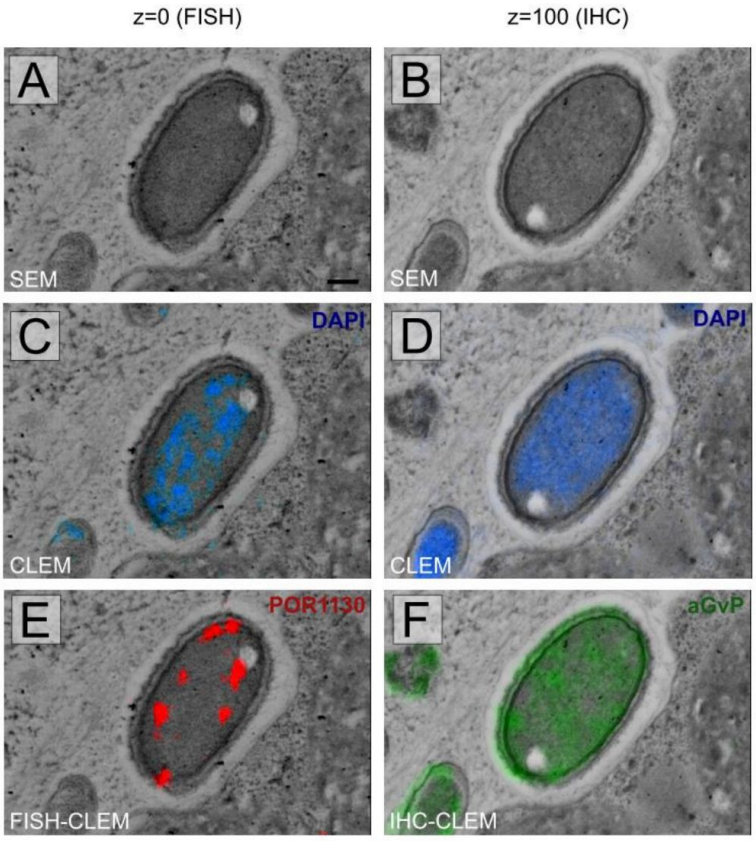


Figure 24. FISH-IHC-CLEM identification of localization of poribacterial gas vesicle protein (GvP) on ultrathin *Aplysina aerophoba* sections. (A-B) SEM images of two consecutive sections of 100 nm distance at 33,000 magnification. (C-D) DNA is counterstained with DAPI (blue). The consecutive sections were stained (E) with the double-labeled probe POR1130 (red; Alexa546; 5'3'), specific for Poribacteria (F) and with FITC labeled donkey anti-rabbit antibody directed against three anti-GvP peptide antibodies (green). Scale bars, 100 nm. (Jahn et al., 2016)

3.3. Cultivation attempts of marine sponges associated-bacterial symbiont.

Many cultivating approaches were conducted on *Poribacteria* and other uncultivable sponges-associated bacteria, but no success up until now in the case of *Poribacteria*. As a matter of fact, the lack of understanding of the complex symbiotic marine life is the main reason that hinders cultivation success. The growth conditions, especially the essential nutrient supply and the ultimate temperature are major obstacles. In addition, H_2O_2 was found to be produced by chemical reactions between different media compositions during autoclaving, thus inhibiting the growth of bacteria. Removing H_2O_2 by the addition of catalase to the culture has improved the growth rate. Furthermore, marine sponges symbiont bacteria are hypothesized to have a slow growth rate and are classified as slow-growing microbes, therefore, other fast-growing microbes compete for nutrient resources (Dat et al., 2021).

An example of a successful cultivation effort on microbial species that were classified previously as "difficult-to-culture" bacterial-sponge symbionts are the cultivation of different *Alphaproteobacteria* strains. The solidifying agar that was thought to be toxic to some microorganisms was replaced with gellan gum, a nontoxic polysaccharide produced by the *Sphingomonas elodea* bacterium used in food processing. They prepared a low-carbon medium and maintained a low incubation temperature of 19 ℃ for 8 weeks incubation period (Karimi et al., 2019). Despite the challenging cultivation efforts, 11 bacterial phyla from marine sponges-associated symbionts have been successfully cultivated.

moreover, a recent cultivation attempt of *A. aerophoba*-derived bacteria took place. The capacity of *Poribacteria* to process and degrade carbohydrates was the basis for designing nine different media formulations with different carbon sources. The identity of successfully cultivated bacteria was identified by the Operational Taxonomic Unit (OTU). This unit is used to classify bacteria based on the sequence similarity of the 16S marker gene. The designed media were productive and efficient in cultivating several different bacterial phyla as follows; *Actinobacteria* (1.2% of cultivated OTUs), *Bacteroidetes* (5.1%), *Planctomycetes* (3.1%), *Proteobacteria* (56.6%), and *Verrucomicrobia* (0.4%). But unfortunately not successful in cultivating Poribacterial phyla (Gutleben et al., 2020).

Objectives of the PhD study

Selenoprotein N has a broad spectrum of functions among different pathways ranging from calcium homeostasis control in the ER and regulation of energy production in mitochondria, to its involvement in muscle regeneration and glucose metabolism. However, many of these biological activities are interconnected and the precise catalytic mechanism(s) behind these functions is yet to be identified. On the other hand, the identification of an orthologous *SELELNON* gene in one group of unclassified bacteria, *Candidatus Poribacteria,* was a great asset in this task and a chance to address SelenoN activity in an original, also natural, ecological and biological context.

The study is divided into three parts:

- In the first part, our plan is to obtain information about SelenoN activity in *C. Poribacteria,* by investigating *SELENON* gene in the context of the *C. Poribacteria* genome. For this, bioinformatics analysis and mapping of the Poribacterial genomes available in the whole genome sequencing WGS database were implemented. This study meant to shade light about the nature of *SELENON* gene and its functional connection to other genes in the bacterial genome.
- The second part of the study intends to perform biochemical characterization of *C. Poribacteria* SelenoN protein. Therefore, first we needed to obtain *SELENON* gene from a natural source. For that, my PhD project focused on the amplification of one operon that is constituted of *SELENON* and its associated gene from its natural host *C. Poribacteria* genome by utilizing a dedicated inverse PCR technique. Then cloning, expression, and purification of SelenoN was optimized. Moving forward, the recombinant SelenoN will be used for further biochemical characterization and *in vitro* enzymatic assays.
- The third part of the study aims to look at the contribution of SelenoN in the *Poribacteria*-sponge symbiotic interaction using fluorescence and electron microscopy. Therefore, I performed Catalyzed Reporter Deposition Fluorescence *in situ* hybridization analyses to study the distribution and localization of *C. Poribacteria in vivo,* using

Aplysina aerophoba as a sponge model. In addition, the expression of SelenoN was detected by immuno-gold labeling technique.

Both second and third parts were conducted in collaboration with Marcelino Suzuki and his team at the Observatoire Océanologique de Banyuls-sur-Mer.

Interestingly, this microscopy part provided critical information on the nature of the host-bacteria functional interactions and suggests a possible function for SelenoN.

RESULTS

I . Bioinformatics

I . Bioinformatics identification of a bacterial *SELENON* gene and analysis of its genomic context

For simplicity, a schematic representation of the bioinformatics study with the results is summarized in (Figure R5).

1. Identification of a *SELENON* homologous gene in one group of bacteria, *Candidatus Poribacteria*

By conducting a survey on the conservation of *SELENON* during evolution, we have identified in the genomic databases repository a homologous gene conserved in a unique prokaryotic group, *Candidatus Poribacteria*. Intriguingly, the peptide sequence of this bacterial SelenoN showed a high degree of homology with the animal protein: 31%, 32%, and 31% identity and 48%, 47%, and 49% similarity between the *Poribacteria* sequence and the Human, zebrafish, and *Nematostella* proteins respectively. In addition, the protein sequence alignment showed the highest degree of conservation with two blocks that correspond to the most conserved domains in animals as well (Figure R1). However, the bacterial SelenoN lacked the N-terminal transmembrane and calcium-binding domain. The sequence and the position of the SCUG catalytic site predicted based on the localization of the Sec residue, are generally conserved, although the first serine (S) and the final glycine (G) residues are sometimes replaced by an alanine (A).

Candidatus Poribacteria is an unclassified group of bacteria that was originally identified in the prokaryotic symbiotic community or holobiont of the sponge A*plysina aerophoba* (Fiesler et al., 2004). This bacteria has not been cultured up until now (Gutleben et al., 2020; Versluis et al., 2017) and the only information accessible is the genomic sequences retrieved from metagenomic analysis (Podell et al., 2019; Siegl et al., 2011; Siegl & Hentschel, 2010). High-throughput sequencing of sponges holobiont generated genomic fragments, assembled into contigs, which sequences are available in the whole-genome shotgun sequencing (WGS) database.

Interestingly, a *SELENON* gene in *Poribacteria* genomes is detected in a subgroup living in symbiosis with different sponges, including *Aplysina aerophoba*, but not in their free-living pelagic counterparts (Podell et al., 2018). This observation suggests that SelenoN activity contributes to bacteria-sponge symbiotic interactions. Considering its limited phylogenic

distribution, it is likely that the bacterial SelenoN gene was acquired from the animal host by horizontal gene transfer, similar to other eukaryotic genes (Kamke et al., 2014).

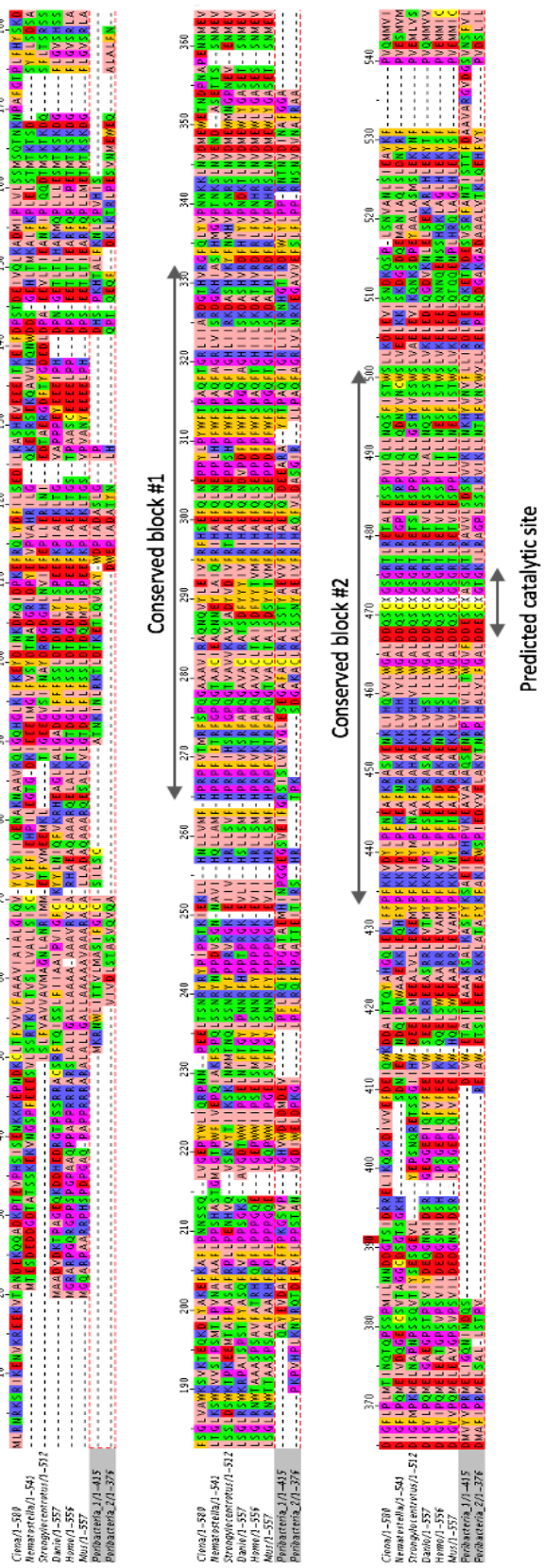
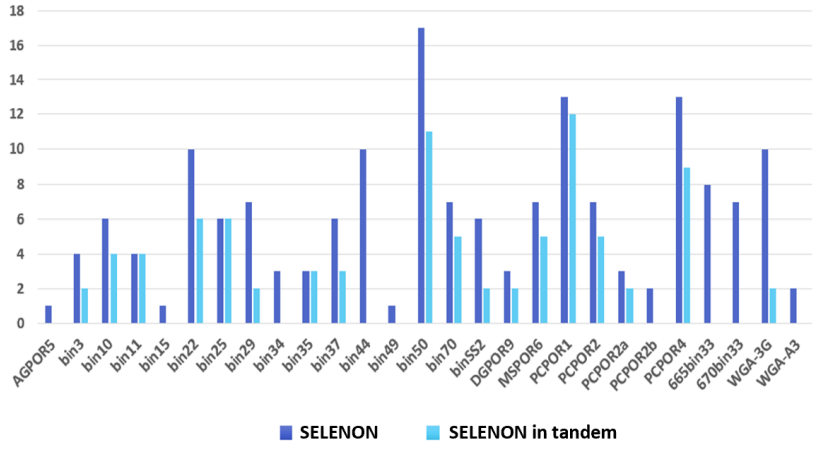


Figure R1: Multiple sequence alignment showing the sequence conservation between the eukaryotic and prokaryotic SelenoN proteins: SelenoN sequences from six animals belonging to various phylogenic groups (tunicate *Cional intestinalis*, sea anemone *Nematostella vectensis*, sea urchin *Strongylocentrotus purpuratus*, zebrafish *Danio rerio*, mouse *Mus musculus*, and Human *Homo sapiens*) were aligned with the ones of two representative *Poribacteria* (highlighted with a dash red box). Functional domains and blocks of conserved residues previously identified in animal SelenoN are indicated (TM=Transmembrane domain; Ca2+ BD=Calcium binding domain). Multiple sequence alignment was obtained through the Muscle Tool (https://www.ebi.ac.uk/Tools/msa/muscle/) and modified with the Jalview application.

2. Contigs search:

The whole-genome shotgun sequencing (WGS) blast search using the Human SelenoN as a query identified 33 *Poribacteria* strains and 149 contigs containing 249 *SELENON* genes. However, redundant contig sequences were detected due to the multiple sequencing of the same strain. After manual curation, up to 72 contig sequences have been deleted to end up with 23 individual strains, 111 curated contigs, and 166 *SELENON* genes. These numbers imply that multiple *SELENON* genes are represented within each strain. Indeed, several *SELENON* copies, up to 17, were identified in the different *Poribacteria* strains. Moreover, in several cases, several *SELENON* genes that were found in the same contig, are organized in tandem repeats (Figure R2A), with up to five genes contained in one repeat. Collectively, these results suggest that SelenoN protein is abundantly expressed in *Poribacteria*, at least under certain conditions.

A)



B)

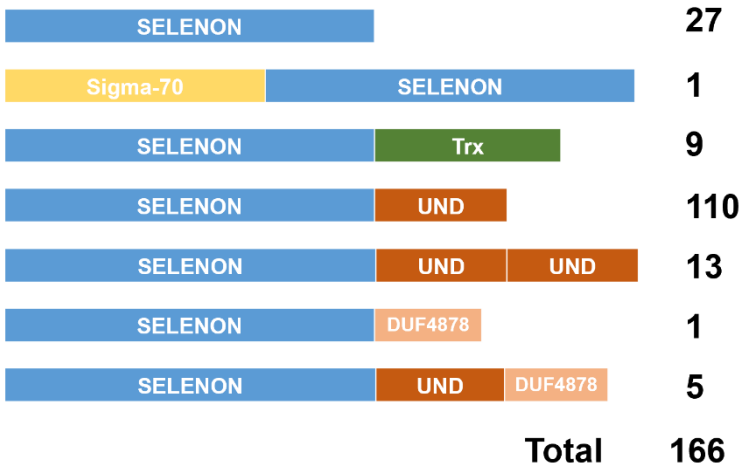


Figure R2. Analysis of *SELENON* genes identified in *poribacteria* metagenomic data. A) The copy number of *SELENON* genes identified in each *poribacteria* genome (dark blue bar) and the number of *SELENON* genes in tandem repeats in each strain (light blue bar). **B)** Multiple alignments of *SELENON* showed its fusion to several additional domains. *SELENON* (blue box), Sigma-70 factor (yellow box), Thioredoxin-like (Trx, green box), Unknown domain (UND, red box), conserved Domain of Unknown Function (DUF4878, orange box) are represented. The number of genes encoding each form of different strains is indicated on the right.

All of the identified sequences contained the expected Sec-UGA codon. We then got interested in the Sec insertion mechanism in this group of bacteria. We observed that the nucleotide sequence downstream of the Sec codon is particularly conserved between the *SELENON* genes, indicative of evolutive constraints on this sequence. A LogoPlot showing the consensus of the 166 *SELENON* SECIS nucleotide sequences is shown in (Figure R3A). Inspection of this sequence showed that it can fold to form a hairpin structure similar to the prokaryotic SECIS element, including a bulged U residue at position +17 downstream of the Sec codon and an apical GU dinucleotide on the top of the hairpin. These two features, characteristics of the prokaryotic SECIS element, are conserved in all SECIS sequences of all of the identified *SELENON* genes. A comparison of the identified *SELENON* SECIS with the *E.* Coli *fdhH* SECIS element is displayed in (Figure R3B). Further supporting this RNA structural model, compensatory mutations maintaining the two stem structures were observed (data not shown). In addition, the length of this SECIS motif is conserved in all *SELENON* sequences.

In addition, a search for Sec-insertion machinery in the *Poribacteria* genomes identified genes coding for the bacterial factors *SELA*, *SELB*, and *SELD*, as well as a gene for the typical bacterial $tRNA^{Ser(Sec)}$ (data not shown). Altogether, these data demonstrate that *SELENON* expression in *Poribacteria* utilizes the classical prokaryotic Sec insertion system.

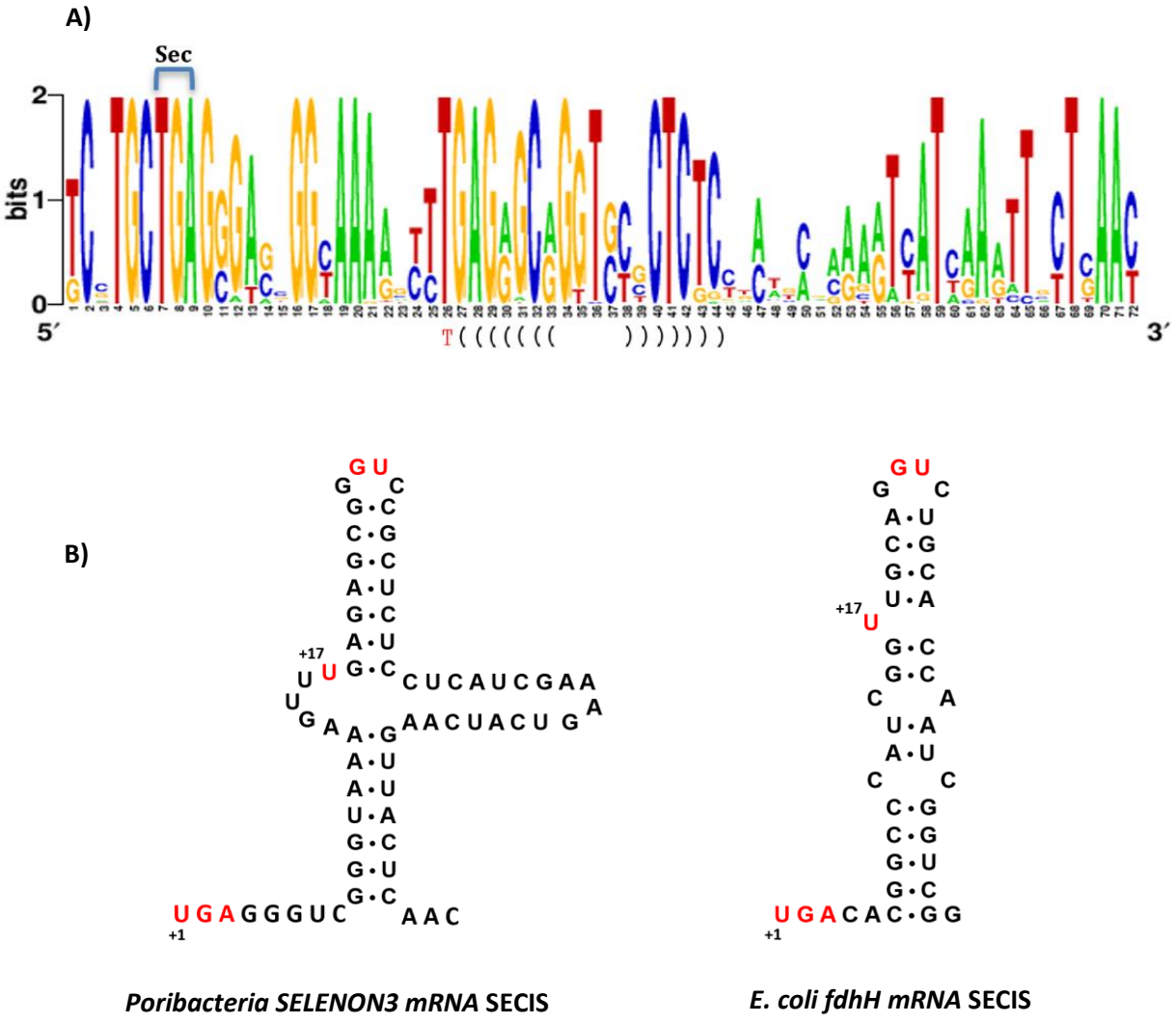


Figure R3. Structure and LogoPlot of SECIS elements. **A)** Sequence alignment identified a SECIS sequence downstream of the Sec codon. This LogoPlot corresponding to 166 aligned *SELENON* sequences shows the conserved residues corresponding to the sequences encoding the catalytic site and the SECIS motif. The LogoPlot was created by https://weblogo.berkeley.edu/logo.cgi. **B)** Structure of the folded SECIS elements on the mRNA of the newly amplified *SELENON* sequence (shown in the upcoming amplification section) from *Poribacteria* is displayed on the (left), compared with *E. coli fdhH SECIS on the (right).* The UGA encoding selenocysteine (Sec), the conserved "U" residue located at +17 downstream of the UGA codon, and the apical GGU sequence are highlighted in red. Nucleotides constituting the apical stem are depicted with brackets on the LogoPlot representation. *E. coli fdhH* SECIS created by Microsoft power point and adapted from (Kim et al., 2009)

3. Analysis of the diversity of *Poribacteria* SelenoN proteins

The protein sequences deduced from the translation of all *SELENON* open reading frames (ORFs) were aligned and clustered using the MUSCLE multiple sequence alignment tool, which creates a neighboring tree. This tree and the sequence alignment were analyzed by Cluster Database at High Identity with Tolerance (*CD-HIT*), a hierarchical clustering and comparison method to compute the *SELENON* genes into subgroups or subfamilies according to their sequence relatedness and the presence of conserved protein domains. This protocol identified 27 contigs containing *SELENON* genes coding for proteins with no additional domain, called unfused-SelenoN, a protein similar to the animal SelenoN. In addition, we identified several translated *SELENON* ORFs fused either at C- or N-terminus with different additional domains, such as a stress-related transcriptional factor Sigma-70 (σ^{70}) and a Thioredoxin-like domain (Trx). However, the large majority of *Poribacteria* SelenoN appeared to be fused with an Unknown domain (UND), present either in one or two copies or with another conserved Domain of Unknown function DUF4878. In five cases, these two domains were combined. A representation of the different *Poribacteria* SelenoNs and the number of the different fusions encountered in the genomes of the different strains are indicated in (Figure R2B).

4. Operon-mapper analysis

Analysis of genes context and genomic organization in prokaryotic can provide valuable functional information, based on the principle of "guilt by association", a particularity of the bacterial phylum, implying that genes present within one genomic unit generally contribute to related functions or metabolic pathways (Shmakov et al., 2019). Indeed, genes residing in the same operon are co-regulated and co-transcribed as they are controlled by a common promoter, and the conservation of a block of genes in different species or strains reveals a co-evolution process referred to as synteny. Laurence Despons has developed a bioinformatics strategy that allows a systematic and automatic analysis of all *SELENON* genes' genomic context in the different *Poribacteria* strains.

Using the Operon-mapper protocol (Taboada et al., 2018), genes located in the vicinity of *SELENON* (10,000 nucleotides upstream and downstream) were extracted and classified according to their common COG (Clusters of Orthologous Genes) and UniProt knowledgebase (UniprotKB) annotations. Next, they were analyzed based on two criteria: first, their number of occurrences and distribution in the vicinity of *SELENON*, second, their presence within *SELENON* predicted operons.

The COG annotations identified two clusters of genes that represent the highest occurrences in the whole analysis referred to as ABC-type-transporters and WD40/WD40-like repeats. However, this high-frequency score could be due to their random occurrences in the *SELENON* neighborhood, since those groups represent large families of genes with numerous copies spread out all over the genome. In addition, the presence of these two genes in operons with *SELENON* is very scarce, since only four cases of WD40-like repeat genes were predicted in operons with *SELENON*.

Significantly, operon mapper identified two clusters of genes in *SELENON*-containing operons, one referred to as ''Redoxin-coding genes'' (*RDX*) and the other one is ''Inosine 5' monophosphate dehydrogenase'' (*IMPDH*). The *IMPDH* gene codes for a NAD-dependent oxidase, a key enzyme of de novo purine nucleotide biosynthesis. However, this gene was found at a variable distance from *SELENON* (130 ± 220 nucleotides) with several genes inserted between the two. On contrary, the *RDX* was always located next to *SELENON*. In addition, *RDX* was predicted to reside in an operon with one specific form of the *SELENON* gene, the one coding for the SelenoN-Trx fusion. This particular genomic organization was intriguing since SelenoN

appeared to be physically linked to two potential redox proteins or domains. Indeed, *RDX* also encodes a protein containing a Trx-like domain containing the classical reductase active CXXC motif (where C stands for cysteine and X for any amino acid), CGFC in case of *RDX*. On contrary, the Trx-like domain fused to SelenoN presents a mutated form of this motif, NXXC (where N stands for asparagine), NSFC in this case. Overall, this striking organization of one *Poribacteria SELENON* form reinforces the notion that SelenoN catalytic activity is taking place in oxido-reductive context.

What can we understand from the organization in the operon of SELENON-Trx. One explanation is that redox proteins are exchanging electrons, so one possibility is that they have different substrates and catalyzing different reactions in a particular pathway, or this is a chain of electron transport. The NSFC is more likely to be glutathione reductase GR, and the GR playing role of the 2^{nd} Cys that is missing in the NSFC, but if this is true, we will be having an oxidized glutathione that needs to be reduced, and the CGFC domain is reducing the oxidized glutathione. Or another hypothesis is that the SCUG reduces the oxidized glutathione, but the U needs to be regenerated, so the redoxin-coding gene does that.

Another cluster, *ROG3500* was found consistently in an operon with the gene coding for the unfused-SelenoN form, the two genes are often on the same strand and interspaced with a short sequence of about twenty-one nucleotides. *ROG3500* corresponds to a group of Restricted Orthologous Genes (ROG) that codes for a short hypothetical protein. Both *ROG3500* nucleotidic and translated amino acid sequences are strongly conserved in the different *Poribacteria* strains.

Next, we got interested in the genomic context of the two most characteristic *SELENON* operons: *SELENONs* and the genes located in their vicinity were mapped on the different contigs and schematized according to their COG annotations; in addition, operons predicted by Operon-mapper are also indicated (see Figure 4A for the genes in the vicinity of the unfused-SelenoN and 4B for the genes in the vicinity of the SelenoN-Trx fusion). The two clusters clearly appeared to be part of a block of conserved genes or synteny, although with several genomic reorganizations, such as the inclusion or deletion of possible small coding units between the most conserved genes. Concerning the operon prediction, both *SELENON* operons are sometimes predicted to include

additional genes, but only the Rdx/SelenoN-Trx or unfused-SelenoN/Rog35000 pairs are consistently predicted together, indicative of their close proximity.

In the following studies, we have focused on the information that will allow us to amplify the *SELENON* gene from the *Poribacterial* genome. Therefore, we have chosen the two candidate genes that are present in an operon with the *SELENON* gene: (i) The *RDX* in an operon with *SELENON*-Trx, and (ii) the *ROG3500* in an operon with unfused-*SELENON*. Our purpose was to amplify these *SELENON* genes in an operon with their related genes from *Poribacteria* DNA prepared from *Aplysina aerophoba* extracts and to express these proteins in a heterologous bacterial system.

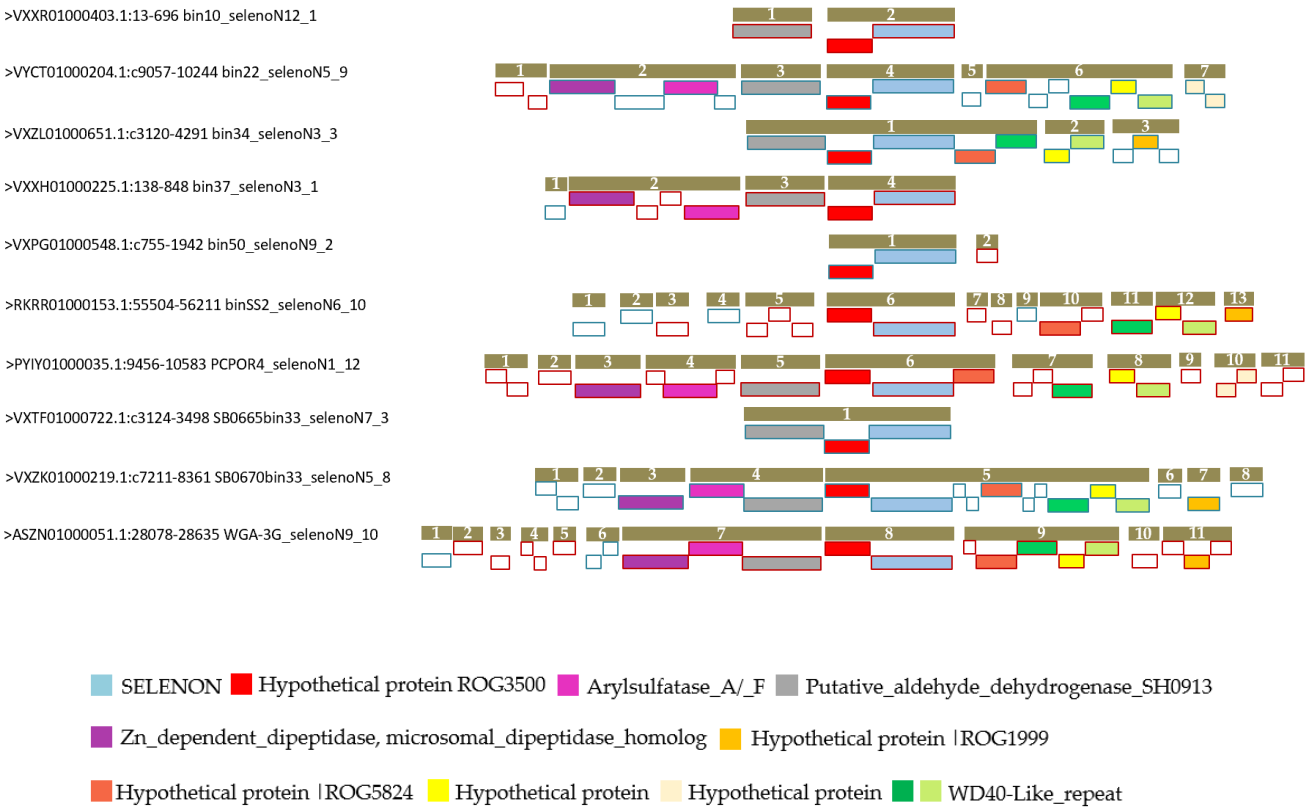


Figure R4A. Contigs map for genes clusters organized in the vicinity of unfused-*SELENON*. Genes were coloured according to the predicted COG function as shown in the lower panel. Empty blocks correspond to genes coding for “hypothetical proteins”. Conservation of genes in the vicinity of *SELENON* (blue box) indicates a synteny over different *Poribacteria* strains. The organization of genes in operons was predicted using Operon mapper and genes contained in the same operon unit are highlighted with a golden bar. Note that the *ROG3500* (red box), a gene of unknown function, is always in an operon with the unfused-*SELENON* in ten strains of *Poribacteria*. Blocks outlined in blue and red are on opposite strands. The name of the poribacterial strains and the contigs ID are indicated for each contig. The ID of each gene is like that of "bin10_selenoN12_1" where bin10 is the name of the contig having the SELENON12 gene and the number "_1" is the position of the gene on the contig.

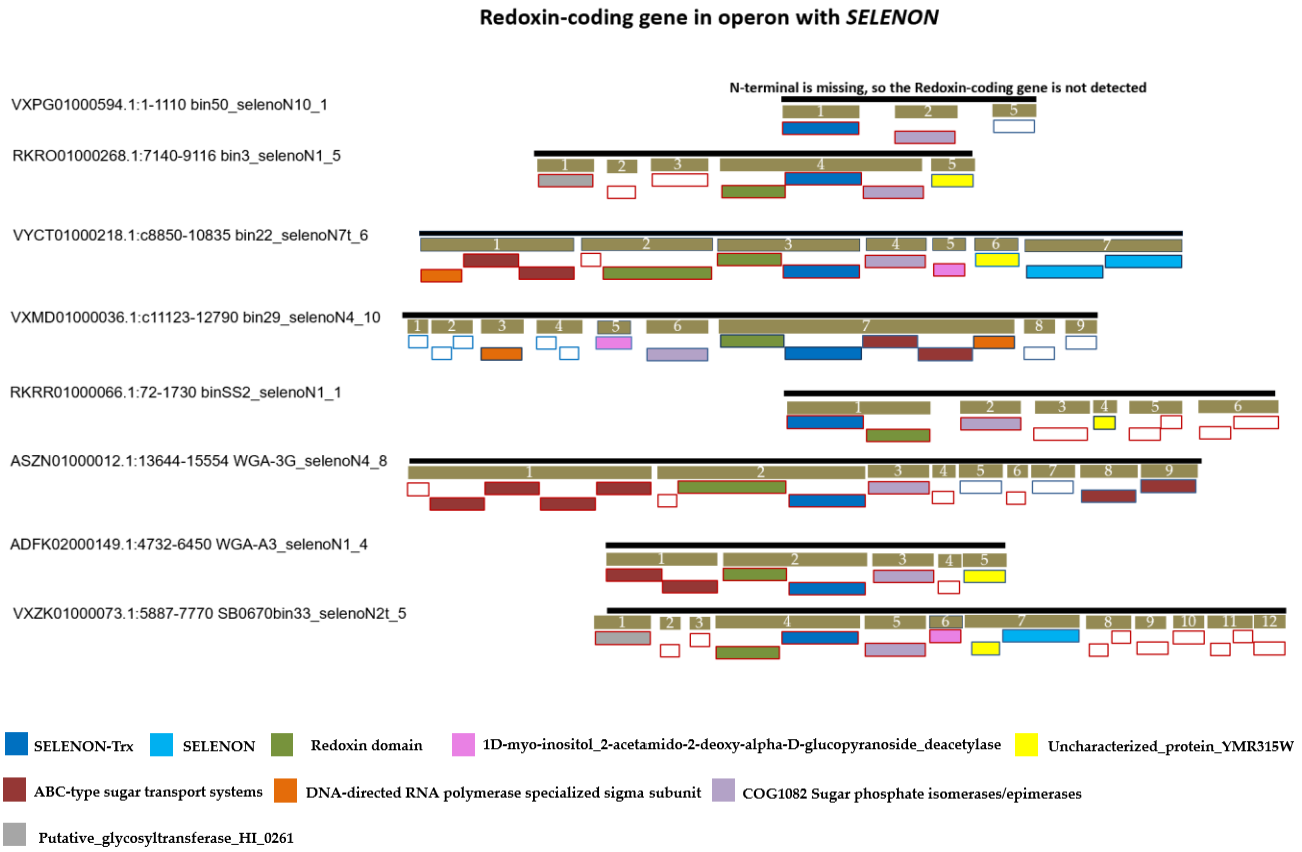


Figure R4B. Contigs map for genes clusters organized in the vicinity of *SELENON* fused to Thioredoxin-like domain. Genes were coloured according to the predicted COG function as shown in the lower panel. Empty blocks correspond to genes coding for “hypothetical proteins” or genes that show no interest in the study. Conservation of genes in the vicinity of *SELENON* (dark blue box) indicates a synteny over different *Poribacteria* strains. *SELENON* in (light blue) is unfused to no additional domain. The organization of genes in operons (golden bar) was predicted using Operon mapper. Note that the redoxin-coding gene (green box), is always in an operon with *SELENON-Trx* in eight strains of *Poribacteria*. Blocks outlined in blue and red are on opposite strands. . The name of the poribacterial strains and the contigs ID are indicated for each contig. The ID of each gene is like that of "bin50_selenoN10_1" where bin50 is the name of the contig having the SELENON10 gene and the number "_1" is the position of the gene on the contig.

1. Identification of a *SELENON* homologous gene in *Candidatus Poribacteria*

High-throughput sequencing of marine sponges holobiont → generated fragments → assembled into contigs, available in the WGS database.

BLAST search on the WGS → human SelenoN protein as a query → identified SelenoN in *Candidatus Poribacteria.*

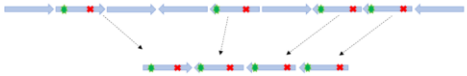
Poribacteria SelenoN degree of homology with the human protein: 31% identity and 48% similarity.

Method

2. Contigs search

Results

Previous studies created → *Poribacteria* genome database by SAGs and MAGs.
In this study: tBLASTn search on the WGS contigs database on *Candidatus Poribacteria* (taxid:265317) → retrieving of all contigs-containing *SELENON* homologous gene by (Python 2.7.17) script + ORF-finder tool → recognizing the Sec-TGA reprogrammed inframe codon + possible next stop codon for translation termination.



Identification of 23 *Poribacteria* strains, 111 curated contigs, and 166 *SELENON* genes. multiple *SELENON* genes are represented within each contig and are organized in tandem repeats.

3. Analysis of the *SELENON* genomic context in the different *Poribacteria* strains

Method

All retrieved *SELENON* ORFs → translated → aligned and clustered using the MUSCLE → creating a neighboring tree → analyzed by *CD-HIT* → *SELENON* genes orgainzed into subgroups or subfamilies according to their sequence relatedness and the presence of conserved protein domains.

Results

Clustering identified 27 contigs-containing *SELENON* genes coding for proteins with no additional domain (unfused-SelenoN), and genes coding for proteins fused either at C- or N-terminus.

4. Operon-mapper analysis

Method

Poribacteria SELENON ORFs →submitted → Operon-mapper → prediction of genes in operons with *SELENON* (10,000 nucleotides upstream and downstream) + prediction of the proteins and their functional UniprotKB or COG descriptions →then, genes in the vicinity of *SELENON* were clustered using Clustal Omega and CD-HIT to find consistent genes represented in *SELENON* environment.

Results

Two clusters of genes were identified in the vicinity of *SELENON* → ABC-type-transporters and WD40/WD40-like repeats.
Two clusters of genes were identified in *SELENON*-containing operons → *RDX* in an operon with the *SELENON-Trx* fusion + *IMPDH*
Another cluster, *ROG3500* in an operon the unfused-*SELENON* form.

WGS: Whole-genome shotgun sequencing
BLAST: Basic Local Alignment Search Tool
SAGs: Single-cell amplified genomes
MAGs: Metagenome-assembled genomes
CD-HIT: Cluster Database at High Identity with Tolerance
MUSCLE: multiple sequence alignment tool
COG: Clusters of Orthologous groups of proteins
UniprotKB: UniProt knowledgebase annotations
RDX: Redoxin-coding genes
IMPDH: Inosine 5' monophosphate dehydrogenase

Figure R5. A brief blueprint of the bioinformatics analysis of the *SELENON* genome in *Poribacteria* and the results.

II . Molecular Analysis

II . Molecular Analysis

1. Partial *Poribacteria*-16S rRNA Amplification

We worked on the *Aplysina aerophoba* samples that were collected previously as explained in the methods section. To validate the presence of *Poribacteria* in the sponge sample, we first amplified the *Poribacteria* 16S rRNA.

The 16S rRNA gene has a slow rate of evolution and is highly conserved in prokaryotes. Therefore, it is largely used in the identification, classification, and quantification of different microbial species in complex samples, such as in metagenomic extraction from environmental samples. The 16S rRNA gene consists of ten conserved regions (C1-C10) surrounding nine variable regions of varied conservation (V1-V9). Nevertheless, the conserved regions have a considerable degree of variation among different microorganisms and are not 100% conserved as shown in (Martinez-Porchas et al., 2017). Universal PCR primers targeting the conserved regions were designed to allow a comprehensive amplification of the 16S rRNA genes of multiple microorganisms, belonging to different taxa from the same sample. An alignment of the 16S rRNA C1 region between *E. coli* and eight *Poribacteria* strains represented in the WGS database displayed several differences. Hence, we modified the universal 27f primer with the best-fit consensus of the alignment to match the difference in *Poribacteria* species. On the other hand, variable regions are more quickly evolving regions and are targeted to differentiate among different genera and species. In fact, different studies examined which variable regions are more preferable over others in discriminating among different species as seen in (Bukin et al., 2019; Yang et al., 2016), but no single consensus on which one is more favorable, and therefore, no single region can differentiate among all bacteria and comparison between different variable regions is required for species identification.

Here, we applied a nested-PCR strategy using two primer sets as was explained in the method section (Figure M2). Amplified fragments were cloned into a plasmid and sequenced. The obtained sequences were aligned to *Poribacteria* reference sequences from the WGS database using the Basic Local Alignment Search Tool (BLASTn). Results are displayed in (Table R1) and the amplified 16S rRNA sequence is available in (Table R2). Analysis of eleven clones revealed that nine of them contained identical sequences, highly similar to partial 16S rRNA of the candidate

phylum *Poribacteria*, as the two other clones corresponded to a different phylum of gram-positive bacteria. An alignment of the amplified partial *Poribacteria*-16S rRNA sequence with the *E. coli* 16S rRNA sequence is displayed in (Figure R6), highlighting the sequence conservation and divergences, confirming the specificity of the primer pairs used for the amplification.

Therefore, we used the *Poribacteria*-specific primers sets previously designed in (Fieseler et al., 2004), POR389f/POR1130r to target the amplification of V4, V5, and V6 regions, and in (Podell et al., 2019), the universal primer set 515Fsp/806Rsp to target the amplification of the V4 region of *Poribacteria* species.

Table R1. Best BLASTn hits of the 16S rRNA sequences amplified from an *Aplysina aerophoba* sponge sample.

Clones	Top BLAST Hit	Accession ID	Query coverage	% Identity
CP1-CP5 CPO1/CPO2 CPO4/CPO5	*Poribacteria* WGA-4G POR4G, and many other strains	AQPC01000005.1	100%	98.86%
CP6	*Cutibacterium* acnes from phylum Actinomycetota	**AP022845.1**	100%	100%
CPO3	*Propionibacterineae* from phylum Actinomycetota	JF733146.1	97%	94.03%

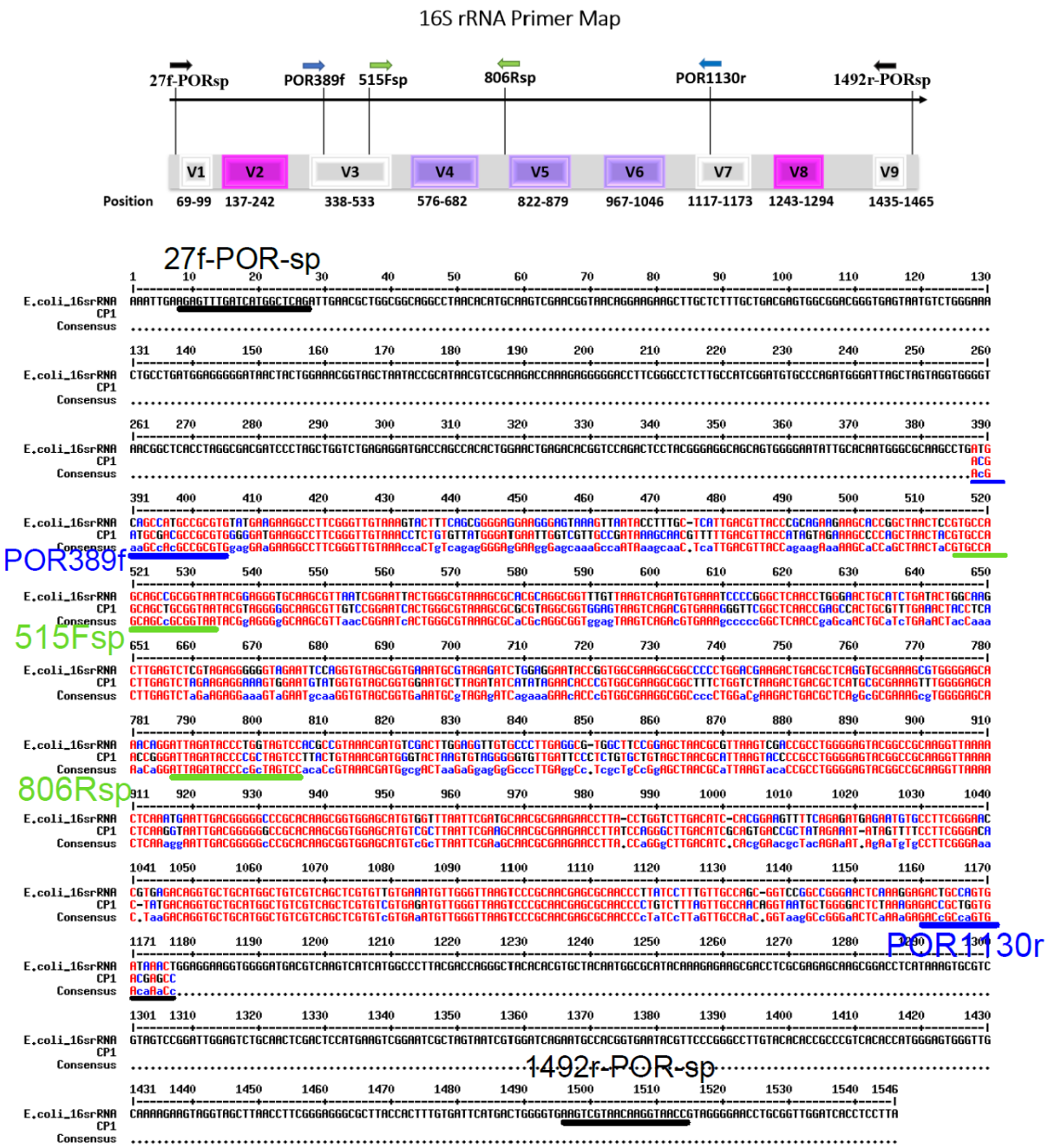


Figure R6. A comparison between *E. coli* and *Poribacteria* 16S rRNAs sequences amplified in this study. 16S rRNA primers specific to *Poribacteria* were designed in accordance with the variable regions positions as shown in the upper panel. All primers used in the amplification are marked on the alignment. The specificity of the poribacterial-specific primers can be observed from the sequence alignment in comparison to *E. coli*. The variable V4, V5, and V6 regions located between the *Poribacteria*-specific primers (in blue) noticeably vary between the two species. *Escherichia coli* 16S rRNA complete sequence reference: **GenBank Sequence ID: J01859.1.** The 16S rRNA primer map shown was created by Microsoft PowerPoint and adapted from (Shahi et al., 2017).

2. *SELENON-ROG3500* Operon Amplification.

The unfused-*SELENON* gene in an operon with the downstream *ROG3500*, spaced by an intergenic distance of 21 nucleotides was amplified from the metagenomic DNA extract of the *Aplysina aerophoba* sample.

The main challenge in this experiment resides in the complexity of the starting material since it was required to amplify one specific *SELENON* gene among multiple copies in one strain, among many poribacterial strains, which are part of a large holobiont containing numerous bacterial species, extracted from the sponge tissue. Therefore, we combined different simple techniques that are useful in increasing specificity, sensitivity, and yield by reducing the nonspecific hybridization of the primers:

- Nested PCR is a simple technique that reduces nonspecific amplification of the DNA region. The process involves two primer sets and two sequential PCR runs. Amplicon resulting from the first run is used as a template in a second PCR run. One downside is that the successive manipulations of the amplicon increase carryover contamination. This explains why in some nested PCRs we observed contamination in the second PCR run and not in the first run. However, the efficiency of the method outweighs this issue.
- Touchdown PCR (TD-PCR) is another simple PCR optimizing technique. TD-PCR employs a cycling program with varying annealing temperatures to increase the specificity of primer annealing. Two stages are programmed in the TD-PCR. The annealing temperature in the first stage is set up between 5-10°C above the calculated Tm of the primers and decreases decrementally by 1–2°C/cycle to reach on average 5°C below the Tm. This decremental decrease in the temperature reduces the chance of non-specific binding and undesired amplification.
- Inverse PCR, also called Outward-directed PCR or inside-out PCR, is a technique utilized in amplifying unknown DNA sequences flanking a known DNA sequence. The primers are designed from the known sequence and pointing outward. The template DNA is digested by restriction enzymes and recircularized by intramolecular ligation. The amplified amplicon should contain the restriction site for the enzyme used for template digestion. This site represents the junction of the ligated recircularized DNA fragment.

The amplification strategy was explained in the methods section (Figure M4). From the ten different restriction endonuclease digestions of the metagenome, as described in the methods section, a complete *SELENON* sequence was obtained only from the EcoR1 endonuclease digestion by the inverse PCR approach. A schematic map of the amplification process is illustrated in (Figure R7). The newly amplified *ROG3500* gene in alignment with ten *ROG3500* genes from BLAST search is displayed in (Figure R8) showing a strong similarity between the newly and previously identified sequences, with only 11 mismatches with the sequence of the *Poribacteria* strain WGA-3G over a 696 nucleotides long sequence. Similarly, the newly amplified *SELENON* gene, called *SELENON3*, in alignment with the top six hits in the BLAST search is shown in (Figure R9), and also demonstrated strong conservation of this gene. The nucleotidic sequences of both *SELENON* and *ROG3500* are available in (Table R2).

Of note, although the *SELENON* sequence of the WGA-3G is not complete in the WGS database, it has the highest percent identity of the partially covered sequence among the other sequenced strains (Figure R9). This corresponds with the results of the amplified *ROG3500*. Therefore, we concluded that both *SELENON* and *ROG3500* were amplified from the same poribacterial strain, but different from the previously sequenced *Poribacteria*, indicating that our amplification operated from one new strain not identified in the metagenomic analyses.

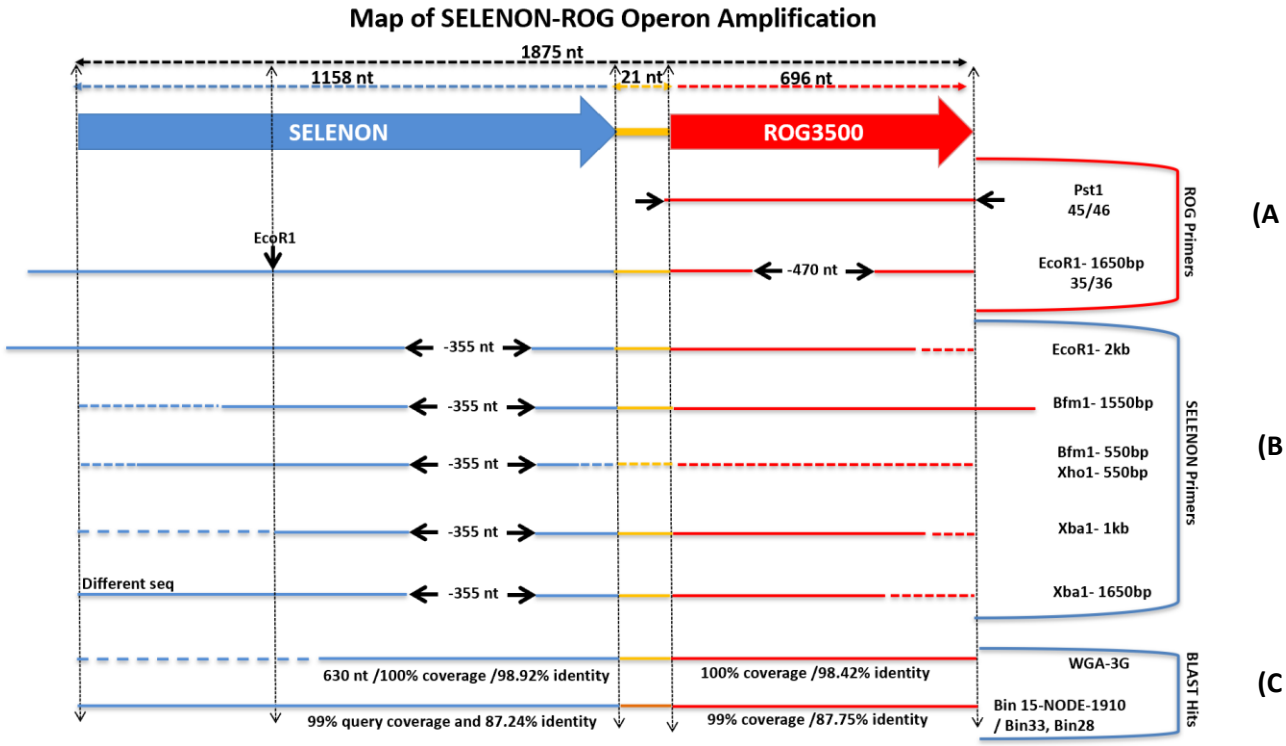


Figure R7. *SELENON-ROG3500* Operon amplification map. The upper panel shows the operon containing the *SELENON* gene, the intergenic space (in yellow), and the downstream *ROG3500*. The length of the two genes and the spacer are indicated. **A)** Using primers designed based on the *ROG3500* sequence we amplified the complete sequence of *SELENON* plus the intergenic distance by nested-inverted PCR. The complete *ROG3500* was then amplified by conventional PCR. **B)** Same nested-inverted PCR approach was used on the DNA template digested with other different restriction enzyme digestion to compare and verify the amplified *SELENON* in each amplification attempt. Each time we have the same *SELENON* sequence but an incomplete sequence. **C)** The highest sequence coverage and percent identity of BLAST search is shown.

Table R2. Nucleotide sequences of different amplified genes from the metagenomic extraction.

Poribacteria-16S rRNA	ACGATGCGACGCCGCGTGGGGGATGAAGGCCTTCGGGTTGTAAACCTCTGTGTTATGGGATGAA TTGGTCGTTGCCGATAAAGCAACGTTTTTGACGTTACCATAGTAGAAAGCCCCAGCTAACTACG TGCCAGCAGCTGCGGTAATACGTAGGGGGCAAGCGTTGTCCGGAATCACTGGGCGTAAAGCGCG CGTAGGCGGTGGAGTAAGTCAGACGTGAAAGGGTTCGGCTCAACCGAGCCACTGCGTTTGAAAC TACCTCACTTGAGTCTAGAAGAGGAAAGTGGAATGTATGGTGTAGCGGTGGAATGCTTAGATAT CATATAGAACACCCGTGGCGAAGGCGGCTTTCTGGTCTAAGACTGACGCTCATGCGCGAAAGTT TGGGGAGCAACCGGGATTAGATACCCCGCTAGTCCTTACTGTAAACGATGGGTACTAAGTGTAG GGGGTGTTGATTCCCTCTGTGCTGTAGCTAACGCATTAAGTACCCCGCCTGGGGAGTACGGCCG CAAGGTTAAAACTCAAGGTAATTGACGGGGGGCCGCACAAGCGGTGGAGCATGTCGCTTAATTC GAAGCAACGCGAAGAACCTTATCCAGGGCTTGACATCGCAGTGACCGCTATAGAAATATAGTTT TCCTTCGGGACACTATGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGT TAAGTCCCGCAACGAGCGCAACCCCTGTCTTTAGTTGCCAACAGGTAATGCTGGGGACTCTAAA GAGACCGCTGGTGACGAGCC
Poribacteria-SELENON FRIHAQ domain in blue. Catalytic site in red.	ATGATGGCAGACTTTAGTGACGTTCTTAGCAGTACAGCGAGTGTGCAAGTGCAAGCATGTTGGG AGCCAATCGAGGATGCCACCTATAACCTGCATCAACTAACGGAAGACCAGTTCGATCGGGTCAA AACAAGATGTCCAGAACTAGTGAATATGGAATGGGAACAAGCGTTAGAACTTTTCAATCTCAAA CCTGTGCACCCATTGAAGAACTACATCCAAACCGACTTTCAAGCTTTTCTACCGCCTTCGACTG CTAATGTAGGGGATATTTGGGAACTTGACTCAGAAAAGATTCTCCCATTCCTCCGCCAGTTTCA CTCGGGCGCAACAACGGAGATCACCATCTTTTCGCACAGAACACCTAAATCTGACGGAGCAAAA GCGTGTTTGCGAGCGATCTCGCCAGACTATGCTGAAATTGTTTTCCGAATTCATGCACAGTTCG TGCTAAACGCACCCGGTGTTCGTTTGTTACCAGCACAGTTTGCAGGACGACTGGTGCTGAATCG CACGGAAGGAGCAGTTGTTGCGTTTTCTCTGTTTCTTCCCTCACGCAATAGCAATGTAGATGTT AATGCGTTCAAGGCAGCAGACATGGCTTTTATTCCTCGTATGGAACTTTCTACTTTATCAAATG AGCCTATACACGAGATAATATGGGAAACAACTATCACTGAAGAAGAAGCCCGCAAAAAATTAGC AACTGCGCTCTACAAGTTCGCTAAAATTAAGTGGGTACCAATTGAAGATGCGGTTGAACTTGGG AAAGGAACAAATCGTCCTATCCATGCGTTAGTTTTATTTGGTGCCCTTGATGATGAATCGTGCT GAGGGTCGGGTAAAAGTTTGAGAGCGGGTCCGCTCTCCTCATCGAAAGTCATCAAGTTACTCAA CACGTACTTTGTCAATGTTTGGGTGCTATTGCGAGAGTTACCCGAACTGCAGGACGGTGCCAAA GGAGCTGCTGCCGCTATCTTAGTGACAAAACTCCAAGAGCATTTTGTATATCCGGTGGACTCTC TAATTCTGACACCAGCGTTGGAGTTTGTCCAACATCTACCGACGAATGAGTTTCATAAATCTTT TCCCAGTAGTGAACGAAAATCCGAATATTTTAGGTGGCTAAAGTCATCTTTAGCGCAAGTAGAG AGATAA
Spacer of 21nt	ACACTGAAGAGGAGTCTGAAT
Poribacteria-ROG3500	ATGATAACGAAAATCGAAAAAGTCTTCCGGTCCCCCTATGCCGTGCCGAACGGTTTGCAAGTTA CGGATGACGGTTTATGGATCGTCGATCAGATTACCGATAGAGTCGCATTGGTCGAAATCACTGC AGACCCGGACTATACCGTGCCGAAACTCATCCGAGACATTCCGAGCGAATCCTCCAACACCAGC GGGATGACGTTCGGTGAAGGCGCGCTCTGGTTAGCCGCAAACGGTTCAGGCGAACGGTGGCGAG CTATCCGAGATACAGATGCCGAGACAGGCGAGATTTTCAAAGTCGATCCCGCTACCGGCGAAAC CCTCGGACGCTACCCGATACCCGATGGCGGCGGCACGCACGGCATCGAATACGACCCCTACGAT GAAGGACATATCTGGGTACAAACGCTCAAGAATCAGGTCACACATAAAATGAGAACTTCGGATT GGTCTATACAAAAAACGCTGCCGTTGCCGCATGGACGTGGACACGGCATGGTACGCGTTGAAGA TGGACTCTGGTCAACCCACACCTCTGATAGGGTCATCGTCAAACTCGACCTTAAAGACGGCACA CCGCTTGACGTGATTGAAGTTCCCGAAGATTGCCCCGAACCCCACGGCTTATCCATCTACGGCG ACGATTTCCTCTATTGCGACGCCGCCTCTGGCTGGGTAGCAAAAATTACATGGTGA
Abbreviation. nt: nucleotide	

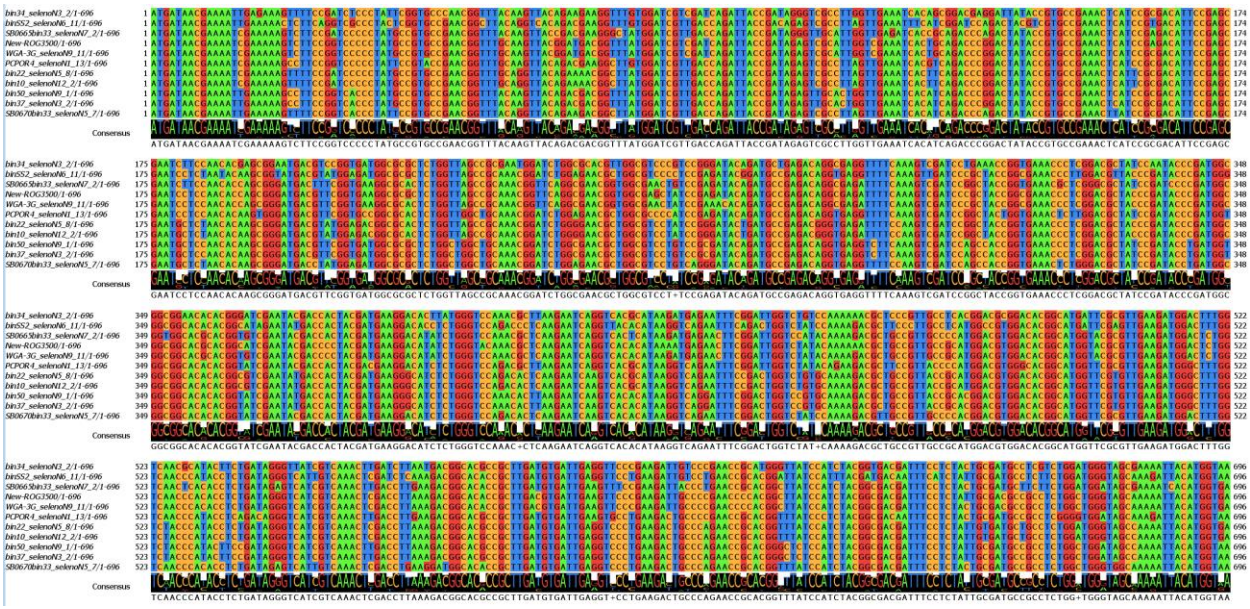


Figure R8. *ROG3500* nucleotides multiple alignments. The newly amplified *ROG3500* gene in alignment with ten *ROG3500* sequences stored in the WGS database. Amplified *ROG3500* from this study shows the highest match with WGA-3G poribacterial strain with 100% sequence coverage, 98.41% identity, and 11 mismatch. The *ROG3500* sequences were extracted by ***Basic Local Alignment Search Tool*** (BLASTn) search on *Candidatus Poribacteria* (taxid:265317) whole-genome shotgun contigs (WGS) from the below listed *Poribacteria* strains.

>SB0665bin33_selenoN7_2 ROG3500 2956_665_bin_33.
>WGA-3G_selenoN9_11 ROG3500 2974_WGA-3G.
>binSS2_selenoN6_11 ROG3500 2828_binSS2.
>bin34_selenoN3_2 ROG3500 1023_bin_34.
>PCPOR4_selenoN1_13 ROG3500 800_PCPOR4.
>bin50_selenoN9_1 ROG3500 1543_bin_50.
>bin37_selenoN3_2 ROG3500 1828_bin_37.
>bin22_selenoN5_8 ROG3500 697_bin_22.
>bin10_selenoN12_2 ROG3500 4159_bin_10.
>SB0670bin33_selenoN5_7 ROG3500 394_670_bin_33.

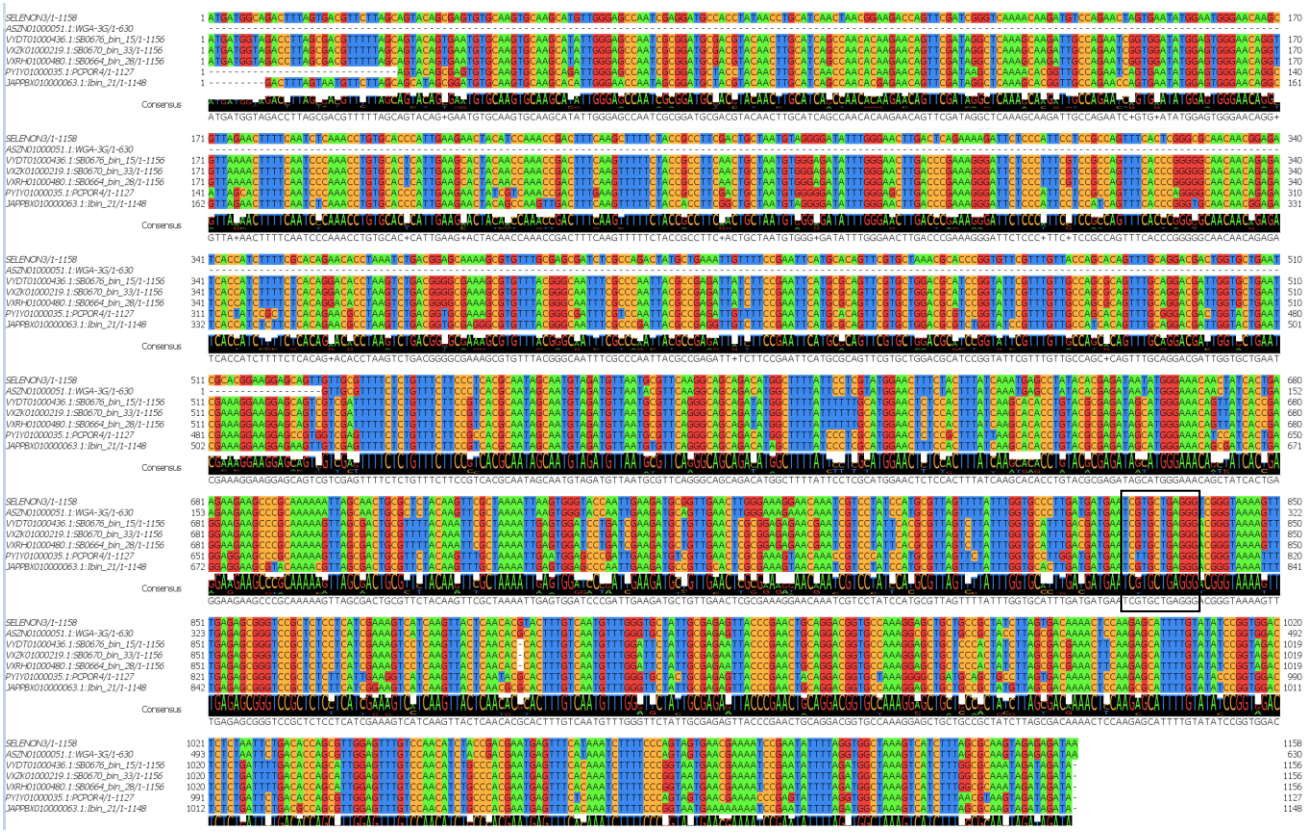


Figure R9. *SELENON* nucleotides multiple alignments. The newly amplified *SELENON3* gene in alignment with top six hits in the ***Basic Local Alignment Search Tool*** (***BLASTn***) on *Candidatus Poribacteria* (taxid:265317) whole-genome shotgun contigs (WGS) from the below listed *Poribacteria* strains. The conserved nucleotide sequence of the catalytic site TCGTGCTGAGGG is highlighted in a black box. The sequence coverage and percent of identity are displayed below :

>JAPPBX010000063.1:40662-41809 MAG:IRC3_bin_21 / query coverage:99% percent identity:87.74%
>VYDT01000436.1:5186-6341 MAG: SB0676_bin_15 / query coverage:99% percent identity:87.24%
>VXZK01000219.1:7209-8364 MAG:SB0670_bin_33 / query coverage:99% percent identity:87.24%
>VXRH01000480.1:7209-8364 MAG:SB0664_bin_28 / query coverage:99% percent identity:87.24%
>PYIY01000035.1:9471-10597 MAG: PCPOR4 / query coverage:97% percent identity:87.62%
>ASZN01000051.1:28009-28638 WGA-3G / query coverage:54% percent identity:98.89%

III. Bacterial Selenoprotein N expression.

III. Bacterial Selenoprotein N expression.

1. 6His-SUMO-bSelenoN-SCU/CG expression and purification

Bacterial SelenoN was expressed from synthetic gene that was designed based on the sequence available in the genomic database for the *Candidatus Poribacteria* taxid. Two *SELENON* genes were selected: The first one is *SELENON* fused with the thioredoxin-like domain, we named it *SELENON*1; the second one is for *SELENON* without any additional domain, we called it *SELENON*2. Nucleotide sequences corresponding to the two ORFs were optimized for *E. coli* codon usage and chemically synthesized for cloning (Genscript). In this study, we focused on *SELENON*2, the bacterial *SELENON* form most related to the animal one.

Two *SELENON*2 constructs were engineered, the wildtype form with its Sec residue and a mutant form with Cys in place of Sec. Both protein sequences were fused to a 6His-SUMO (Small Ubiquitin-like Modifier) tag to improve solubility and facilitate their detection and purification. The two constructs were expressed in *E. coli* C321Δ, a genetically modified strain engineered and optimized for selenoproteins expression (see Material & Methods, Cheng et Arnér, 2017), and the recombinant proteins were purified by Immobilized Metal Affinity Chromatography (IMAC) by applying the soluble protein fraction to a Ni-IDA column. In the total soluble extracts, the two recombinant proteins were expressed, but at a low level, and could only be detected by western blot using an antibody directed against the His-tag. The mutant Cys was mainly expressed as a protein of the expected size, in contrast to the native Sec form that was always expressed as truncated products as shown in (Figure R10). The truncation could be due to a protease degradation or a stop of translation at the UGA codon. Termination of translation was unlikely since the expression system was designed to optimize selenocysteine insertion and proved to be efficient in producing several other selenoproteins. Supplementing the media with 5µM selenite improved the expression of the Sec full-length form, but still, more than half of the expressed protein was in a truncated product and it could not be further improved by increasing selenite concentration, up to 20µM (Figure R11). Another issue was the general low expression level of the protein, including for the mutated Cys form, for which selenocysteine insertion is not a limiting factor. Several optimization strategies were tested to increase the expression by modifying growth and expression factors (temperature, medium composition, OD at induction …), but all failed. Moreover, retention on the affinity column was weak, since most of the protein did not bind to the column and remained

in the flow-through (FT) (Figure R10). These results suggested that the recombinant SelenoN2 protein was not properly folded, leading to its instability and the possible masking of the N-terminal His-tag.

a) Synthetic 6His-SUMO-SelenoN2-SCU/CG Expression without Selenite in the media.

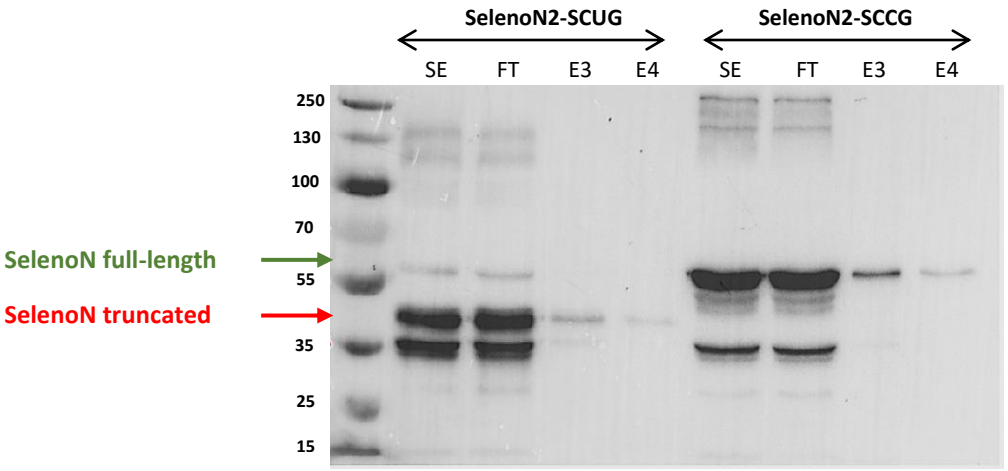


Figure R10. Expression in a heterologous *E. coli* system and purification of the synthetic bacterial SelenoN2. 8% SDS-PAGE was analyzed by Western Blot using an anti-His antibody. The wildtype full-length Sec-containing was detected at 57kDa (green arrow) in a faint band, while we observed accumulation of truncated forms at 44.4kDa (red arrow) and below. For the Cys mutant, it is mainly expressed in full-length at 57kDa with one lower band corresponding to a degraded product. Two fractions E3 and E4 eluted from the Ni-IDA purification column were loaded on the gel and analyzed in parallel. Abbreviation; S, Ser; U, Sec; C, Cys; G, Gly; SE, soluble extract; FT, flow-through; E3 & E4, eluted fractions 3 & 4.

b) 6His-SUMO-SelenoN2-SCUG Expression optimization with Selenite and different media.

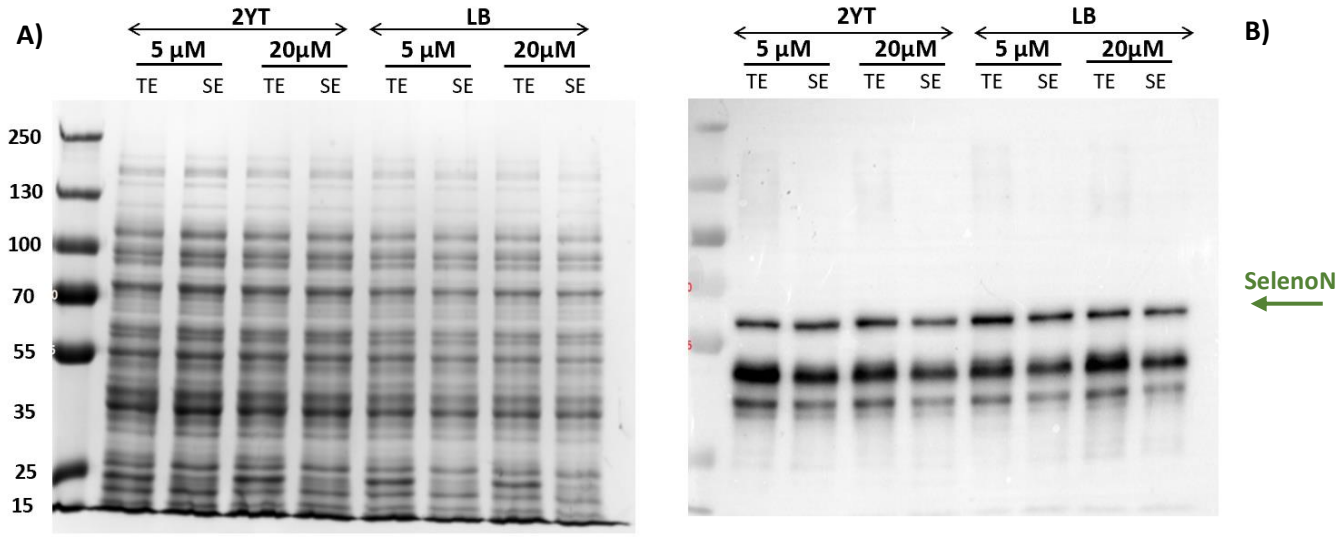


Figure R11. Optimization of the expression of wildtype SelenoN2 with increasing concentrations of selenite in LB or 2YT media. **(A)** 8% SDS-PAGE stained with Coomassie blue, the expression of the recombinant protein is very low and cannot be detected directly. **(B)** 8% SDS-PAGE was analyzed by Western Blot with an anti-His antibody and showed an improvement in the production of the full-length Sec-containing form of 57kDa (green arrow) at the two concentrations of selenite. Media composition had only a minor contribution to the expression and still, a large amount of the protein is expressed as a truncated form. Abbreviations; S, Ser; U, Sec; C, Cys; G, Gly; TE, total extract; SE, soluble extract

Actually, the synthetic sequence strategy is not fully reliable and lacks credibility, since metagenomic sequencing, from which our sequence was retrieved, only generated partial bacterial genomic sequences with a low coverage index, and therefore sequence mistakes cannot be excluded. We took advantage of the bioinformatics analyses and PCR amplification of *SELENON* genes mentioned previously to work on an original and native sequence, to improve the quality of the protein expression. We called the newly amplified sequence *SELENON3*, and this gene was expressed either alone or in its natural genomic context, in an operon with *ROG3500*. Five constructs were designed: an operon containing *SELENON3* in the Sec (U) or Cys (C) forms together with *ROG3500*, a construct for *SELENON3* alone for U and C forms, and a construct for *ROG3500* alone, as illustrated below. A sequence coding for the 6His-SUMO tag was introduced at the N-terminal of *SELENON3* and the *ROG3500* was strep-tagged at the C-terminus.

- pABC2-**6His-SUMO-SELENON3-U/C-ROG3500-Strep**
- pABC2-**6His-SUMO- SELENON3-U/C**
- pABC2-**ROG3500-Strep**

We compared the expression of SelenoN3 in an operon with Rog3500 and the individual SelenoN3 to address if the co-expression of Rog3500 and SelenoN3 could trigger the expression and stabilize SelenoN3. Analyses of expression profiles revealed that it was no impact of Rog3500 on the expression of SelenoN3 since the expression levels of SelenoN3 were the same in an operon as in an individual expression. The distinct feature of this natural SelenoN3 sequence is that the full-length expression of wildtype improved significantly compared to the synthetic sequence Selenon2 (Figure R12.A). Here, the full-length protein for the Sec form was expressed to a level equivalent to the Cys mutant, showing only a minor truncated product. This result confirmed that the engineered expression system that we used, was efficient for selenocysteine insertion into over-expressed selenoproteins. Conversely, we also examined if SelenoN3 influences Rog3500 expression, but as shown from the results, Rog3500 was better expressed as an individual gene rather than in an operon with SelenoN3 (Figure R12.B). There are two possible, although non-exclusive, explanations for this result: (i) the expression of SelenoN3 competed with the one of Rog3500, slowing down its production, or (ii) the presence of SelenoN3 destabilized Rog3500. Of

note, it was no aggregation of Rog3500 since the expression level in the total (T) and soluble (S) fractions were equivalent.

c) 6His-SUMO-SelenoN3-SCUG/SCCG expression both individually or in an operon with Rog3500

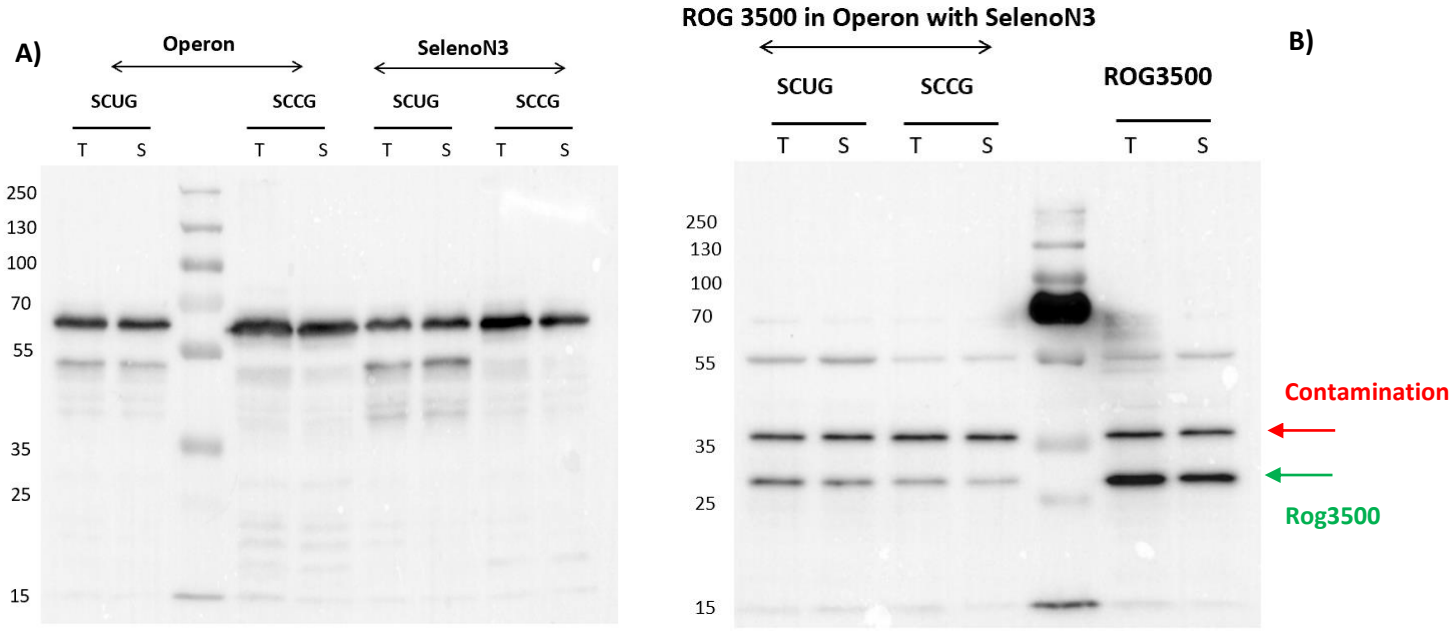


Figure R12. SelenoN3 expression in an operon and as an individual gene. (A) 10% SDS-PAGE was analyzed by Western Blot using anti-His antibody for SelenoN3 co-expressed with Rog3500 (Operon) or as an individual gene. The full-length at 55.8kDa for both Sec form (SCUG) and Cys mutant (SCCG) has improved remarkably compared to the synthetic protein, but no difference in the expression level between selenoN3 in an operon or individually was observed. **(B)** 10% SDS-PAGE for Western Blot with anti-strep tag antibodies for the detection of Rog3500 expression in an operon with SelenoN3 or as an individual gene at 26.7 kDa (green arrow), we observe that Rog3500 is better expressed as an individual gene rather in an operon. A strong cross reaction is observed at approximately 38kDa (red arrow).
Calculated molecular weight: 6His-SUMO-SelenoN3, 55.8kDa; SelenoN3, 43.2kDa; Rog3500-St, 26.7 kDa. Abbreviations: T, total extract; S, soluble extract.

d) 6His-SUMO-SelenoN3-SCU/CG Purification

The purification of Sec and Cys forms of SelenoN3 by Immobilized Metal Ion Affinity Chromatography (IMAC) as detailed in the method section is displayed in (Figure R13, R14). Coomassie staining showed an enrichment of the recombinant proteins, but low recovery. As confirmed by the western blot analysis, the protein showed low binding affinity to the column even after 1hour batch incubation at room temperature. However, retention on the column was slightly increased when the protein was incubated with the beads at 4°C overnight with rotation. This suggested that the recombinant proteins can adopt alternative conformations with different affinities for chromatographic support.

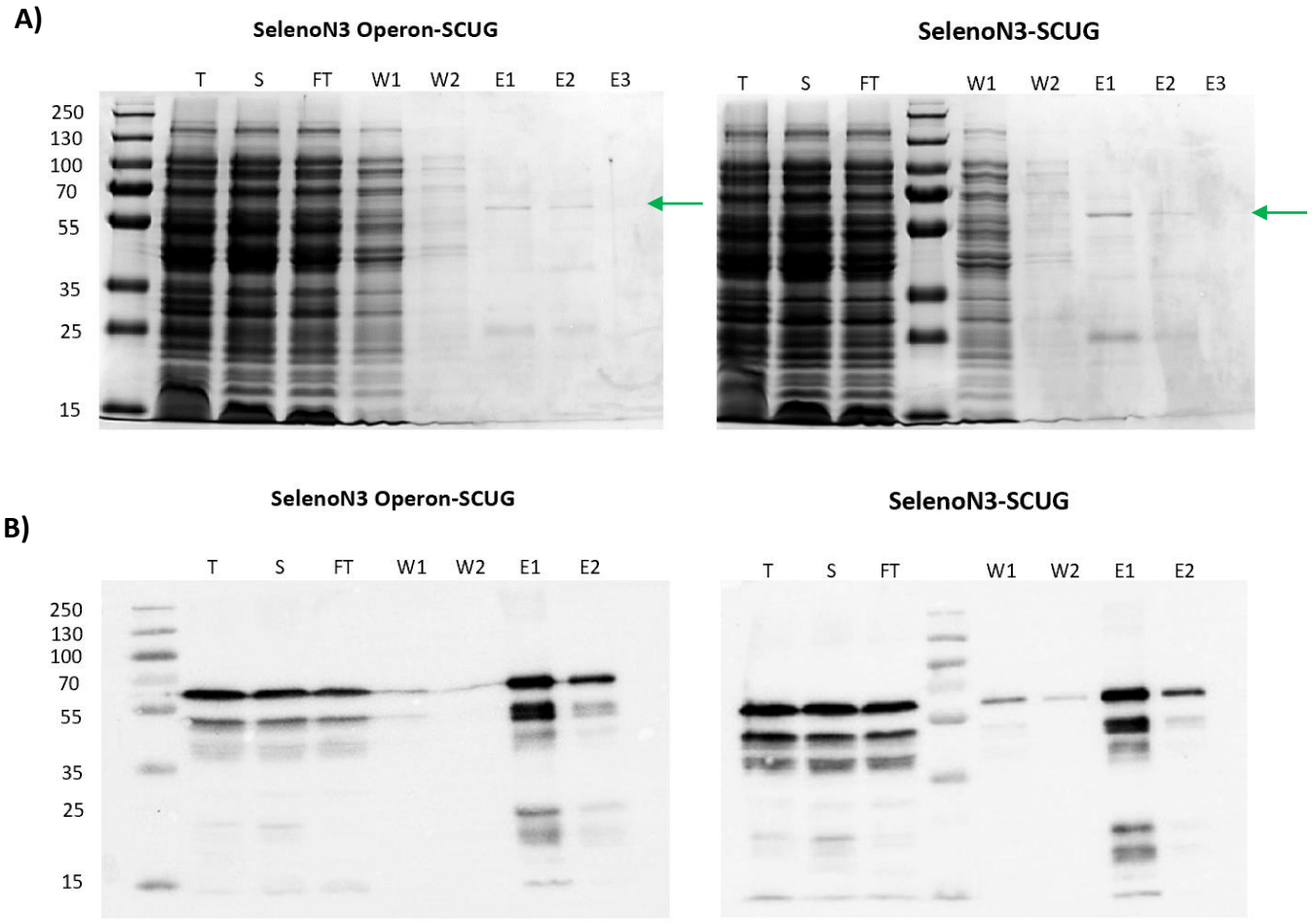


Figure R13. Purification of 6His-SUMO-SelenoN3-SCUG by Ni-IDA IMAC. (A) 10% SDS-PAGE stained with Coomassie blue, the purified SelenoN3 appeared as a very faint band in the eluted fractions and is not detected in the total protein extract. (B) 10% SDS-PAGE for Western blot with anti-His antibody, the protein is expressed in full-length both in an operon or individually, with a large portion remaining in the flow-through fraction. Degradation of the protein is also observed.

Molecular Weight: 6His-SUMO-SelenoN3, 55.8kDa. Abbreviation: T, total extract; S, soluble extract; FT, flow-through; W, wash fraction; E, eluted fraction.

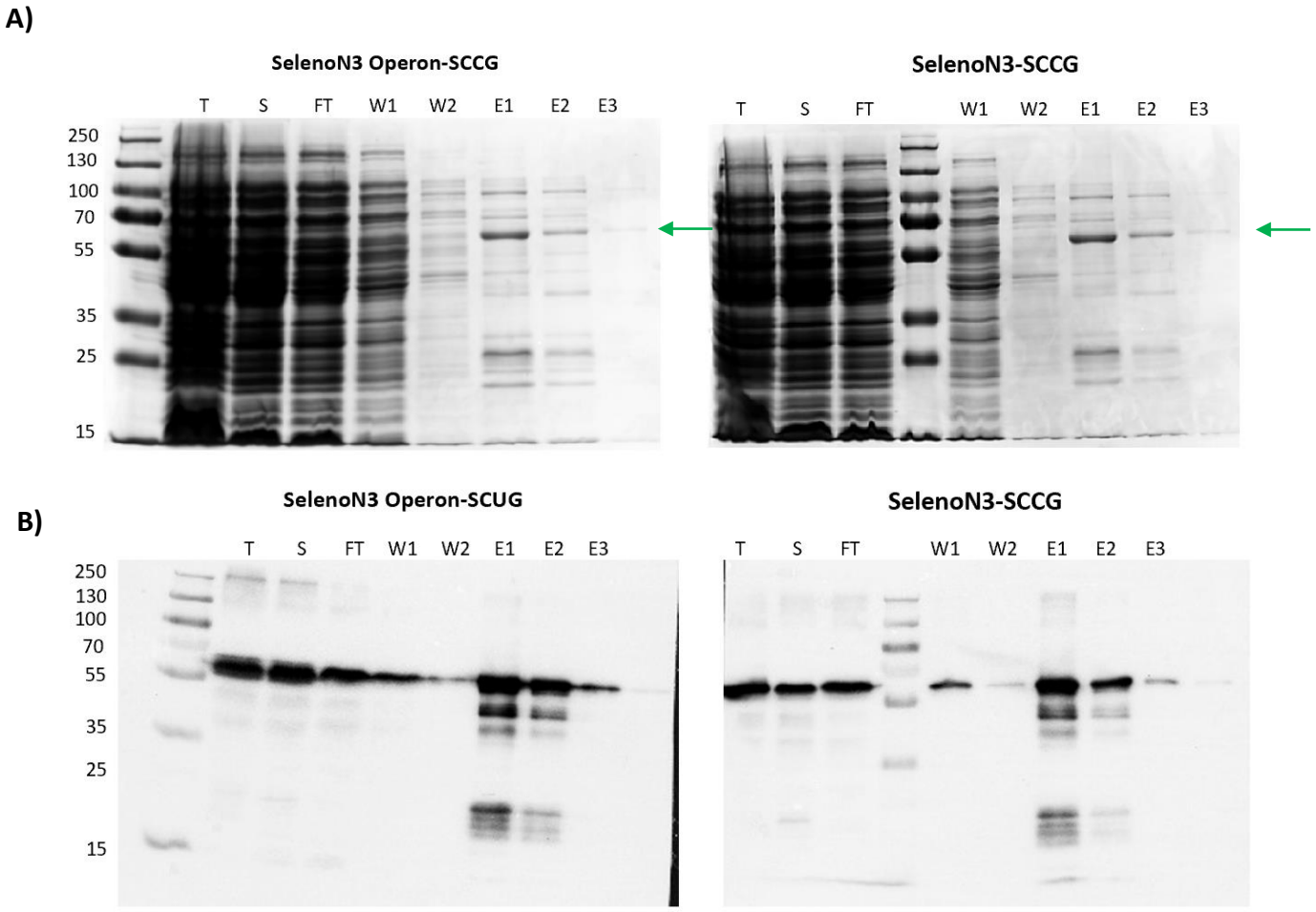


Figure R14. Purification of 6His-SUMO-SelenoN3-SCCG by Ni-IDA IMAC. (A) 10% SDS-PAGE stained with Coomassie blue, the protein band is detected and is better expressed than the Sec form. (B) 10% gel for Western blot with anti-His antibody, the protein is expressed in full-length either in an operon or individually with a large portion left in the flow-through fraction. Degradation of the protein is also observed.
Molecular Weight: 6His-SUMO-SelenoN3, 55.8kDa. Abbreviation: T, total extract; S, soluble extract; FT, flow-through; W, wash fraction; E, eluted fraction.

Preliminary experiments showed that SelenoN is co-purified with two chaperons, likely GroEL and DnaK, whether in the mutant or the wildtype form. Therefore, an extra chaperon removal wash step was implemented. In addition, since the size of the chaperon co-purified with SelenoN3 was very close to the size of the tagged recombinant protein, it was necessary to the cleavage of the 6His-SUMO tag using the Ulp1 protease. This step allowed us to compare the relative amount of the two proteins. After removal of the tag, the SDS-PAGE showed a strong residual band at 57.3 kDa corresponding to the chaperon, as SelenoN3 after tag cleavage migrating at 43.2 kDa only represented a faint band (Figure R15). Considering the impressive accumulation of the chaperone compared to the limited amount of SelenoN, it seemed that the chaperone is not only favoring the expression of SelenoN, but more likely, it was protecting the *E. coli* from the production of the exogenous protein. This toxicity could be related to the redox activity of SelenoN, therefore suggesting that the protein is active in the bacteria.

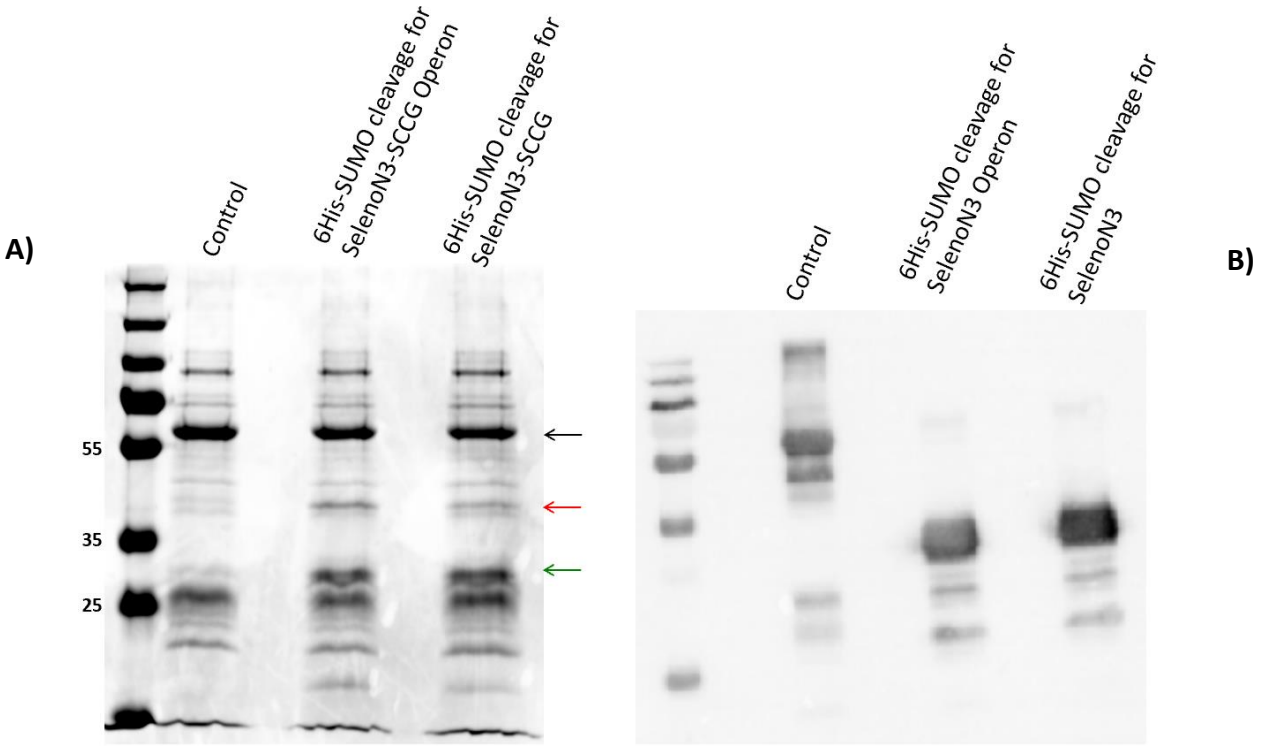


Figure R15. 6His-SUMO tag cleavage from SelenoN3. After tag cleavage by the specific protease Ulp1, the protein was loaded on the gel to detect the amount of protein in regard to the chaperone. **(A)** 10% SDS-PAGE stained with Coomassie blue. 1st lane is a non-cleaved control at 55.8kDa corresponding to the expected size for SelenoN3. The 2nd and the 3rd lanes correspond to the cleaved form of SelenoN3 expressed either in an operon with Rog3500 or individually respectively. The position of the chaperon is marked with a black arrow, the protein after tag cleavage at 43.2kDa is marked with a red arrow, and the Ulp1 is marked with a green arrow. Notice the huge amount of the expressed chaperone compared to SelenoN. **(B)** western blot analysis with anti-His antibody detecting only the Ulp1 protease in the cleaved fractions, which demonstrates the successful and complete removal of the tag. Calculated molecular weight: 6His-SUMO-SelenoN3, 55.8kDa; SelenoN3, 43.2kDa; 6His-SUMO, 12.6kDa; Ulp1, Ubiquitin-like-specific protease 1, 28.1 kDa

e) Mass Spectrometry Analysis

To validate the expression of SelenoN3 in its native form, mass spectrometry analysis for two samples of trypsin in-gel digestion and in-solution digestion of the purified SelenoN3 expressed from the operon construct was conducted. MASCOT algorithm for the in-gel digested sample detected SelenoN3 as the dominant protein with 161 spectra and 57 spectra for the *E. coli* chaperon GroEL with a molecular weight of 57.3 kDa. For the in-solution digestion 63 spectra for SelenoN3 and 122 spectra for another *E. coli* contaminant protein, FKBP-type peptidyl-prolyl cis-trans isomerase SlyD, which also has a chaperone-like activity. These two results confirmed the strong interaction of the purified SelenoN3 with different bacterial chaperons.

Moreover, 6 spectra for one peptide of Rog3500 were detected in the in-solution digestion, not enough to confirm the presence of Rog3500 co-eluted with SelenoN3 in the purified fraction. Therefore, we concluded that the co-expressed SelenoN3 and Rog3500 are unlikely to form a stable complex.

This mass spectrometry analysis was also used to assess Sec incorporation in the catalytic peptide using the MASCOT protein search engine by the proteomics platform. Of note, due to its ionization, the Sec residue is generally converted to dehydroalanine (DE) in mass spectrometry analysis, because of the instability of the Se atom (Ma et al., 2003). Indeed, the results from the in-gel sample digested with trypsin confirmed the insertion of DE/Sec into the catalytic peptide in the full-length protein. In addition, the preceding Cys residue was found carbamidomethylated (CA) by iodoacetamide treatment (Figure R16.a/b)**.**

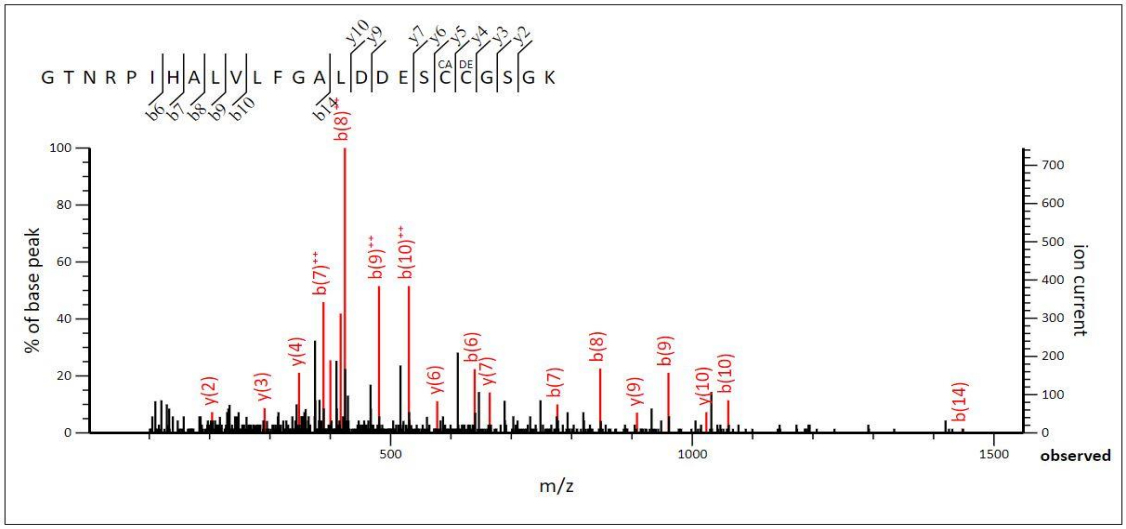


Figure R16.a. Sec incorporation assessment by mass spectrometry. Identification of the catalytic peptide in the InGel-sample confirmed the insertion of the Sec residue that is detected as a dehydroalanine (DE). In addition, the preceding Cys was carbamidomethylated (CA).

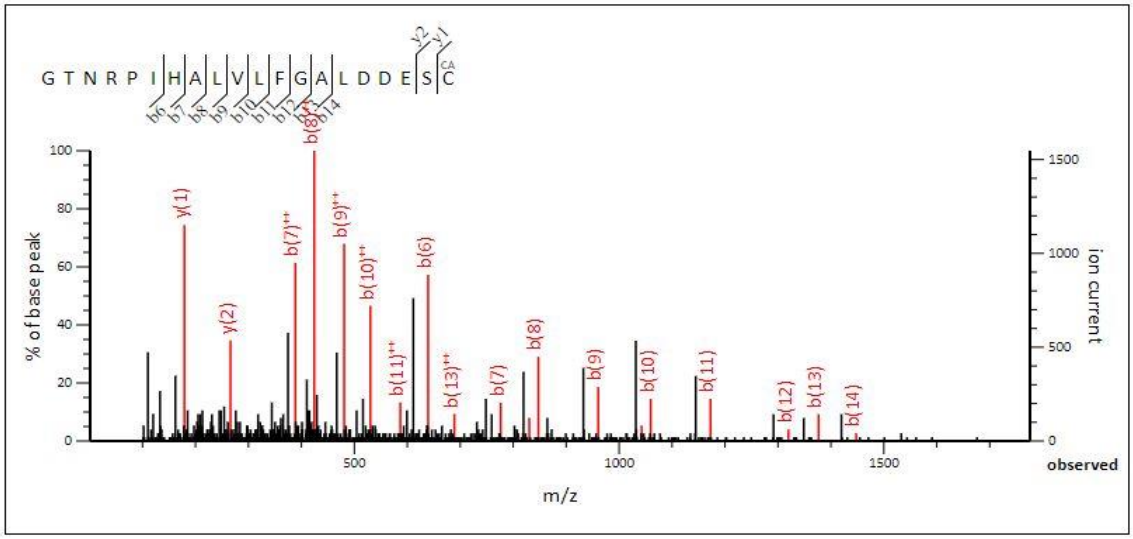


Figure R16.b. Sec incorporation assessment by mass spectrometry. Identification of the catalytic peptide in the InSolution sample showed that the peptide stopped at the Cys residue at the C-terminal that was carbamidomethylated (CA) by iodoacetamide. This result indicated a truncated form of the protein.

f) Optimization of SelenoN3 Expression

Since our main obstacle was the low expression level of the proteins, different conditions were tested to improve the expression for both SelenoN3 and Rog3500. To find the optimal incubation temperature and induction time, a series of different temperatures, and incubation periods, were tested. From this, it was concluded that lowering the expression temperature was facilitating the production of folded, soluble protein, but it also decreased the yield of the proteins even for a prolonged incubation period. Different concentrations of IPTG and induction at different OD^{600} were tried, but no difference in the expression level was obtained. In the case of SelenoN3-Sec form, expressions with different concentrations of selenite were assessed to increase Sec insertion into the polypeptide, but the expression remained the same between the minimum 5μM or higher concentrations.

Considering the strong association of the bacterial SelenoN with *E. coli* chaperons, Arctic express cells (Agilent Technologies) were tested to express the SelenoN3-Cys mutant. These cells are specially designed to allow the expression of heterologous proteins at low temperatures with high yield and correct folding by co-expression of the cold-adapted chaperones Cpn60 and Cpn10 from the psychrophilic bacterium, *Oleispira antarctica,* to improve the protein folding process. Those chaperones show high protein refolding activities at temperatures of 4–12°C and substitute the *E. coli* GroEL and GroES chaperones which lose their activity at these temperatures. Unfortunately, no improvement in the expression level was noticed.

Tag change was another thing to try, and new constructs of the SelenoN3 were created carrying a GST tag or a simple 6His tag without SUMO. However, it was no improvement in the SelenoN3 expression. Actually, the removal of the SUMO sequence didn't show any expression of the SelenoN3 in C321Δ for both wildtype and Cys mutant. The GST tag had the same expression level for the SUMO-tagged Cys mutant but not in the case of Sec wildtype.

Unfortunately, none of the above-listed attempts worked out for improving the expression level of a soluble protein. Hence then, both SelenoN3 and Rog3500 were expressed in small-scale culture in BL21De3 at 37°C for 6h in inclusion bodies, and different nonionic detergents were used to solubilize the proteins, but no solubility was obtained. Then, since the anionic detergent sodium dodecylsulfate (SDS) was found to solubilize overexpressed proteins successfully (Schlager et al., 2012), Sarkosyl (N-lauroylsarcosine) anionic detergent was added to the lysis buffer to solubilize

the proteins expressed in inclusion bodies. At 0.5% the solubilization of SelenoN3 was successful, but not in the case of Rog3500. Then purification was done to test the capacity of SelenoN3 binding to the Ni-column in the presence of Sarkosyl, only in the lysis buffer, which was found to be not affected by Sarkosyl. The washes and elution of SelenoN3 were Sarkosyl-free (Figures are not shown).

h) SelenoN oxidation test

Parallel experiments in the laboratory conducted by Ahmad Rida on the expression and purification of the Human SelenoN led to the observation that the purified recombinant protein was massively oxidized, with multiple amino acids scattered over the full sequence being modified. Oxidation of proteins modifies the side chains of many amino acids, including methionine, histidine, and tyrosine, and leads to the formation of disulfide bonds between cysteine residues. In addition, metal-catalyzed oxidation of proteins introduces carbonyl groups into protein side chains at lysine, arginine, proline, or threonine residues. Hence the oxidation possibility of the bacterial SelenoN3 was tested using an OxyBlot protein oxidation detection kit. The principle behind relies on the conversion of the carbonyl groups in the protein side chains to 2,4-dinitrophenylhydrazone (DNP-hydrazone) in a reaction catalyzed by 2,4-dinitrophenylhydrazine (DNPH). The DNP moiety attached to the protein is detected by antibodies specific to the DNP. The Human SelenoN was used as a positive control since it was previously tested positive with OxyBlot, but the results were negative for the bacterial SelenoN3 (Figure R17).

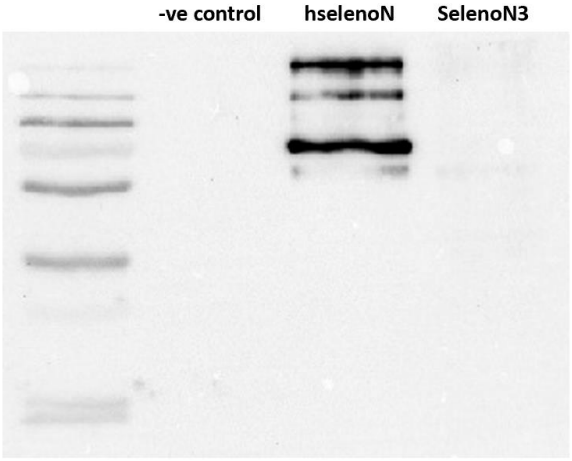


Figure R17. Evaluation of SelenoN3 oxidation by OxyBlot detection kit. Purified bacterial SelenoN3 oxidation level was compared to the one of the Human SelenoN (hSelenoN) and a negative control.

IV. Alphafold Structure Prediction.

Recently, in a collaboration between our laboratory and the team of Shuya Fukai (Kyoto university), the three-dimension structure of the Human SelenoN was solved by crystallographic X-ray diffraction. Importantly, the diffraction map was resolved thanks to a 3D structure prediction of the rat SelenoN provided by Alphafold (Jumper et al., 2021). Alphafold is a bioinformatics platform modeling protein structures from their primary amino acid sequence. Comparing the X-ray structure and the Alphafold model revealed that the prediction was very accurate, since the two structures overlapped by up to 98%. Therefore, we used the Alphafold program to predict the protein structure for both Rog3500 and SelenoN3 (Figure R18).

A comparison of the proposed models for SelenoN3 and the Human SelenoN showed that the two proteins are very similar (Figure R19). The conserved core is composed of three domains- a semi β-barrel structure (in blue); an intermediate domain composed of three antiparallel β-strands (in orange), and a classical thioredoxin-like domain composed of four β-strands surrounded by five alpha helices (in pink) and containing the two catalytic Cys and Sec residues almost entirely superimposed between the two proteins. The main difference resided in the more compact structure of the bacterial protein, the Human ortholog presenting several additional extensions protruding from the core domains.

Concerning Rog3500, it was proposed to adopt a unique β-propeller structure, which belongs to β protein structures. It is a highly symmetrical structure characterized by a supersecondary structure of multi-bladed β-sheets arranged sequentially through hydrophobic interactions between the sheets. Each β-sheet is composed of four anti-parallel β-strands. The number of β-sheets ranges from 4-12 repeats organized around a central core; The variations in the number of sheets determine the stability and the function of the β-propellers domain (Afanasieva et al., 2019). The Rog3500 model is composed of five β-sheets. Classically, β-propeller exists as a domain unit in larger proteins, even though Rog3500 structure prediction as an independent β-propeller protein is very unique. An example of β-propeller domain-containing proteins is the WD40 repeats which are implicated in a variety of functions including signal transduction, transcription regulation, control of cell cycle, and apoptosis, in which the β-propeller domain mediate protein-protein interaction and protein assembly into complexes (Afanasieva et al., 2019). Interestingly, the

genomic analysis of the SelenoN context in *Poribacteria* identified enrichment of WD40 and WD40-like genes in the vicinity of SelenoN (see previously).

Complete regular repeats of β-sheets proteins are aggregation-prone, due to their intrinsic tendency to form β-Hydrogen bonding with another β-strand they encounter. However, this aggregation is controlled in nature by an edge-protection mechanism of free edge strands (Richardson & Richardson, 2002). As we can see in the case of the Rog3500 expression we have a solubility issue most of the time, this could refer due to the repeated β-sheets in a closed circle.

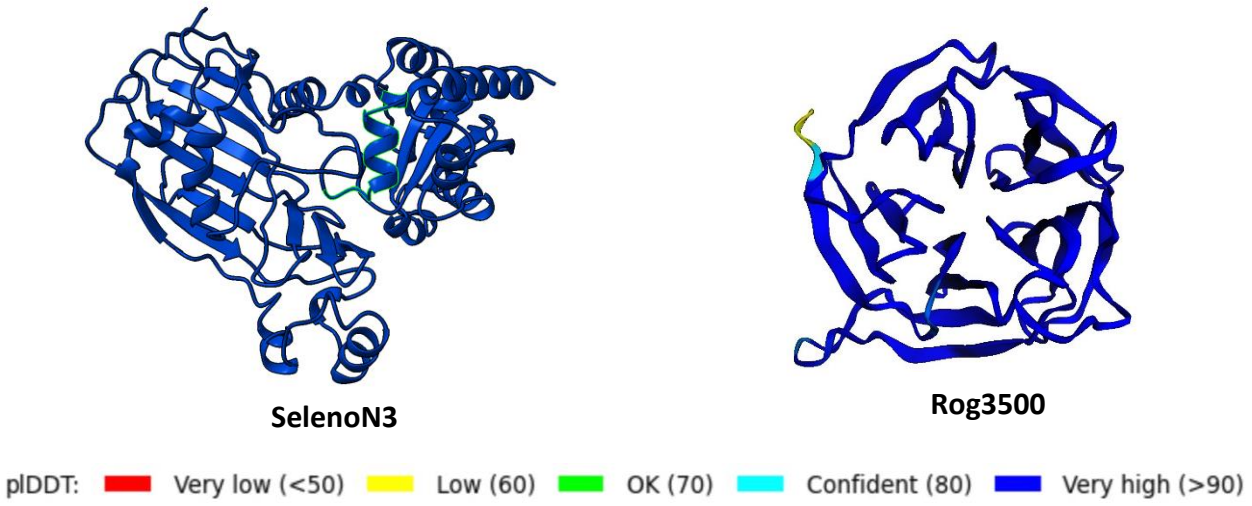


Figure R18. Alphafold simulation of SelenoN3 and Rog3500 structures. In SelenoN3 the SCUG catalytic domain is highlighted in green. Lower panel shows the prediction confidence (plDDT) scores according to which the model was coloured. Overall, the two structures were predicted with a high score.

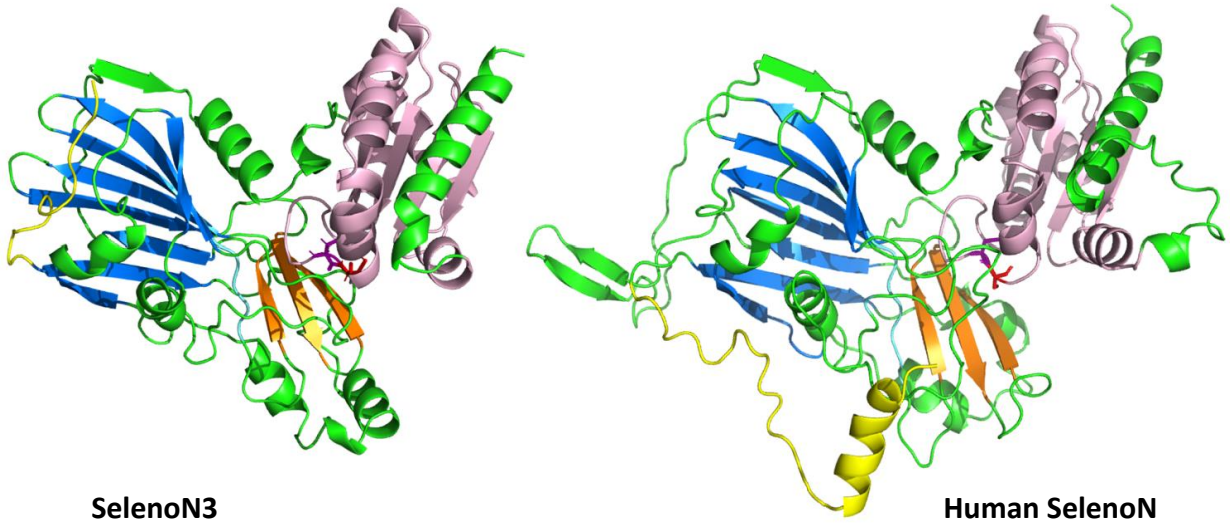


Figure R19. Alphafold modeling of bacterial and Human SelenoN. Protein 3D structures were predicted for the Poribacteria and Human SelenoNs using Alphafold. The common domains were highlighted as follows: N-terminal extension in yellow (truncated of the EF-hand and transmembrane domains for the Human protein); semi beta-barrel in blue, intermediate antiparallel beta sheets in orange; Thioredoxin-like domain in pink with the catalytic residues Sec in red and Cys in magenta. A conserved peptide linking the beta-barrel and the intermediate domain is depicted in light blue. The bacterial SelenoN is predicted to adopt a similar, also more compact, structure compared to its eukaryotic ortholog.

V. Distribution and colocalization of the Candidate phylum *Poribacteria* and Selenoprotein N in *Aplysina aerophoba*.

V. Distribution and colocalization of the Candidate phylum *Poribacteria* and Selenoprotein N in *Aplysina aerophoba* marine sponge.

The purpose of this study is to address whether SelenoN expression is associated with a specialized cell type of the sponge. Therefore, first we had to identify the distribution and localization of *Poribacteria* in the context of the sponge *A. aerophoba* holobiont by using fluorescence *In situ* hybridization techniques and electron microscopy imaging. Next, immune-gold staining was used to visualize if SelenoN expressing poribacteria are localized within defined spatially differentiated areas of the sponge This study is expected to provide hints about the sponge-bacteria interaction and hopefully can indicate a possible role for the *Poribacteria* in supporting SelenoN expression and activity. As matter of fact, and as presented in the introduction part, *Porifera* animals lack organized tissue structure, but still display an architecture that organizes specialized functions.

1. CARD-Fluorescence *in situ* hybridization (FISH)

CARD-FISH is a sensitive technique for the identification and visualization of microorganisms in their natural habitat using 16S rRNA nucleotide probes. CARD-FISH is more sensitive than conventional FISH because this technique relies on an enzymatic-mediated amplification of the fluorescent signal that makes it suitable to identify bacterial cells even with low ribosomal contents or low abundant microorganisms. These experiments were conducted as part of a collaboration with the laboratory of Marcelino Suzuki at the Observatoire Oceanologique de Banyuls-sur-Mer and under the expert supervision of Nyree West.

First, we analyzed *A.aerophoba* sections from animals that were maintained in seawater tanks at 20°C for several weeks, but similar results were obtained with freshly collected samples (data not shown). The first hybridization was conducted with a universal EUB338I probe that is complementary to a 16S rRNA sequence conserved in most bacterial species, and as a control for the specificity of the hybridization, we used the antisense probe, NON338. As expected, only background signals were obtained for the NON338 control probe, as specific signals with the bacterial universal probe EUB338I showed a remarkable distribution of different bacterial species in the mesohyl of the sponge (Figure R20). These signals were compared to the DAPI staining for all bacterial DNAs indicating that almost every bacteria was fluorescently labeled. Next, another

section of the same sample was hybridized with the *Poribacteria*-specific probe POR1130 (see Material and Methods), here again highlighting many bacteria, but less than with universal probe EUB338I. Interestingly, the *Poribacteria* labeled with the POR1130 probe appeared larger than the other bacteria identified with the EUB338I probe.

We noticed the following; in the control sample a remarkable distribution of different bacterial species using the universal EUB338I probe, and the enrichment of the sponge with *Poribacteria* using the *Poribacteria*-1130 specific probe (Figure R20).

a) Thermoneutral control sample kept at 20°C in seawater aquarium.

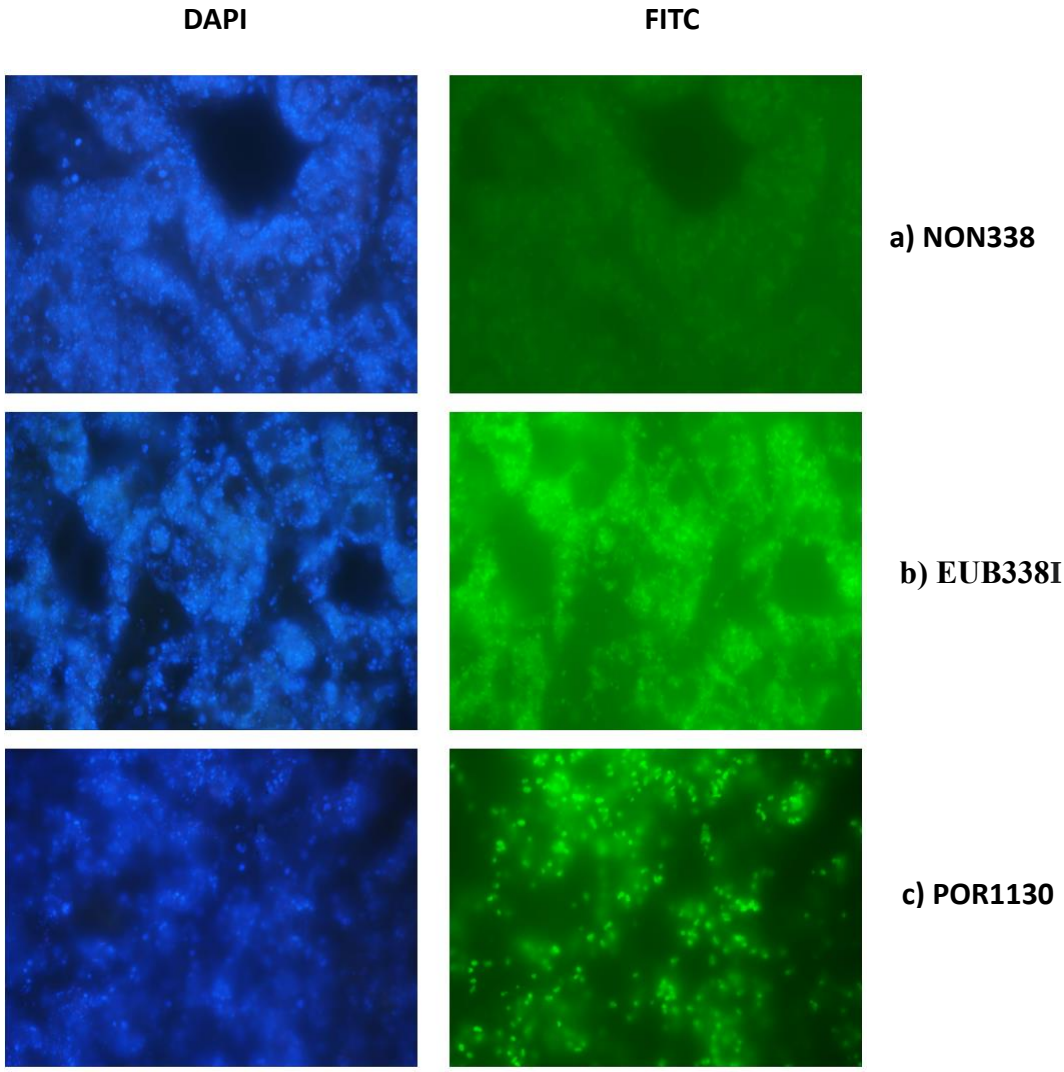


Figure R20. Distribution of *Aplysina aerophoba* associated-bacteria in a control sample. A) An autofluorescent green colour is shown with the negative control using NON338-HRP conjugated probe. B) a positive control for most bacterial species using the universal EUB338I-HRP conjugated probe. C) *Poribacteria* distribution using POR1130-HRP conjugated probes. DAPI nucleic acid staining is shown for each probe. Abbreviation; HRP, horse radish peroxidase; FITC, Fluorescein isothiocyanate filter.

Next, we got interested in the distribution of *Poribacteria* in the tissue of sponges exposed to stress. Indeed, the multicopy of the poribacterial *SELENON* genes and their organization in tandem repeats suggested an abundant expression at least under specific conditions. For example, the fusion of one *SELENON* ORF with a stress-related σ^{70} transcription factor pointed to a possible link with a stress response. Therefore, we conducted CARD-FISH experiments on sections of sponges that were submitted to heat stress (raising water temperature from 20°C to 27°C), either acute (24 hours) or chronic (two weeks), and compared to a control maintained at thermoneutrality (20°C) (see Methods for details). In the acute heat-stressed sample, we noticed a strong reduction in the different bacterial species, including *Poribacteria* (Figure R21). In the chronic heat stress sample, we noticed a partial recovery of the *Poribacteria*, which indicates the adaptation of the bacteria to the heat stress (Figure R22).

b) Acute heat stress sample kept at 27°C in seawater aquarium for 24h.

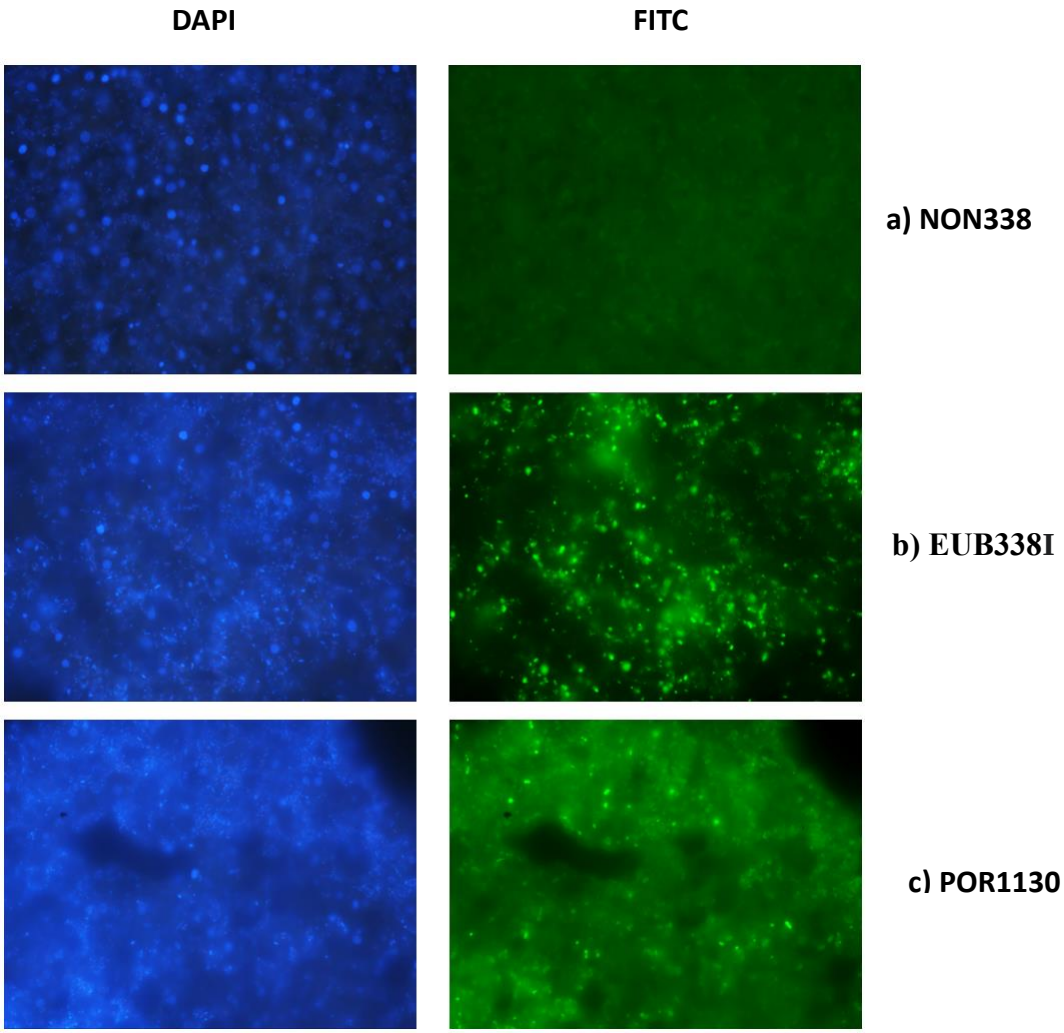


Figure R21. Distribution of *Aplysina aerophoba* associated-bacteria in acute heat stress sample. A) An autofluorescent green colour is shown with the negative control using NON338-HRP conjugated probe. B) a positive control for most bacterial species using the universal EUB338I-HRP conjugated probe. C) *Poribacteria* distribution (less than in the control sample) using POR1130-HRP conjugated probes. DAPI nucleic acid staining is shown for each probe. Abbreviation; HRP, horse radish peroxidase; FITC, Fluorescein isothiocyanate filter.

c) Chronic heat stress sample kept at 27°C in seawater aquarium for 2 weeks.

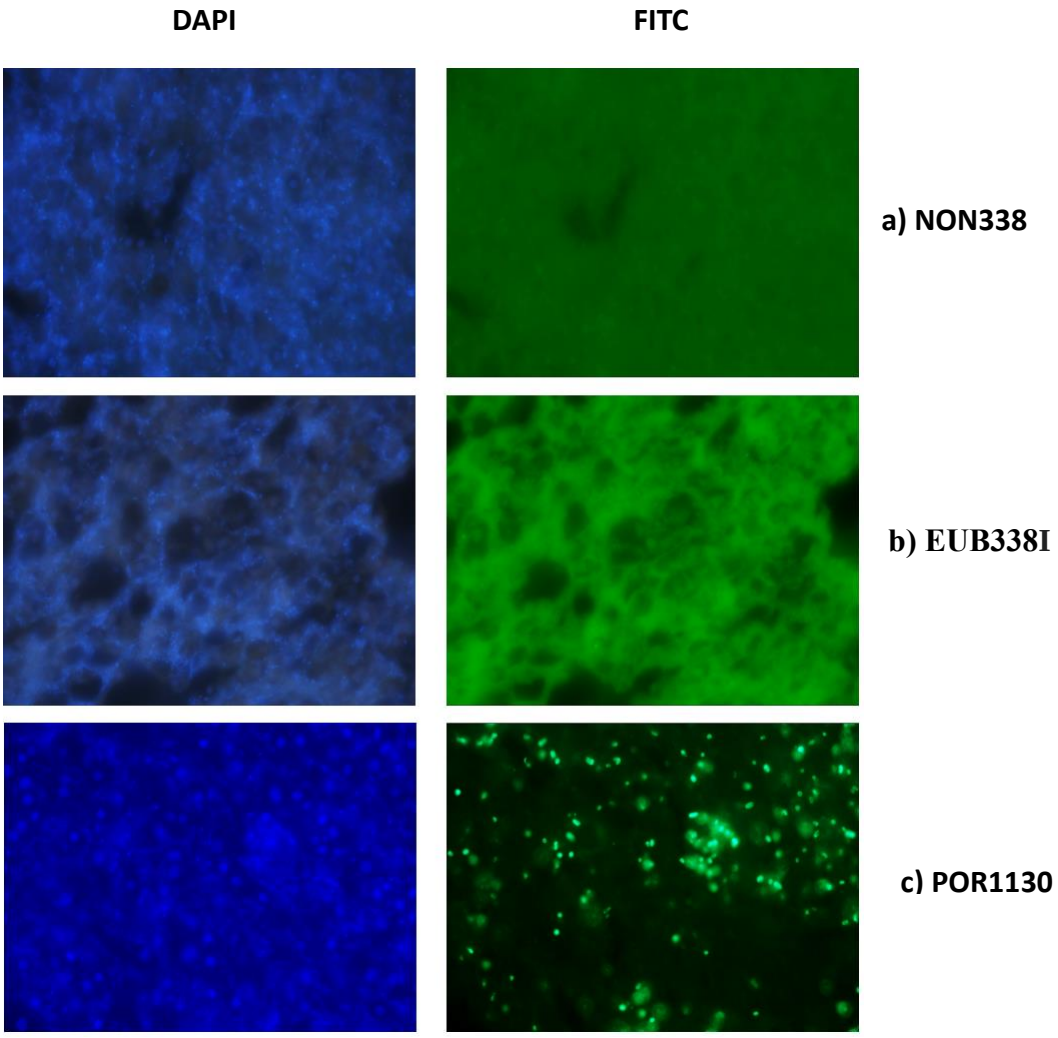


Figure R22. Distribution of *Aplysina aerophoba* associated-bacteria in chronic heat stress sample. A) An autofluorescent green colour is shown with the negative control using NON338-HRP conjugated probe. B) a positive control for most bacterial species using the universal EUB338I-HRP conjugated probe. C) *Poribacteria* distribution (bacteria numbers are recovered compared to heat stress) using POR1130-HRP conjugated probes. DAPI nucleic acid staining is shown for each probe. Abbreviation; HRP, horse radish peroxidase; FITC, Fluorescein isothiocyanate filter.

Confocal Microscopy images were taken for the heat stress samples hybridized with a *Poribacteria* 1130-specific probe to confirm the results of the heat stress effect on *Poribacteria* distribution (Figure R23).

In fact, we faced a problem with background fluorescence due to the autofluorescent from the sponge tissue, then this was explained due to the wide distribution of the photosynthetic *Cyanobacteria* in *A. aerophoba* sponge. C*yanobacteria* are a rich source of biopigments such as chlorophylls, carotenoids, phycobiliproteins (fluorescent proteins of various colours), phycoerythrin, phycobilin, and more. The emitted colours depend on the dominant photosynthetic pigment in the cells of particular species (Mandal et al., 2020).

d) Confocal Microscopy photos for the distribution of *Poribacteria* in the control, acute, and chronic heat stress samples.

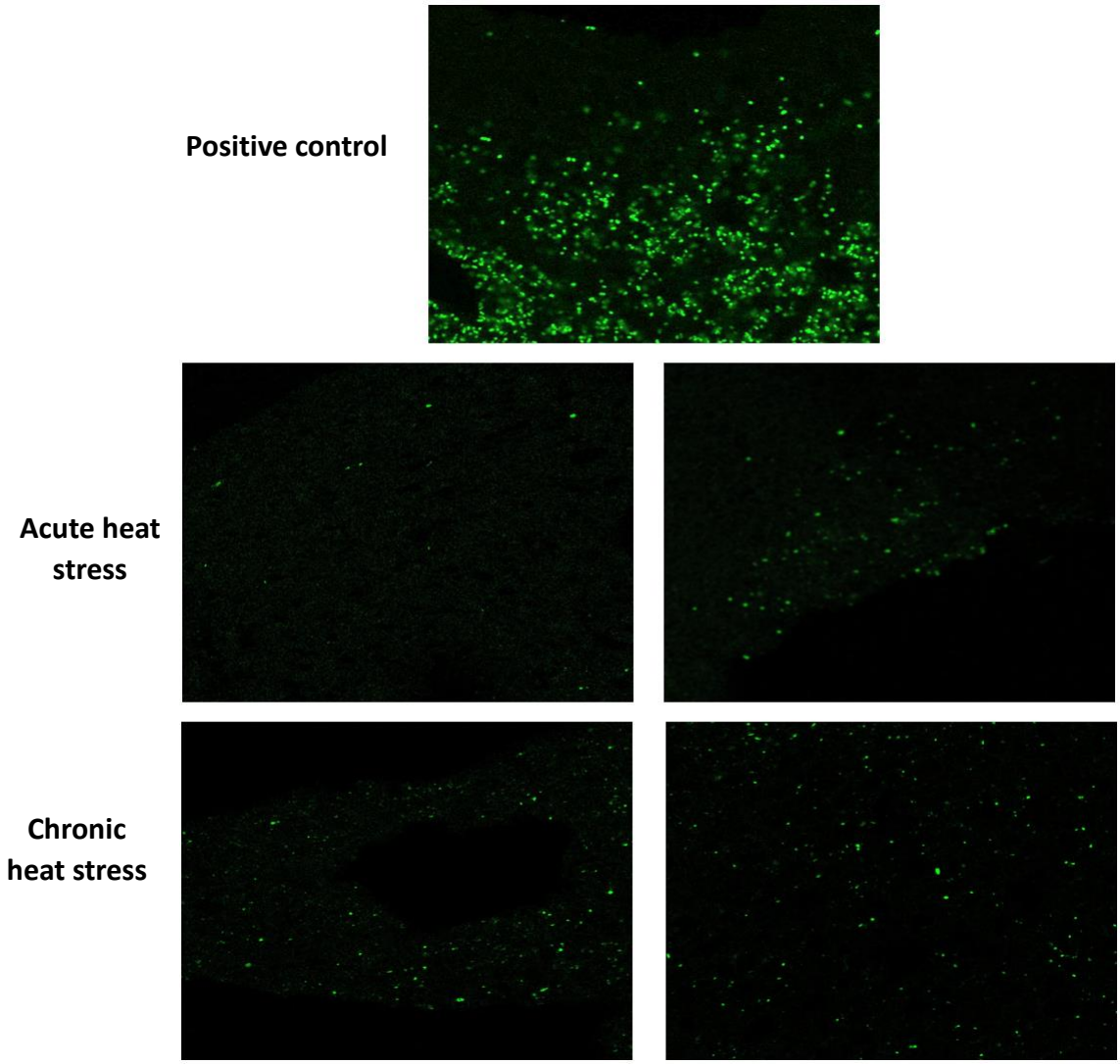


Figure R23. Confocal microscopy for *Poribacteria* distribution in *Aplysina aerophoba* tissue in stress conditions. The positive control at 20°C sample shows high numbers of *Poribacteria* distributed in the sponge tissue. In the acute heat stress sample two images were taken from two different spots and both show a reduction in *Poribacteria*. In the chronic heat stress, two images were taken from two different spots and both show a recovery in *Poribacteria* numbers compared to the acute stress but still less than from the control sample. However, in both acute and chronic stress, we observed the effect of heat stress on *Poribacteria* proliferation compared to the control.

2. Electron microscopy

a) Transmission Electron Microscopy (TEM).

To gain more detailed information about the *Poribacteria* and *A. aerophoba* interaction, we performed an electron microscopy imaging of the sponge tissue. A comparison of different sections showed clear differences in sponge organization and the composition of the symbiotic bacterial community. Refer to (Figure M1.b) in the methods section for sponge collection. The first picture (Figure R24, top) shows a section from the top of one sponge digitation, close to the apex. In this section near the ectoderm, we observed an accumulation of sponge cells (spherolous cells) surrounded by a rich environment of monotypic bacteria. The sponge spherolous cells contain large intracellular compartments, called spherules, in which brominylated alkaloid compounds, that act as a defense mechanism of the sponge, are stored (Turon et al., 2000). In the pictures we obtained, the spherolous cells show a large number of spherules with different densities, likely to correspond to different concentrations of the alkaloids. On the other hand, in the middle side section of the sponge which corresponds to the mesohyl, we observed only a few spherolous cells and a distribution of many different bacterial species, according to the large variety of morphology, with few examples of the monotypic bacteria previously observed in the apical part (Figure R24, bottom).

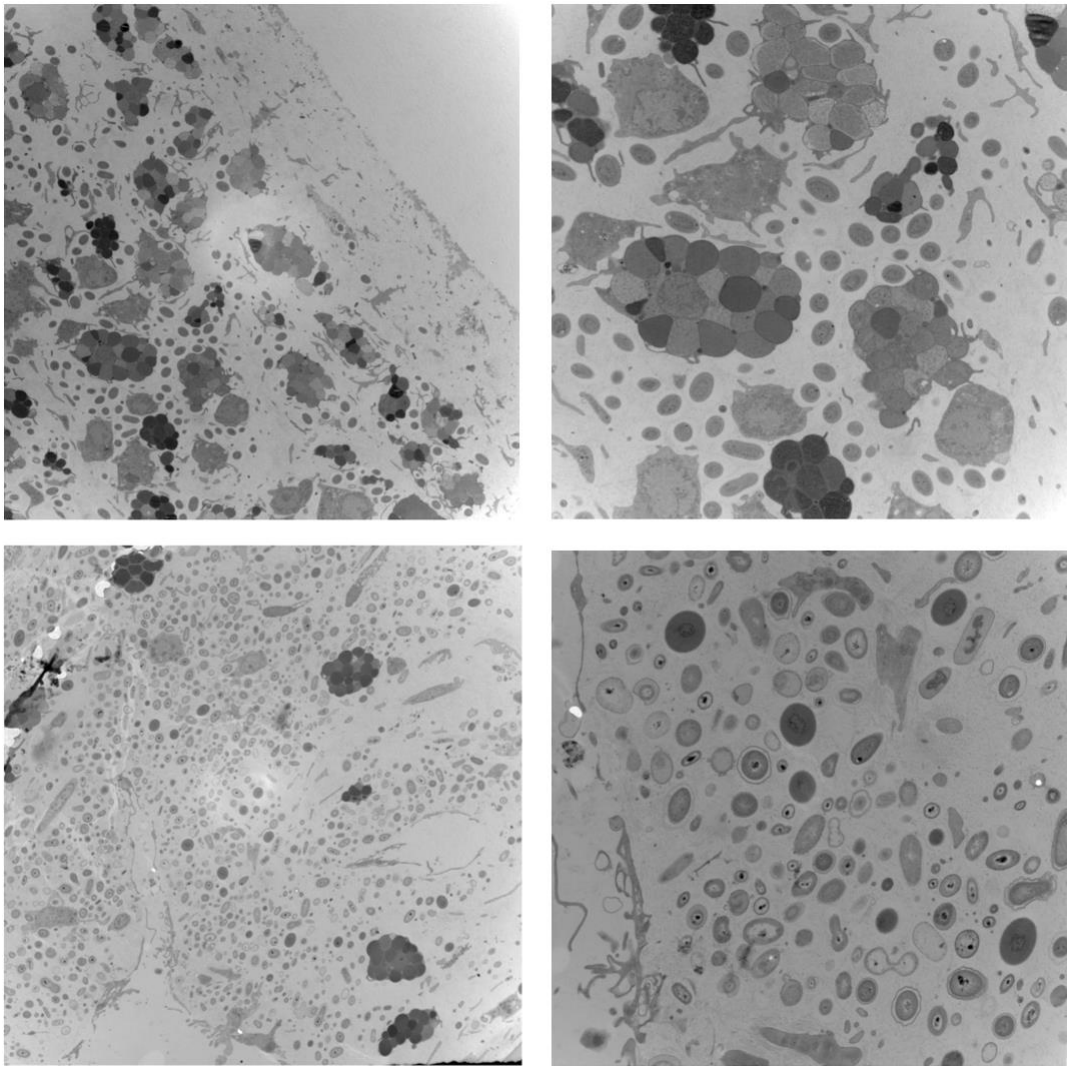


Figure R24. Transmission Electron Microscopy (TEM) of *Aplysina aerophoba* tissue sections. The top pictures display the apex part from the top of the sponge showing a remarkable distribution of sponge spherolous cells and monotypic bacteria. The bottom pictures display the mesohyl from the middle of the sponge displaying a distribution of different bacterial species and a reduction in the number of sponge cells.

A closer view of the sponge spherolous cells and the surrounding bacteria is displayed in (Figure R25). It showed that the monotypic bacteria display internal membranes and a size that ranges from 0.8-2.5 microns, which are characteristics of *Poribacteria*. Comparing these morphological features with the ones described in previous studies on *Poribacteria* using TEM or CELM-FISH (Fieseler et al., 2004; Jahn et al., 2016), we hypothesized that these bacteria correspond to *Poribacteria.*

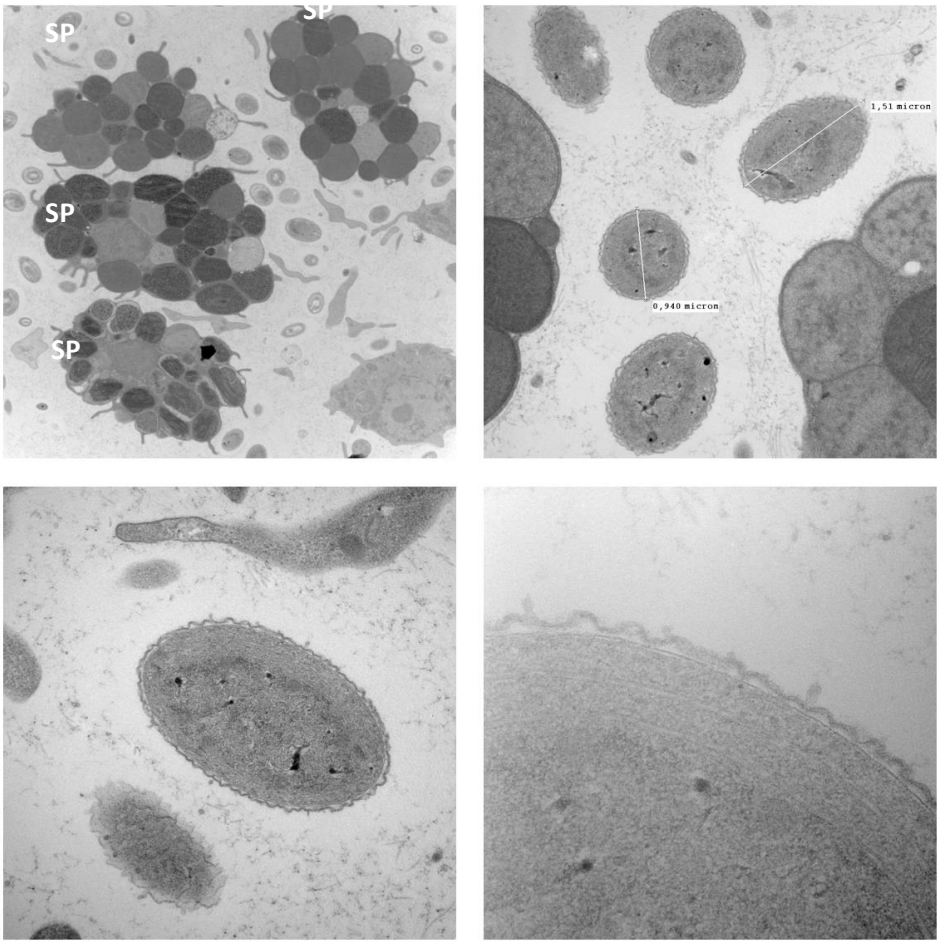


Figure R25. Transmission Electron Microscopy (TEM) of *Aplysina aerophoba* tissue sections of the bacteria nearby the spherolous cells (SP). Different magnifications show the ultrastructure, especially the presence of internal membranes in the bacteria. The size of the selected bacteria is indicated in the last panel.

In addition to the remarkable accumulation of the predicted *Poribacteria* around the spherolous cells, close contact between the prokaryotic and eukaryotic cells was observed, with bacteria being engulfed by protrusions of the spherolous cells in some instances (Figure R26). This close contact suggested a host-bacteria interaction in which both the host and the symbiont share mutual functions. The key point here is how this could relate to SelenoN activity. The accumulation of *Poribacteria* nearby the sponge spherolous cells might have some indications of the SelenoN implicated in a reaction involved in the defense of the sponge.

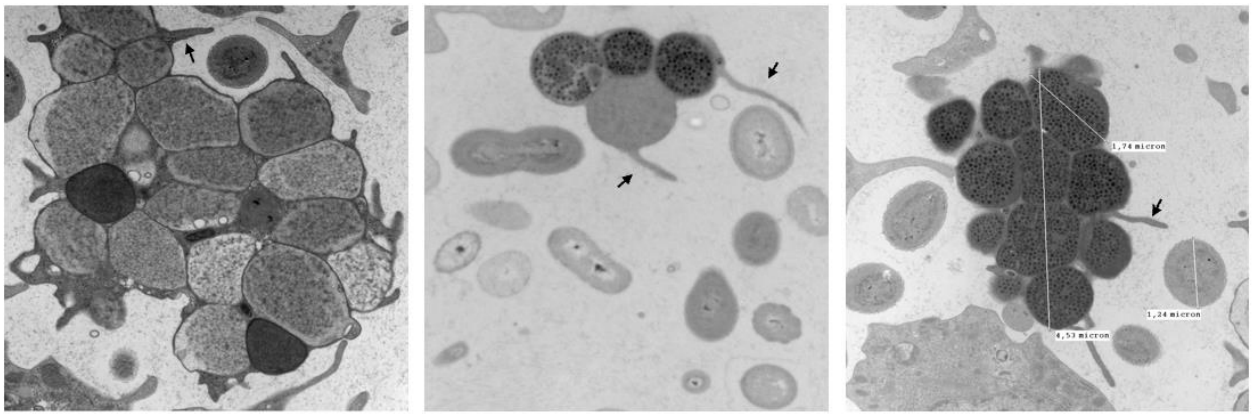


Figure R26. Transmission electron microscopy images showing the close contact between animal spherolous cells and the predicted *Poribacteria*. Notice the protrusions of the sponge cells engulfing the predicted *Poribacteria* (black arrow).

b) Immunogold for detection of SelenoN expression in *Poribacteria*.

To confirm the identity of the monotypic bacteria that were found in close proximity to *A. aerophoba* spherolous cells, we pursued our investigations on the localization of *Poribacteria* by performing immunogold-labeling using an antibody developed in the lab directed against one copy of SelenoN fused to the thioredoxin-like domain. A negative control using only the secondary-gold conjugated antibody is shown in (Figure R27). The gold particles were found concentrated in some bacteria, which revealed the high expression level of SelenoN (Figure R28). Of note, the gold-labeled antibodies cannot penetrate the sections and only proteins present at the surface of the section can be detected. Therefore, the intensity of the signal detected in these experiments is quite high and equivalent to what is observed in a glandular secretion system. We also noticed in some cases the distribution of the gold particles on the outer membrane of the *Poribacteria* (Figure R29). We also observed differences in the morphology of the SelenoN-expressing *Poribacteria*, oval, round, and elongated shapes distributed around the spherolous cells, in addition to either smooth or wavy outer membranes. Moreover, Immuno-gold staining showed that the bacteria that make close contact with the spherolous cells also express SelenoN (Figure R30).

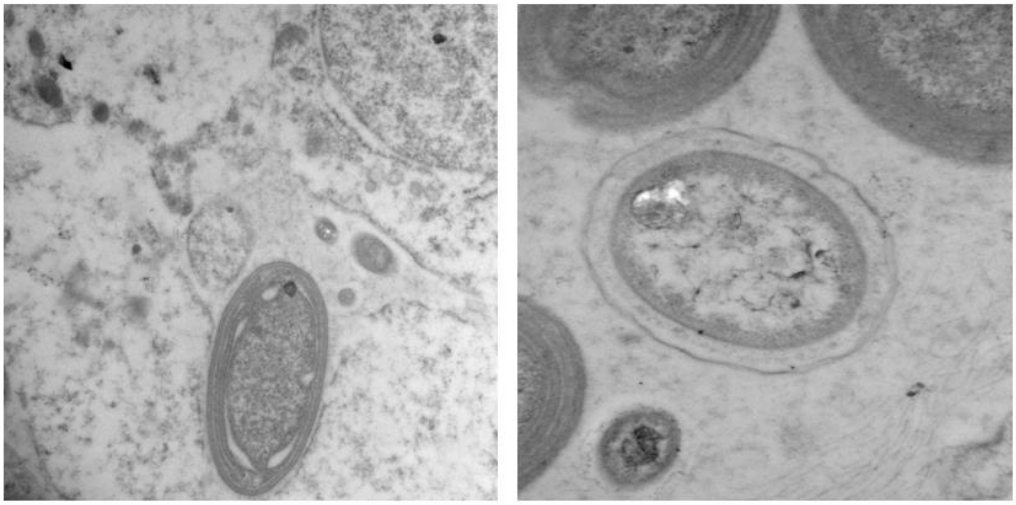


Figure R27. Immunogold-labeling Electron Microscopy negative control for Poribacteria in Aplysina aerophoba tissue. No primary antibody, only the secondary-gold conjugated antibody is used.

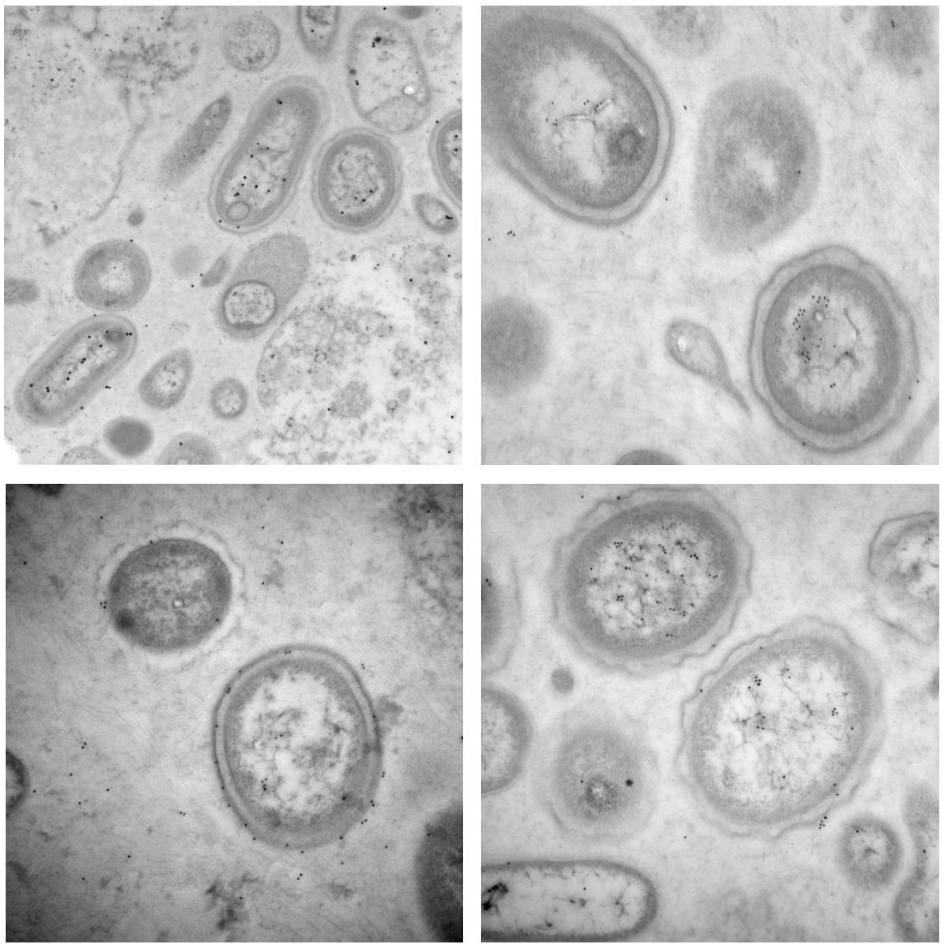


Figure R28. Immunogold-labeling Electron Microscopy for *Poribacteria* in *Aplysina aerophoba* tissue using anti-SelenoN antibody result 1. The four images display the detection of SelenoN expression by gold particles in different *Poribacteria* with different morphologies. 10nm gold particles were used.

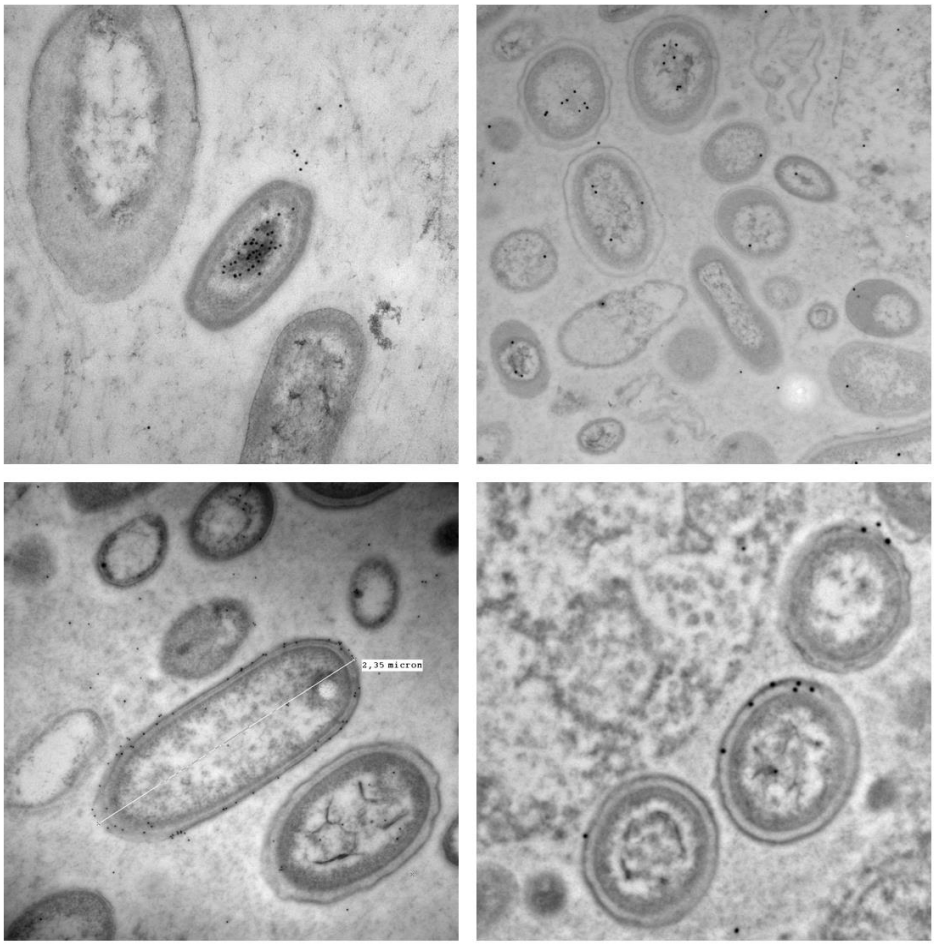


Figure R29. Immunogold-labeling Electron Microscopy for *Poribacteria* in *Aplysina aerophoba* tissue using an anti-SelenoN antibody result 2. The images display the concentration of the gold particles either in the center of the bacteria or on the outer membrane. Different morphologies are observed here too. 10 or 25 nm particles were used.

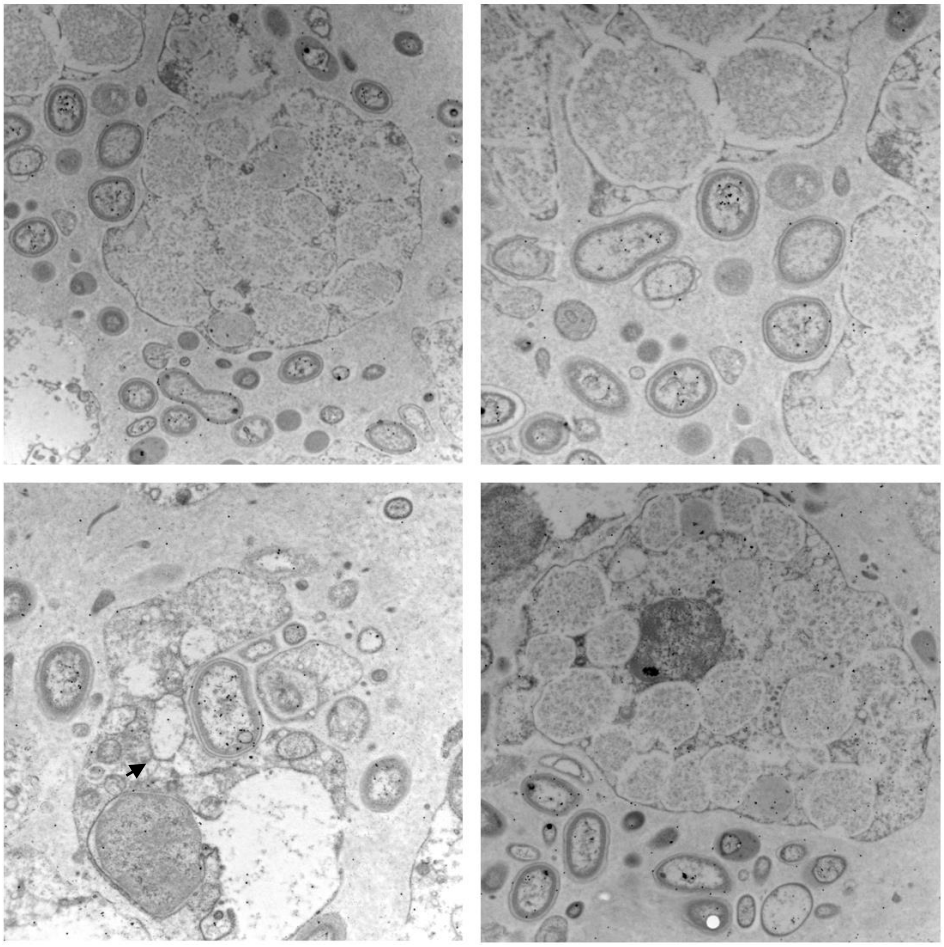


Figure R30. Immunogold-labeling Electron Microscopy for *Poribacteria* in *Aplysina aerophoba* tissue using an anti-SelenoN antibody result 3. The images display SelenoN-expressing *Poribacteria* stained with gold particles in close proximity to spherolous cells. In the lower left picture we can observe an engulfing event of *Poribacteria* (black arrow).

DISCUSSION

I . Bioinformatics Studies on the Candidate Phylum *Poribacteria* Genome.

Since the function of eukaryotic SelenoN could not be predicted based on the sequence and there is no homology to any known protein, it was important to broaden the screening for orthologous genes in other species to provide information about the evolution of this gene. Computational means have led to the identification of *SELENON* genes in one single yet-to-be-cultured group of bacteria, the candidate phylum *Poribacteria*. To our knowledge, we are the only lab that is conducting this kind of bioinformatics screening on the *SELENON* gene in bacteria.

To gain knowledge into *SELENON* molecular function we aimed to map the *Poribacteria* genomic organization of different strains and analyze *SELENON* genomic context based on prokaryotic properties to identify specific features such as genes in the vicinity or spatially linked to *SELENON*. Our identification of numerous *SELENON* copy numbers in many strains and the occurrence of *SELENON* genes in the same contig organized in tandem repeats reveals a duplication events. So why bacteria tend to duplicate a gene of interest is actually due to many reasons. First, it could be due to the abundant expression required, the emergence of paralogous genes associated with related or different diversification of functions in *Poribacteria,* or it could be for specialization with different substrates. Second, a strategy of the bacteria against loss of function of the gene. In addition *SELENON* in tandem repeats suggests high expression of the protein at least under certain conditions.

Furthermore, inspection of SelenoN amino acid sequence in the different *Poribacteria* strains reveals that SelenoN is in fusion to additional N- or C-terminal domains which suggests that specific genes have strong genetics link to *SELENON*. According to the “Guilt by association” theory (GBA) genes co-expressed and co-transcribed within one operon generally code for proteins contribute to related functions or metabolic pathways, and genes found in synteny are important in genome comparison to reveal genomic evolution of related strains or species. So the fact that SelenoN is in operon or in a synteny with other genes suggests that the two proteins contribute to the same metabolic or signaling pathway, and here it is more likely to be a metabolic pathway because SelenoN is a redox enzyme involved in metabolism. Therefore, operon amplification/recovery is the first step toward testing the functional relation between *SELENON* and linked genes.

II . *Poribacteria SELENON* Gene Amplification.

We got very interested in the bacterial version of SelenoN because of the sequence similarity between the bacterial and the eukaryotes SelenoN which suggests a horizontal gene transfer (HGT) event of *SELENON* gene from the sponge to the bacteria supported by the degree of identity and similarity. This requires an evolutionary adaptation of the translation mechanism in the bacteria**.** Another asset of the extensive analysis of the *Poribacteria* genomes is to reach the point of amplifying the correct and natural bacterial *SELENON* from the *Poribacteria.*

In fact, encountered some difficulties in *SELENON* amplification by inverse PCR technique from a metagenomic extract. The holobiont structure and its associated metagenome implicate a large complexity of different genomes with different sizes. Inverse PCR requires steps of digestion and re-ligation of the template genomes. Thus, a complex mixture of genomes from different species could limit the accessibility of the restriction enzymes to their recognition sites or be saturated in digesting DNA fragments from untargeted genomes. Moreover, endonucleases recognition sites are unevenly distributed across genomes, and the inverse PCR applied to a genome that is digested by restriction endonucleases is not always productive. This limitation became obvious in our *SELENON* amplification from the metagenomic extraction. As mentioned previously, ten restriction endonucleases were used separately to digest the metagenomic extract. However, only inverted PCR performed on the EcoR1 digested template provided us with the complete *SELENON* sequence, while the complete *ROG3500* was obtained from the Bfm1 digestion. Surprisingly, there is an EcoR1 recognition site in the first third of the *SELENON* ORF, and a Bfm1 site at the beginning of the *ROG3500* sequence, for which, if the digestion was successful, amplification wouldn't be achieved. Of note, none of the endonucleases without any recognition sites on either *SELENON* or *ROG3500* yielded a complete sequence amplification for either of the two genes in the inverse PCR. Therefore, upon the complexity of our sample, digestion by endonucleases was ineffective due to the abovementioned reasons. Our conclusion was that shearing of the genomic material caused by the tissue homogenizer generated small fragments (a few kilobases, kb), which were circularized independently of the digestion. On the other hand, the limited size of the genomic DNA fragments hampered the amplification of larger genomic fragments containing multiple or longer genes. This was deduced from the fact that we never had an amplification over 3kb at all, and some *SELENON* operons are more than 4kb in size. Hence,

the homogenizer and endonucleases digestion steps could be replaced with a gentler fragmentation method.

In a report from (Alquezar-Planas et al., 2021), the Fast-Prep homogenizer step was avoided not to cause damage to the DNA, and the enzymatic-based fragmentation was replaced by a sonication-based shearing and blunt-end repair for the DNA fragments, that was next used in an inverse PCR method. They concluded that this method effectively identified unknown sequences flanking a known sequence, avoiding the limitation of the use of restriction endonucleases digestion.

III. Recombinant Selenoprotein N Expression.

The bacterial version of SelenoN displays several features that constitute crucial advantages for its expression in a bacterial system. Contrary to eukaryotic protein, the bacterial version lacks a transmembrane domain and might be expressed as a soluble protein, the EF-hand calcium domain, and it is not glycosylated. Therefore the bacterial SelenoN was a good candidate for expression in a heterologous bacterial system such as *E. coli*. Numerous heterologous proteins fail to fold into their native conformation when expressed in the heterologous *E. Coli* expression system; instead, they are either degraded by the cellular proteolytic machinery or they form insoluble aggregates as inclusion bodies (Georgiou & Valax, 1996). The bacterial SelenoN seems to possess intrinsic properties that affect its biosynthesis and folding capacities in conditions other than its natural environment.

In this work, our main objective was to produce the bacterial SelenoN in fair quantities to proceed with the downstream application for functional assays, such as *in vitro* enzymatic assays. Comparing the SelenoN expression between the synthetic and amplified *SELENON* genes, we can notice the improvement in the expression of the full-length wildtype protein. Then we needed to develop a protocol that allows the production of highly pure, soluble, full-length. Different optimal expression and purification conditions were undertaken for the recombinant SelenoN protein in the *E. coli* expression system, including induction at different ODs, induction time, different host cell lines in the case of the Cys mutant, tag change, eventually no appreciable change in protein expression level was obtained though. Noticeably, inspecting codon usage in ORFs at the metagenomic level, we have checked that the codon usage in *Poribacteria* and *E.*

coli are similar. The only effective option was to lower the temperature to compromise between solubility and quantity.

At the cellular level, to increase the solubility, the protein is normally co-expressed with chaperones which assist the protein to adopt its correct folding. In the case of bSelenoN, expression in the *E. coli* strain C321ΔA, the chaperone GroEL was over expressed, largely exceeding the amount of the bSelenoN, and co-purified with the bSelenoN, as it was confirmed by mass spectrometry analysis. We hypothesized that the chaperonin was not expressed to help in protein folding, but is a result of a cellular stress-response of *E. coli* to the production of insoluble heterologous protein, which could imply toxicity of the bSelenoN as the chaperone would protect the bacteria from the protein itself during stationary phase. Or due to improper folding of the bSelenoN.

IV. Electron Microscope and Fluorescence *in situ* hybridization.

The distribution of the candidate phylum *poribacteria* and the SelenoN in *Aplysina aerophoba* marine sponge was examined using CARD-FISH, and transmission electron microscopy.

1. CARD-Fluorescence *in situ* hybridization (FISH)

FISH is a useful cytogenetic technique for the visualization and localization of microorganisms in their natural habitat using specific DNA sequence fluorescently labeled probes that bind to their complementary sequence of RNA with a high degree of sequence complementarity, like the 16S rRNA nucleotide sequence of the target cells. Even though FISH sensitivity for detection can be minimized due to some issues, like the low number of target molecules in cells like in mRNA transcripts, non-specific binding, and weak hybridization to the target molecule. Tyramide Signal Amplification (TSA) or also called catalyzed reporter deposition (CARD) was developed to increase signal intensities. CARD is an immunohistochemistry technique that is used in FISH to detect and quantify nucleic acids even in the case of low abundant mRNAs or short oligonucleotide probes. CARD-FISH was designed for signal amplification and was used over 20 years in immunoassays and later in FISH techniques for signal amplification purposes. The principle of CARD mainly relies on the ability/activation of tyramide in the presence of H_2O_2 produced by horseradish peroxidase (HRP). The active tyramide nonspecifically reacts with aromatic compounds, such as tyrosine and tryptophan residues near the HRP molecules,

resulting in a deposition of the active or radical tyramide around the HRP molecules. CARD reaction can occur by different methods depending on the target, whether it is a nucleic acid or a protein. For nucleic acids, the method relies on direct detection where tyramide conjugated-fluorophores deposit around HRP-labeled probes; for proteins indirect detection is used with biotinylated tyramide and HRP-labeled streptavidin. Both approaches showed significant signal amplification for the CARD reaction, with more than 10-fold stronger than classical FISH (Kubota, 2013; Shakoori, 2017).

In addition, the permeability of the cell, and the three-dimensional structure of the ribosomal RNA can block the access of the HRP-conjugated probes due to their large size. To address this issue, a systematic study on the accessibility of 16S rRNA target sites had been conducted (Fuchs et al., 1998, 2000) using several *E. coli* species. An *in situ* accessibility map of *E. coli* was obtained that can be used as a blueprint to design 16S rRNA oligonucleotide probes for other species; therefore, helper DNA based oligonucleotides were used in this study as proposed in the publication to aid in the accessibility of the oligonucleotide probes in FISH, and strength the hybridization signals. The probes were designed from the adjacent nucleotidic residues of the target sites of the HRP-labeled probes.

In addition, the decision on which fluorophore to use in the CARD-FISH is very critical, as previously pointed out to the autofluorescent problem from the *Cyanobacteria*, this could be avoided by changing the fluorophore to another one that will not interfere with the autofluorescence signal of *Cyanobacteria.* Moreover, to be able to observe the accumulation of the *Poribacteria* around the spherolous cells on FISH slides, we can use Nomarski microscopy (differential interference microscopy) in a brightfield filter to observe the structures of the spherolous cells and then locate the surrounding *Poribacteria* and their proximity to the sponge cells.

2. Electron microscopy (EM)

Concerning the immunogold labeling, we tested a combination of different fixatives in tissue preparation to balance the efficiency of gold labeling with maximum preservation of the subcellular structure. Glutaraldehyde fixation is indicated to preserve the ultrastructure of tissues in EM, that is why we used 2.5% glutaraldehyde followed by 1% OsO4 for Epon embedding for the transmission electron microscopy (TEM). For immunogold labeling, this fixative is “too

strong": The free aldehyde groups introduced by the two aldehyde groups (–CHO) of the glutaraldehyde make links between proteins and form networks of proteins that prevent antigens accessibility to the antibodies. However, fixation with formaldehyde (single –CHO) is better to preserve antigenicity, but poorly preserve the ultrastructure. Therefore, a mixture of the two aldehydes was used; 3% formaldehyde was added with 0.5% of glutaraldehyde. This makes a compromise allowing a minimum level of conservation of ultrastructure and the preservation of antigenicity. Moreover, fixation by glutaraldehyde introduces free aldehyde groups in the sample which are responsible for the non-specific binding of the antibodies. Because of that, the free aldehyde groups were deactivated by a glycine solution, but many residual aldehyde unbound groups that could not be washed out of the tissue can still cause non-specific binding of the antibodies.

V. Conclusion

The Ph.D. project aimed to work on the expression of the bacterial SelenoN protein for further downstream enzymatic and structural analysis. However, since the structure of the Human SelenoN was recently solved by the group of Shuya FUKAI (Kyoto University) in collaboration with our team, we oriented the research project toward a different area. In the beginning, we were working on bSelenoN amino acid sequences that were derived from synthetic genes based on the bioinformatics analysis of the genomes of the candidate phylum *Poribacteria* from the WGA database. Unfortunately, the amino acid sequences were unreliable even though frameshift mutations were manually corrected. Hence, after mapping the *Poribacteria* genomes and the identification of the organization of genes occurring in the vicinity of *SELENON*, we analyzed the cluster of genes. As a result, we took advantage of the genes present in operon with *SELENON* as a marker to amplify the *SELENON* gene. Two genes were consistently present with *SELENON* in one operon; *ROG3500*, a gene of unknown function, presents in ten copies of *SELENON*, with no additional domain (unfused), in different poribacterial strains, and a Redoxin-coding gene presents in eight copies of *SELENON* fused to a thioredoxin-like domain at the C-terminal. The *SELENON-ROG3500* operon was successfully amplified using the inverse PCR technique. The Redoxin-coding gene was amplified as well, but not its associated *SELENON* copy due to the large size of the operon. The Redoxin-coding gene was amplified by Leo Tekotte. After that, the unfused-SelenoN was expressed and purified. However, we hoped that Rog3500 might have an impact on SelenoN and can trigger the expression to higher levels in

a soluble form, but we did not observe any difference between SelenoN expression either individually or in operon with Rog3500.

So, what we conclude from the preliminary results is as followed: there is a high conservation between bacterial and eukaryotic SelenoN structure in which a conserved motif was identified in both proteins and was detected with pathogenic miss-sense mutations. In addition, the bacterial SelenoN is predicted to adopt a similar but more compact, structure compared to its eukaryotic ortholog

Even though, SelenoN and Rog3500 are co-expressed in an operon, but no co-purification of the two proteins were shown in the results since operon expression reduced Rog3500 expression which suggests that they are not part of a complex. So, why these two proteins are in operon? The first possibility is that the two proteins contribute to the same pathway and this can be tested using in vitro assay in the future or Rog3500 acts as a partner to support the correct folding and stability of SelenoN, but this is not supported by the expression results, or maybe we are missing a factor in the expression.

In the next part of the project, we employed the sponge *Aplysina aerophoba* model to study the distribution of *Poribacteria* in the sponge tissue, both under standard and heat stress conditions. In a control sample of *A. aerophoba,* the electron microscopy images showed a remarkable distribution of *Poribacteria,* and an accumulation of the bacteria around specialized sponge cells, the spherolous cells. These cells were demonstrated to produce compounds that act as a defense mechanism for the sponge, even though this is still a matter of dispute since the same compounds were isolated from some bacterial species. In addition, immunogold labeling was performed to confirm the identity of the poribacterial cells by detecting the expression of SelenoN inside the cells. The antibody used in immunogold-labeling are homemade anti-SelenoN antibody designed to target SelenoN-Trx protein copy. However, this antibody only detects one copy of SelenoN, and we aim to examine the expression level of a maximum of SelenoN copies in the context of the sponge tissue. Therefore, newly developed antibodies were designed to target the catalytic site SCUG with the upstream and downstream most conserved amino acids, although tolerating some degree of degeneracy to detect the expression of most of the SelenoN by immunogold-labeling. Moreover, by implementing CARD-FISH techniques on the heat stress sample, we observed a reduction in the proliferation of the *Poribacteria* compared to the control sample

using *Poribacteria* 16S rRNA-specific probes. Furthermore, to examine the effect of the heat stress on the expression of the SelenoN we planned to amplify the signals of the mRNAs in the sponge tissue using CARD-FISH techniques. For that, we designed three degenerated probes targeting the catalytic site UGA and the SECIS sequence to detect the expression level of SelenoN. These experiments are still undergoing. So, what we concluded from the microscopy studies is that *Poribacteria* are enriched in sponge tissue, and are preferentially gather around spherolous cells, that store brominated compounds that exert medicinal properties. Our main question is that does SelenoN participate in the synthesis of the these brominated products? Still lots of investigation need to be performed to answer this question.

VI. Some suggestions on how to improve the previous experiments

To avoid working on a complex metagenome extraction, we can try to restrict the genomic extraction to the sponge-associated prokaryotes (SAPs), purified by density gradient centrifugation or differential centrifugation (Lasken et al., 2005). Next, the *Poribacteria* could be isolated from the purified SAPs by Fluorescence-activated cell sorting (FACS) as previously described in (Siegl & Hentschel, 2010) or based on the protocol from (Lasken et al., 2005). According to Siegl, the *Poribacteria* were sorted according to the cell size without the use of fluorescently labeled antibodies. Then, to confirm the identity of the isolated *Poribacteria* we can run a PCR screening using the 16S rRNA poribacterial-specific primers designed in our study. However, working on the individual isolated *Poribacteria* is still going to be challenging, because not all *Poribacteria* strains have the same copies of *SELENON* genes. Using degenerated *SELENON* primers, it will still be possible to amplify at least some of the distinct identified *SELENON* operons. Then we can run a reverse transcription PCR (RT-PCR) on the RNA extraction from the isolated *Poribacteria* cells to examine the expression level of *SELELNON* mRNAs. Moreover, the isolated *Poribacteria* will be very useful to try to extract the metabolic compounds produced by the *Poribacteria,* or to test the presence of brominated alkaloids, that were suggested to be produced by the sponges in some studies. If we succeeded, it will be very interesting to develop enzymatic assays to reproduce the SelenoN activity *in vitro* and test for the contribution of SelenoN in the biosynthesis of these molecules. A blueprint to optimize the experiment is illustrated in (Figure P.1).

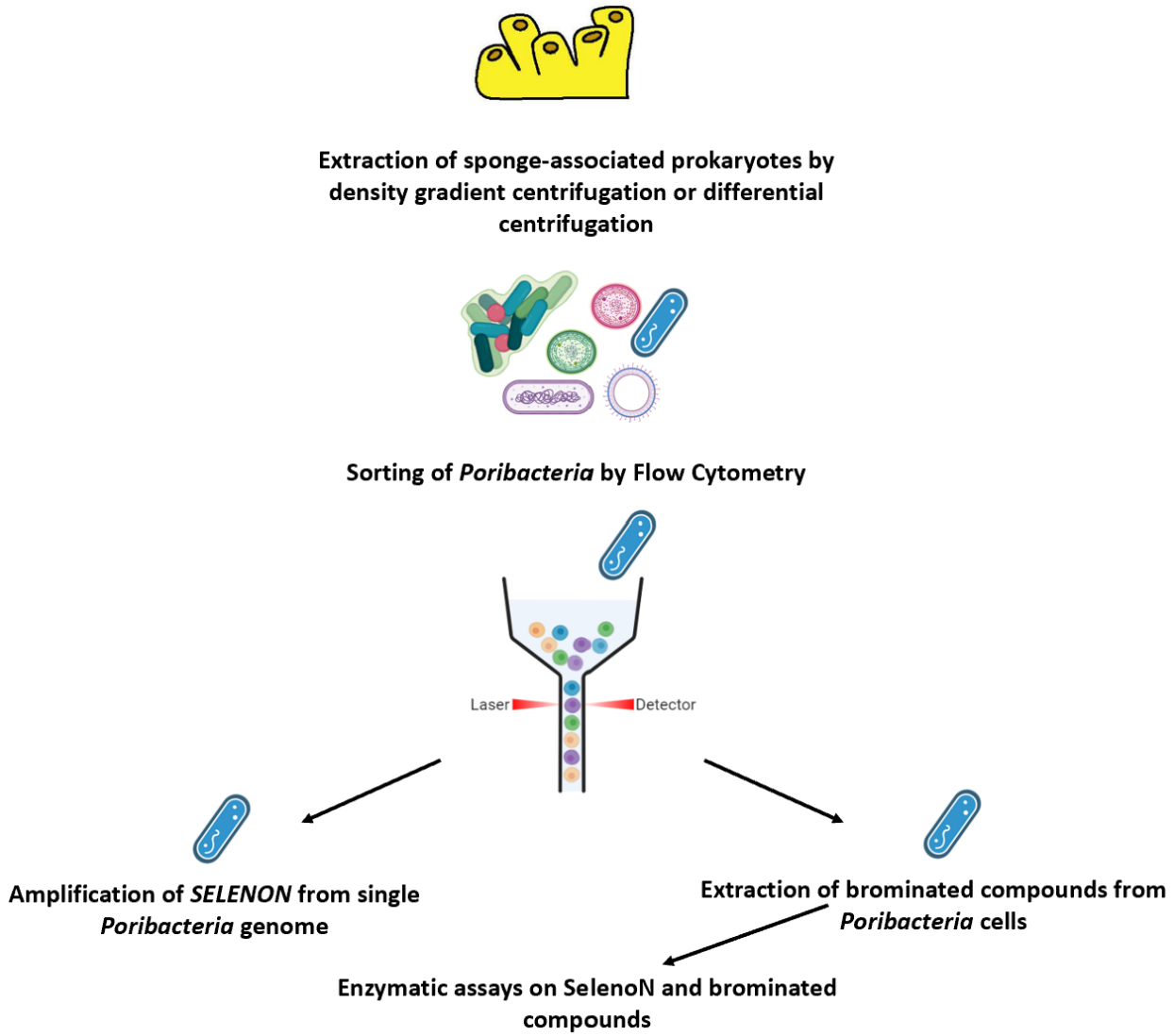


Figure P.1. A blueprint for isolating the *Poribacteria* and further experiments.

VII. Future perspective for the project

This multidisciplinary project involved different research areas ranging from bioinformatics studies, molecular biology, and microscopic investigations and aimed to identify the importance of SelenoN activity for the sponge-*Poribacteria* interaction and to gain in understanding of SelenoN molecular function.

Bioinformatic inspection of *SELENON* genomic context in *C. poribacteria* genomes provided valuable information in regard to the organization of *SELENON* multicopy genes and the combination of *SELENON* in operons with other related genes that suggested their co-expression and their involvement in one common metabolic or signaling pathway. In addition, fusion of most of SelenoN proteins encoded by the different copies to additional domains indicated possible alternative, also related, functions and divergent evolution of *SELENON* into several paralogs.

Heading toward implementing the information obtained from the bioinformatics studies, one natural *SELENON* gene, with its associated gene, was amplified and cloned into expression vectors. After optimization of proteins expression and purification strategies, the recombinant proteins will be used for further biochemical characterization and *in vitro* enzymatic assays, to test different proposed SelenoN activities. In addition, results from the fluorescent *in situ* hybridization (FISH) and electron microscopy studies are expected to contribute to the understanding of a possible role of SelenoN in the sponge-poribacteria interaction, but also it is expected to provide us with essential information about SelenoN molecular activity. Therefore, the results obtained will be translated to eukaryotic models and tested on cellular and animal models. In addition, this information will be used to understand the pathophysiological muscular dysfunction that will help in the design of targeted pharmacological interventions.

Furthermore, as mentioned previously in the introduction, limited fraction of sponge-associated bacteria could be cultivated on laboratory developed media, and all attempts to design culture conditions to grow *C. poribacteria* outside of its sponge host failed so far (Gutleben et al., 2020). However, further investigation in collaboration with specialists in marine sponges could alleviate this present limitation to cultivate *C. poribacteria* in the lab and provide opportunities to conduct genetics/mechanistic analyses.

MATERIALS & METHODS

I . Materials

1. Cloning

Table 1. Plasmids and Enzymes

		Specification	Supplier
Plasmids	pABC2	Expression vector for Sec insertion, streptomycin resistance	(Cheng et Arnér, 2017)
	pJET 1.2	PCR cloning, ampicillin resistance	Thermo-scientific
Restriction Endonuclease	Nde1		Thermo-scientific
	Eco52I (EagI)		Thermo-scientific
	NotI		Thermo-scientific
	EcoR1		Thermo-scientific
	BAMH1		Thermo-scientific
	Xba1		Thermo-scientific
	Xho1		Thermo-scientific
	Bgl II		Thermo-scientific
	Bfm1		Thermo-scientific
	Alu1		Thermo-scientific
	Mlu1		Thermo-scientific
	Pst1		Thermo-scientific
	Nsi		Thermo-scientific
	Taq polymerase		Thermo-scientific
Polymerases	10x Taq polymerase buffer		Thermo-scientific
	High Phusion DNA polymerase		Thermo-scientific
	5X High Phusion DNA polymerase buffer		Thermo-scientific
Ligases	T4 DNA ligase		Thermo-scientific
	T4 DNA ligase buffer		Thermo-scientific
	deoxyribonucleoside triphosphates (dNTPs)		Lab preparation

Table 2. 16S rRNA primers

Probe	Sequence (5'-3')	Specificity	Reference
27f *	AGAGTTTGATCMTGGCTCAG	Universal modified	(Fieseler et al., 2004)
1492r	GGTTACCTTGTTACGACTT	Universal	(Fieseler et al., 2004)
27f-POR-sp	AGAGTATGATACCGGCTCAG	Universal modified	This study
1492r-POR-sp	GGCTACCTTGTTACGACTT	Universal modified	This study
POR389f	ACGATGCGACGCCGCGTG	*C. Poribacteria*	(Fieseler et al., 2004)
POR1130r	GGCTCGTCACCAGCGGTC	*C. Poribacteria*	(Fieseler et al., 2004)
515Fsp	GTGCCAGCAGC**Y**GCGGTAA	Universal modified	(Podell et al., 2019)
806Rsp	GGACTA**S**CGGGGTATCTAAT	Universal modified	(Podell et al., 2019)

* The universal 27f is modified from the universal 8f : 5' AGAGTTTGATC**C**TGGCTCAG 3' (Turner et al. 1999)

Table 3. SELENON and ROG3500 Primers for inverse PCR

Probe	Sequence (5'-3')	Specificity	Reference
ROG31f	TGGATCGTYGAYCAGATTACCGA	ROG3500	This study
ROG32r	TCAATCACRTCAAGCGGYGTGC	ROG3500	This study
ROG33f	ACTCGACCTTAAAGACGGCACA	ROG3500	This study
ROG34r	AGTGATTTCGACCAATGCGACTC	ROG3500	This study
ROG35f	GCACACCGCTTGATGTGATTGA	ROG3500	This study
ROG36r	TCGGTAATCTGATCAACGATCCA	ROG3500	This study
ROG45f	GAGATAAACACTGAAGAGGAGTCTG	ROG3500	This study
ROG46r	CGGAGGTTCTCTTTAACTGACAAC	ROG3500	This study

2. Protein expression & purification

Table 4. Cells and Media

C321ΔA	**Expression strain. (Lajoie et al., 2013). Addgene (catalogue no. 488998).**
BL21De3	Expression strain
Arctic cells	Expression strain
DH5α	Cloning strain
LB medium	10g/L Trypton, 5g/L Yeast extract, 5g/L NaCl. pH = 7
2YT medium	16g/L Trypton, 10g/L Yeast extract, 5g/L NaCl, pH = 7
2YT plates	10g/L Trypton, 5g/L Yeast extract, 5g/L NaCl, 15g/L Agar Agar, pH = 7
Antibiotic	Ampicillin antibiotic 100µg/ml
	Streptomycin antibiotic 50µg/ml
	Gentamycin antibiotic 20µg/ml
Expression inducer	Isopropyl-1-thio-β-D-galactopyranoside (IPTG) 0.5mM
	Agar-Agar, Kob1from ROTH/ Trypton from ROTH/ Yeast extract from SIGMA-ALDRICH

Table 5. Lysis buffers and competent cells suspension

Standard 1x lysis buffer for bacterial culture 50 mM Tris HCL pH8, 250 mM NaCl, 0.5 mM TCEP, 1 mM AEBSF, 1 mM Benzamidine, 10% Glycerol, milliQ. 10µg/ml DNase (optional if the lysate is viscous, with 10-20 min incubation on ice). 100µg/ml Lysozyme (optional, incubate lysate 30 min on ice before sonication). 1mM Cacl2 / 2.5mM Mgcl2 (Optional to reduce protein contamination and prevent ionic interaction).
Chemo-competent cells suspension solution 10% PEG6000, 5% DMSO, 10mM MgCl2, 10mM MgSo4, 10% Glycerol
4X Radio-Immunoprecipitation Assay (RIPA) buffer used in immunoprecipitation. 400mM Tris-HCL pH 7.5 600mM NaCl 4% Triton 0.4% SDS 8mM EDTA 4% Deoxychilate

Table 6. Western blot

Semi dry transfer: 25mM Tris.HCl; 192mM Glysine; 0.1%SDS; 20% Ethanol in 1 liter 1xbuffer. Wet transfer buffer: 25mM Tris.HCl; 192M Glycine in 1 liter 1xbuffer (in immunoprecipitation) PDVF Immobilon-P Transfer membrane from Millipore; ref, IPVH00010 Whatman Filter paper from GE Healthcare Life Sciences
Blocking solution: 3%BSA in 1x PBS-Tween 0.05% **Washing solutions**: 1x PBS Tween 0.05%, 1x PBS, 3%BSA in PBS-Tween 0.05%
Primary antibodies: 1/4000 dilution of anti-His6 Abs / anti-selenoN1 Abs in 3% BSA in PBS-Tween 0.05%.
Secondary antibodies: 1/ 50,000 dilution of anti-mouse-HRP* / anti-rabbit-HRP* Abs in 3%BSA in PBS-Tween 0.05% *Anti-mouse IgG, horseradish peroxidase linked whole antibody (from sheep) GPR. LNXA931/AE *Anti-rabbit IgG, horseradish peroxidase linked whole antibody (from donkey).
Protocol: **@His Antibodies / @Streptag Antibodies** • Equilibration of the PDVF membrane with 1x PBS-0.05%tween. • Blocking the membrane with 3%BSA in 1x PBS-0.05%tween. • First Antibodies incubation ON @4°C. • 3 washes with 1x PBS 0.05%tween each for 10 min @rt. • Secondary antibodies incubation 1h @rt. • 2 washes with 1x PBS-0.05%tween and 2 washes with 1x PBS for 10min each. HRP substrate incubation 1min and filming.
ON, overnight; rt, room temperature; HRP, horseradish peroxidase

Table 7. Immobilized metal affinity chromatography (IMAC)

Native purification by nickel-coupled Imino-diacetic acid (Ni-*IDA*) column. His60 Ni Superflow Resin 25ml (TaKaRa)	
Lysis buffer	1x lysis buffer (from table 5)
Wash buffer	1x lysis buffer + 20mM Imidazole
Elution buffer	1x lysis buffer + 300mM Imidazole
Denatured purification with nickel-coupled Imino-diacetic acid (Ni-*IDA*) column according to QIAexpressionist handbook (Qiagen)	
Lysis buffer	100mM NaH2PO4 + 10mM Tris.HCl + 8M Urea / adjust pH to 8 by NaOH
Wash buffer	Buffer C: 100mM NaH2PO4 + 10mM Tris.HCl + 8M Urea / adjust pH to 6.3 by HCL
Elution buffers	Buffer D: 100mM NaH2PO4 + 10mM Tris.HCl + 8M Urea / adjust pH to 5.9 by HCL Buffer E: 100mM NaH2PO4 + 10mM Tris.HCl + 8M Urea / adjust pH to 4.5 by HCL

Table 8. Polyacrylamide gel electrophoresis

Sodium dodecyl sulfate–Polyacrylamide gel electrophoresis (SDS-PAGE)
12% Separation gel pH 8.8 3.3ml Milli-Q, 4ml ROTIPHORESE Gel 30, 2.5ml Tris HCl pH 8.8, 100 µl APS, 50 µl TEMED **10% Separation gel pH 8.8** 4ml Milli-Q, 3.3ml ROTIPHORESE Gel 30, 2.5ml Tris HCl pH 8.8, 100 µl APS, 50 µl TEMED **8% Separating gel pH 8.8** 4.6ml Milli-Q, 2.7ml ROTIPHORESE Gel 30, 2.5ml Tris HCl pH 8.8, 100 µl APS, 50 µl TEMED **5% Stacking (resolving) gel pH 6.8** **Composition missing** Before adding APS and TEMED to the separating gel take out 600 µl of the mixture, add 60 µl APS, 3 µl TEMED. **SDS-PAGE running buffer:** 1x TGS (Tris-Glycine-SDS) buffer pH 8.3-8.9 (do not adjust)
Agarose Gel:
1%Agarose gel: 2.5g in 250ml 1x TBE (Tris/Borate/EDTA) **0.8% Agarose gel**: 2g in 250ml 1x TBE **Agarose gel running buffer:** 1x TBE buffer
Coomassie blue stain: 0.1 % G250 Coomassie Brilliant Blue + 20 % Ethanol + 8 % acetic acid
PageRuler Plus Protein Ladder: ThermoFisher SCIENTIFIC

3. Cell Transformation reaction.

Table 9. Chemo-competent cells Transformation
5 µl of the assembly reaction (or try different vol) + 12 µl 5X KCM + 43 µl Milli-Q incubate on ice for 10min, add 60 µl of ice thawed chemo competent DH5α *E. coli*, mix gently by aspiration- incubate on ice 30min to 1h - place @ RT for 5min for heat shock - complete to 1ml with LB – mix gently – incubate @ appropriate temp °c water bath for 30 min. **For ligation transformation:** centrifuge 20sec with mini-spin – remove 850 µl of the supernatant and resuspend the pellet with the 150 µl left then plate the full transformation on a 2YT agar plate with appropriate antibiotics. **For plasmid transformation:** plate 1/10th volume of the culture. **DH5α *E. coli* optimum temperature *is*** 37°c. **C321ΔA optimum temperature is** 30°c.

Table 10. Electro-competent cells Transformation
• Thaw electro-competent cells on ice. • Transfer 10 to 20ng of recombinant plasmid or ligation reaction to a pre-chilled electroporation cuvette followed by 50 µl of electrocompetent cells (the vol of competent cells increased to 60 µl for ligation reaction **and the insert:vector ratio is calculated by NEB calculator website**). • Mix gently by pipetting up and down. • Carry out electroporation immediately by electroporator for the electric shock. • Add 500 µl of room temperature LB or 2YT media to the cuvette immediately after electroporation. • Place the tube in a water bath at the appropriate temp °C for the used cell line and incubate for 30- 60 minutes, invert the tube every 15 min. • Centrifuge 1min at 11000 rpm, remove 400 µl of the supernatant, resuspend the cells in the remained supernatant and plate 100 µl. • Warm selection plates to the appropriate temp °C before plating and incubate the plated petri-dish overnight at the appropriate temp °C.

4. Chemicals

Table 11. Chemicals

Chemical	Reference	Supplier
NaCl	3957.1	ROTH
AEBSF hydrochloride	FA29609	BIOSYNTH
TCEP	Reducing agent.	B59-10G
Benzamidine	B6506-25G	SIGMA-ALDRICH
H_2O_2	STBJ8210	SIGMA-ALDRICH
Sucrose	1.07687.1000	Millipore
Lysozyme	89833	Thermo-Scientific
DNase	10104159001	SIGMA-ALDRICH
Triton X-100	Detergent- SLBJ81129v	SIGMA-ALDRICH
N-Lauroylsarcosine (Sarkosyl)	61739	SIGMA-ALDRICH
Tween	Detergent t.9127.1	ROTH
Np40	492016	SIGMA-ALDRICH
CHAPS	Detergent 26680	Fluka
ROTIPHORESE®Gel 30 (37.5:1)	3029.2	CARL ROTH
TRIzol	T9424	SIGMA
SDS	436143-25G	SIGMA-ALDRICH
BSA Bovine serum albumin	04-100-812-C	EUROMEDEX
Phenol:chloroform:isoamyl alcohol (25:24:1)	A156.1	(ROTH;25:24:1)
Glycerin	3783.1	ROTH
Ethanol absolute	Alcohol 20821.310	VWR
Chloroform	1731042	ROTH
Sodium Acetate (NaAc)	127-09-3	SIGMA-ALDRICH
Isopropyl β-d-1-thiogalactopyranoside (IPTG)	I6758	SIGMA-ALDRICH
Proteinase K	49185724	Roche
Hydrogen peroxide solution	STBJ8210	SIGMA
Ampicillin sodium salt	A9518-5G	SIGMA-ALDRICH

5. Commercial Kits.

Table 12. Commercial kits

Commercial kit	Supplier
Western blot Super Signal TM Western Blot Enhancer	Thermo-Scientific
(Nucleosping plasmid kit-MN)	740588.250-MN
PCR clean up kit	11992242-MN
FastDNA SPIN Kit (MP)-Lot No.99219	Millipore
Bicinchoninic Acid Protein Assay Kit (BCA)	Thermo-Scientific
OxyBlot^{TM} Protein Oxidation Detection Kit (S7150)	Millipore

6. Northern blot.

Table 13. Selenoprotein N specific Probes / degenerated sequences

Probe	Sequence (5'–3')	Reference	Company
Probe1	TCCCTCAGCANGMYTCRTC	This study	IDT
Probe2	TTTRCCRSTCSCTCAGCA	This study	IDT
Probe3	ACCTGCCCTCARRYYTTT	This study	IDT

Table 14. Northern blot solutions

- 10%SDS to wash all utilities.
- 0.5% SDS sterilized to strip the blot.
- 0.5 M EDTA 50ml.
- 50mM Sodium acetate NaAc 1L.
- 2N NaOH → 4g NaOH in 50ml
- 0.05 N NaOH to hydrolyze RNA and improve the efficiency of transfer.
- DEPC treated water. (replaced with Milli-Q).

Gel running buffer-5x MOPS 1L: 0.1M MOPS (pH7) 20.6g + 800ml 50mM Sodium Acetate (NaAc) (adjust to pH7 with 2N NaOH) + 10ml 0.5M EDTA pH8, complete to 1L with milli-Q. (Final Conc = 40mM NaAc, 5mM EDTA). Sterilize solution by filtration using 0.2 micron Millipore, Keep the solution in dark, @4°c or RT.
Gel Running buffer 1L = prepare 1X MOPS/ add formaldehyde only in agarose and sample.

Membrane transfer buffer 20x SSC: 3M NaCl 175.3g + 300 mM Sodium Citrate dehydrogenase 88.2g. Dissolve NaCl in 800ml H2O, then dissolve Sodium Citrate into the NaCl solution (adjust pH to 7.4), complete to 1L H2O, autoclave.

Wash solution: 20x SSPE for 5 L: 870g NaCl +138g NaH2PO4 + 37g EDTA adjust pH to 7.4

Agarose gel preparation with formaldehyde: 200ml volume: 1.6g agarose dissolved in 124.3ml Milli-Q → boiled, cooling down to 60°c then 40ml of 5xMOPS and 2.2 M formaldehyde 35.7ml (stock12.3 M) was added just before pouring.

Formaldehyde gel Loading buffer: 50% Glycerol + 1mM EDTA pH8 + 0.25% bromophenol blue (stock 1%) + 0.25% xylene cyanol FF (stock 2.5%).

Hybridization solution for 50ml: 2.5 ml Denhardt 100X + 15 ml 20X SSPE + 2.5 ml SDS 10% + Milli-Q complete to 50 ml.

Sample preparation for loading into the gel: RNA up to 30µg = 4.5µl + 5x formaldehyde gel-running buffer 2µl + Formaldehyde 3.5µl + Formamide 10µl + EtBr (1mg/ml) 1µl → denature sample at 65°Cfor 5 min, chill on ice, then add formaldehyde gel-loading buffer 2µl. Ready to load.

Probes labeling:
Mix: 10p mol oligo (60ng for a 20 mer) + 1.5µl 10x Polynucleotide kinase buffer + 1.5 µl 100 mM DTT + 2.5µl [gamma-^{32}P]ATP + H_2O to 15µl final volume +1µl T4 Polynucleotide kinase.
Incubate 30min at 37°C →Denature 5min at 65°C.
Complete the volume to 50µl by adding 35µl of hybridization solution.
Spin the filtration column down @4000rpm 30s to get rid of the storage sol.
Load the labeled probes into the column and place an empty Eppendorf underneath, spin down @4000 30s. Collect the probes and perform the counting cpm count per minute as in the picture using the instrument LS 6500 Multi-Purpose Scintillation Counter.

7. Fluorescence *in situ* hybridization (FISH)

Table 15: CARD-FISH horse-radish peroxidase (HRP) labeled oligonucleotide probes, and the DNA base helper (H-) probes of 16S rRNA specific to *Poribacteria.*

Probe	Sequence (5′–3′)	Label	Reference	Company
POR600	CCGAACCCTTTCACGTCT	HRP	(Fieseler et al.,2004)	BIOMERS
POR600	CCGAACCCTTTCACGTCT	Cy3	(Fieseler et al.,2004)	BIOMERS
POR1130	GGCTCGTCACCAGCGGTC	HRP	(Fieseler et al.,2004)	BIOMERS
POR1130	GGCTCGTCACCAGCGGTC	Cy3	(Fieseler et al.,2004)	BIOMERS
POR619H	CGCAGTGGCTCGGTTGAG	NONE	This study	BIOMERS
POR582H	GACTTACTCCACCGCCTA	NONE	This study	BIOMERS
POR1131H	TCATCCCCACCTTCCTCT	NONE	This study	BIOMERS
POR1112H	TCTTTAGAGTCCCCAGCA	NONE	This study	BIOMERS
EUB338I	GCTGCCTCCCGTAGGAGT	HRP	Amann et al. (1990)	BIOMERS
NON338	ACTCCTACGGGAGGCAGC	HRP	Amann et al.(1995)	BIOMERS

Table 16: CARD-FISH Solutions

Fixative Solution: 4% paraformaldehyde in 0.02 M phosphate buffer pH 7.4 stored at 4°C
Stock solutions to prepare in Milli-Q H2O • (1 L) 1M Tris. HCl pH 7.5 pH • (0.5 L) 5M NaCl • (250 ml) 0.5M EDTA pH 8.0 • 10x PBS Adjust pH to 7.6 All solutions should be autoclaved, then keep small sub-stocks (50 ml) of each one in falcon tubes.
Lysozyme preparation. Lysozyme buffer 100mM Tris.HCl, 50mM EDTA pH 8.0 Measure 10mg/ml Lysozyme in 1ml lysozyme buffer.
Low melting point agarose solution for embedding filters. LMP agarose of 0.1% in 20ml in a small beaker. The solution is heated carefully in a microwave and checked frequently for dissolution by swirling. It is then left to cool (<30 °C). Note that LMP agarose starts to gel at about 26 °C.
Blocking buffer 10% blocking reagent (Roche ref: 11096176001: 50 g) in 10X Maleic acid buffer Add 10g blocking reagent to 100ml 10x Maleic acid buffer (below) Heat to ~60 °C on a heated stirrer for up to 1 hr or until dissolved. Autoclave and then store aliquots (10ml) in the fridge Open aliquots aseptically as this solution can become easily contaminated.
10X Maleic acid buffer 100mM Maleic acid + 250mM NaCl, pH 7.5 Adjust pH with solid or concentrated NaOH. Autoclave.

Table 17: Hybridization buffer (30%-10% formamide)

Final Concentration	Stock	Vol. stock for 20 ml hybridization buffer
0.9 M NaCl	5M	3.6ml
20 mM Tris pH 7.5	1M	0.4ml
10 % dextran sulfate (500 Kd)	Powder	2g
1% blocking buffer	10%	2ml
Formamide 30% / 10%	>99.5%	6ml / 2ml
0.02% SDS	10% SDS	40µl
milli-Q		8ml / 12ml

- Put the NaCl, Tris, H_2O and dextran sulfate in a small beaker and stir with a magnetic bar with gentle heating until the dextran sulfate is dissolved
- Add the formamide and blocking buffer (cover with parafilm to reduce evaporation of the formamide) and stir well before adding the SDS
- Make 1 ml aliquots and store at –20 °C

Table 18: CARD-FISH Wash buffer

Final Concentration	Stock	Vol. stock for 100 ml wash buffer
X mM NaCl (see table below)*	5M	Xµl*
20 mM Tris pH 7.5	1M	2ml
5 mM EDTA	0.5M	1ml
0.01% SDS	10%	100µl

*Volume of NaCl to use to give equivalent stringency of formamide concentration in hybridization buffer. More details can be found on www.arb-silva.de, FISH & Probes section Bremen, Germany, October 2010, v 2.2

Formamide in hybridization buffer	NaCl in wash buffer (48 °C)	Volume NaCl in 100ml wash buffer
30 %	0.112M	2.04ml
10 %	0.45M	9ml

Table 19: CARD-FISH Amplification buffer and Reaction mix

CARD Amplification Buffer (final volume 40 ml)
2ml of 20x PBS, pH 7.6 (important for proper enzyme function) (0.1%) 0.4 ml Blocking Reagent (stock 10%, see above) (2M) NaCl 16ml (stock 5M) add Milli-Q to a final volume of 40ml add 4g of dextran sulfate (10% [w/v]) Heat (40 to 60°C) and shake until the dextran sulfate has dissolved completely. The amplification buffer can be stored in the refrigerator for several weeks.
CARD Reaction Mix
The CARD substrate mix was prepared by mixing amplification buffer (above) with a freshly prepared H_2O_2 solution (0.15% in 1x PBS) at a ratio of 100:1 then the substrate **Alexa Fluor™ 488 tyramide reagent (Invitrogen by Thermo-Scientific-LOT 2301004)** was added at a ratio of 1 part tyramide to 200-1000 parts amplification buffer. Here we used 1:200 ratio.
Staining
Citifluor:Vectashield, 4:1 ratio with DAPI [2.5µg/ml]), DAPI stock = 250µg/ml

Table 20: CARD-FISH List of reagents

Reagent	**Reference**	**Supplier**
Formamide	SI-F9037-100ml	Analytic Lab (Sigma)
Blocking Reagent	11096176001	Roche Diagnostics
Lysozyme	62971-10G-F	Analytic Lab (Sigma)
Formamide deionized, molecular biology-100ml	SI-F9037-100ml	Analytic Lab (Sigma)
Dextran sulfate 10 g	SI-D8906-10g	Analytic Lab (Sigma)
SDS 10% 100 ml	SI-71736-100ML	Analytic Lab (Sigma)
TSA Reagent, Alexa Fluor 488 Tyramide	15251116	Fisher Scientific (Life Technologies)
Hydrogen Peroxide	SI-216763-100ML	Analytic Lab (Sigma)
3x flat ended tweezers	11714474	Fisher Scientific (Millipore)
Vectashield mounting medium 10ml	H-1000	Clinisciences
Citifluor:Vectashield	AF1	Citifluor Ltd
DAPI	D1306	Invitrogen by Thermo-Scientific
Maleic acid (Sigma ref: M0375)	M0375	Sigma
Tissue Section Adhesive	REF 86014	Thermo-Scientific
Alexa Fluor™ 488 tyramide reagent	LOT 2301004	Invitrogen by Thermo-Scientific

8. Electron Microscope

Table 21: Electron Microscope Solutions

Fixative-1. 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer pH 7.4 **Fixative-2. 3% paraformaldehyde and 0.5% glutaraldehyde in 0.02 M phosphate buffer pH 7.4**
Wash solution: 0.1M sodium cacodylate buffer pH 7.4.
Osmium fixation solution: 2ml 4% OsO4 + 2ml sodium cacodylate buffer 0.4M (pH 7.4) + 4ml d.H2O = final volume 8ml.
Impregnation/ Embedding solutions: Propylene oxide pure 99.5% (cover tightly volatile). LR WHITE pure (for Immune-cytochemistry EM) EPON 812 (there are four reagents → Epikote 812/ DDSA / MNA/ DMP30, see preparation table 22).
Incubation buffer: 10mM PBS +150Mm Nacl + 0.2% (AURION BSA 10%) (pH 7.3), filter with 0.2 μm pore size. Note adjust the pH after adding the BSA.
Blocking buffer: AURION BLOCKING SOLUTION
Staining: Uranyl acetate and lead citrate

Table 22: EPON 812 preparation:

Reagent	EPON A	EPON B
Epikote 812	25ml	50ml
DDSA	50ml	//////
MNA	///////	50ml
For 50ml Vf of EPONE = 40% epon A (20ml) + 60% epon B (30ml) + 1ml DMP30		
The solutions Epon A et epon B can be stored at 4°c. The final Epon mix (aliquotes) can be stored at -20°c.		

9. Instruments

Table 23: Instruments utilized in all experiments

Instrument	Supplier
Micro-Ultrasonic Processor.	Vibra Cell 75022
Medium-Ultrasonic Processor.	Ultrasons Annemasse
NanoDrop	ND-1000 Spectrophotometer
Electrophoresis power supply	BIO-RAD / amilabo ST3002
Agitateur Mini Rocker-Shaker PMR-30	Grant-Bio
Block heater	Stuart-SBH130D
Magnetic Stirrer	SB161-Stuart
Water bath	BIOBLOCK SCIENTIFIC-Polystat 86602
Cell density meter-Ultrospec 10	Amersham Biosciences
3510 pH Meter	JENWAY
Spectrophotometer	(DeNovix spectrophotometer/Fluorometer DS-11 FX+)
TRANS-BLOT SD SEMI-DRY TRANSFER CELL	BIO-RAD
SimpliAmp Thermal cycler	Applied biosystems by life technology
T100 Thermal cycler	BIO-RAD
Chemi-Doc Touch Imaging System	BIO-RAD
Centrifuge 5424 R	Eppendorf
Avanti- J-E Centrifuge	BECKMAN COULTER
SORVALL RC 6+ Centrifuge	Thermo SCIENTIFIC
Incubator Shaker Series	INNOVA 42-New Brunswick
FastPrep® instrument homogenizer FP120	BIO101-Thermo Savant
LS 6500 Multi-Purpose Scintillation Counter	BECKMAN COULTER
Microplate Reader-iMark	BIO-RAD
Gel Doc EZ Imager	BIO-RAD
Weight Balance	METTLER TOLEDO
Weight Balance	Sartorius
Ultramicrotome Leica Ultracut R	Leica Microsystems
Microtome	(MICROM HM 340E Thermo-SCIENTIFIC)
Widefield UPRIGHT FLUORESCENCE microscope	(OLYMPUS-PROVIS)

10. Bioinformatics tools

Table 24: Bioinformatics tools

BLAST https://blast.ncbi.nlm.nih.gov/Blast.cgi
MUSCLE https://www.ebi.ac.uk/Tools/msa/muscle/
Clustal Omega https://www.ebi.ac.uk/Tools/msa/clustalo/
WebLogo https://weblogo.berkeley.edu/logo.cgi
Operon-mapper https://biocomputo.ibt.unam.mx/operon_mapper/index.cgi
CD-HIT Suite http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit
Genomic sequences are available on: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/Candidatus%20Poribacteria
Protein sequence analysis: https://web.expasy.org/protparam/

II . METHODS

1. Bioinformatics studies on the *Poribacteria* genome.

Bioinformatics analysis of *SELENON* genomic context by mapping the *Poribacteria* genomic organization.

Searching for orthologous genes for *SELENON* in other species, identified *SELENON* in prokaryotes, in the specific candidate phylum *Poribacteria*, through a Basic Local Alignment Search Tool (BLAST) using Human SelenoN protein sequence as a query. In fact, studying the genomes of the different *Poribacteria* strains was important to analyze genes linked to *SELENON* as part of operon(s) (co-transcribed genes) or part of a conserved block of synteny (co-evolved genes), and to have a primary prediction of SelenoN possible molecular function. The bioinformatics studies were conducted by Luc Thomes and Laurence Despons.

Contigs search: *Poribacteria* genomes were previously identified and sequenced by single-cell amplified genomes (SAGs) and metagenome-assembled genomes (MAGs) (Podell et al., 2018; Engelberts et al., 2020). As a first step all contigs containing at least one *SELENON* homologous gene were retrieved from a tBLASTn search (NCBI) on the whole-genome shotgun (WGS) contigs database using the organism search set *Candidatus Poribacteria* (taxid:265317).

Next, a python script (Python 2.7.17) was developed using the ORF-finder tool to recognize the Sec-TGA reprogrammed inframe codon and search for the possible next stop codon for translation termination, as without this script the recognition of the *SELENON* gene will only be limited to the N-terminal part. In addition, possible frameshifts within some ORFs were curated by manual annotation using ApE-plasmid editor 2.0.61.

Then *SELENON* ORFs flanked by 10,000 upstream and downstream nucleotides were extracted from each contig, using a developed Python script that replaced *SELENON* Sec-TGA with a Cys-TGT codon. In addition, this script is enforced with an artificial start or stop codon to compensate for the missing coordinates at the extremities of the contigs.

Operon-mapper* analysis*: Two criteria were used for operon detection in each contig, the coordinates for the bait *SELENON* gene within each contig and the encoded protein sequence. ORF-finder tool with the ATG and alternative initiation codons parameters were used to determine gene coordinates using the *Poribacteria* WGS sequences (NCBI).

The modified nucleotidic sequences containing the *Poribacteria SELENON* ORFs were submitted to the Operon-mapper tool (https://biocomputo.ibt.unam.mx/operon_mapper/) using default parameters. Operon-mapper generated data concerning the predicted operons with their corresponding genes and coordinates, as well as the predicted proteins and their functional UniprotKB or COG (clusters of orthologous groups of proteins) descriptions extracted from the relevant databases. Finally, genes in the vicinity of *SELENON* were clustered using Clustal Omega (Clustal Omega < Multiple Sequence Alignment < EMBL-EBI) and CD-HIT Suite (CD-HIT Suite (ucsd.edu)) (Biological sequence clustering and comparison) to find consistent genes represented in *SELENON* environment.

2. Amplification and cloning of *Poribacteria* Selenoprotein N genes

2.1. *Aplysina aerophoba* sample collection

For molecular experiments, sponge individual samples were collected in collaboration with Marcelino Suzuki by the Diving Service of the Observatoire Océanologique de Banyuls-sur-Mer (OOB, Bruno Hess) from the Mediterranean sea in April 2021 at the Plage Sana near the Sphynx. https://user-images.githubusercontent.com/20677440/140489614-1447fa5f-282d-43a5-b02b-3a642b3cb688.png. Some sponge samples were immediately stored in liquid nitrogen at -20°C for molecular analysis, while sections were preserved in 4% paraformaldehyde for electron microscopy (Figure M1) shows an example of *A. aerophoba* individuals and the cut made for electron microscopy experiment.

Another collection of sponge samples was carried out for the sponge-stress experiment by the same diving team on the 30th of Jun 2022. Sponges were maintained in a seawater aquarium at 20°C for one week before being exposed to different stress conditions as detailed in (Table M1).

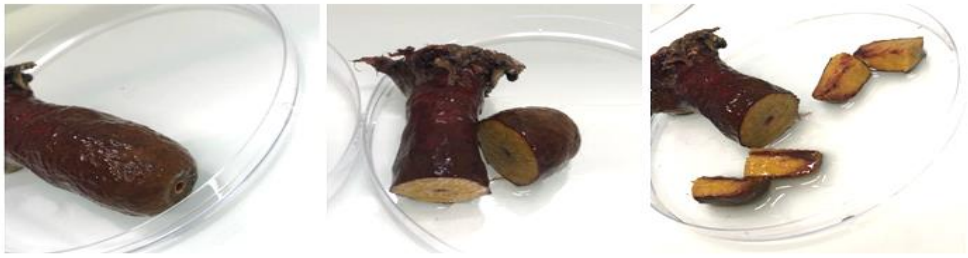


Figure M1.a| ***Aplysina aerophoba sample.*** Is a yellow bright sponge, that turns black when exposed to O_2.

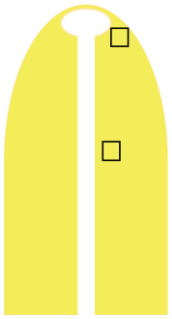


Figure M1.b| A schematic view of the cut made on Aplysina aerophoba for electron microscopy examination. The squares represent the site of the incisions. At the top (apex), and in the center of the sponge.

Table M1. Aplysina aerophoba **collection and stress conditions.**

Sample	Stress	Time
IND2A	Scar	An incision was made from the center of the scar after 2 weeks of the cut.
IND2B	No stress (control for IND2A)	An incision was made from the unwounded half of the sponge individual.
IND2E	Scar	An incision was made from the center of the scar after 24h
IND2F	No stress (control for IND2E)	An incision was made from the unwounded half of the sponge individual.
IND5	No stress (control at 20°C)	An incision was made as a control for the heat stress sample.
IND6	Acute heat stress at 27°C	An incision was made after 24h exposure to 27°C
IND9	Chronic heat stress at 27°C	An incision was made after 2 weeks of exposure to 27°C
Abbreviation. IND, individual; h, hour.		

2.2. Metagenome Extraction from *Aplysina Aerophoba* sponge holobiont

Meta-genomic DNA was extracted by tissue homogenization and centrifugation from 200mg of *A. Aerophoba* sponge tissues frozen in liquid nitrogen using 1ml of FastDNA SPIN Kit (MP) CLS-TC lysis solution. RNase A, DNase-free (100 µg /ml) was added and the lysate was incubated at 37°C for at least 1 hour with shaking every 20 min. Next Proteinase K (40 µg/ml) was added to eliminate contaminating proteins and nucleases and incubated at 50°C for 1 hour. Lysate was then transferred to a 2ml Lysing Matrix A tube (MP-Bio) containing two 1/4th inch ceramic spheres, and homogenized twice in a tissue and cell homogenizer FastPrep® instrument (MP-Bio) for 40s at a speed setting of 6 m s−1 and return on ice in between for 3 min to cool down the sample. This step allows a mechanical lysing of the cellular membrane and extraction of the molecular material. Of note, the ceramic sphere's mechanical action can cause shearing and fragmentation of the genomic DNAs. Then, to precipitate cellular debris, the sample was centrifuged at a maximum speed of 14,000 g at room temperature for 10 min. The DNAs were then extracted from the supernatant three times with one volume phenol:chloroform:isoamyl alcohol (ROTH; 25:24:1), and the aqueous phase was transferred to an empty Eppendorf tube. Metagenomic DNA in the aqueous phase was then precipitated with three volumes of pure ethanol and 1/10 volume of 3M sodium acetate, incubated overnight at -20°C. Next, the sample was centrifuged at a maximum speed of 14,000 g at 4°C for 1 hour, the supernatant was removed carefully, and 500µl of 70% ethanol was added to the pellet, and centrifuged again at the same speed for 10min. Finally, the supernatant was removed and the pellet was air-dried for 20-30 min at room temperature with observation. The pellet was then dissolved in 50 µl milli-Q. Then, since the OD 260:230 nm ratio was too high in the prepared sample, overnight dialysis of the resuspension was performed to get rid of the salts, and the precipitation procedure was repeated with ethanol and sodium acetate. Finally, the DNA pellet was resuspended in 50 µl TE buffer and incubated overnight at 4ºC for complete resuspension. A dilution of 10ng and 50ng was prepared for downstream molecular analysis, and stored at -20ºC.

2.3. 16s rRNA amplification of *Poribacteria*.

To search for the *SELENON* gene, we first needed to validate the presence of the *Poribacteria* in the sponge sample. To address this issue, the 16S rRNA gene of the *Poribacteria* was PCR amplified from the extracted metagenome of the *A. aerophoba* sponge sample using poribacterial-specific 16S rRNA primers.

First, we used the universal eubacterial primers set 27f/1492r to validate the quality of the extracted genomic DNA. Since this universal eubacterial primers set displays multiple mismatches with the *Poribacteria* 16S rRNA sequence, we modified the primers set to adapt to the *Poribacteria* sequence. These modified primers were then used in an external PCR run of a nested PCR approach. The resulting amplicon from the external PCR run was then used as a template in an internal nested PCR with two primer sets specific for *Poribacteria* (POR389f/POR1130r and a modified universal primer set 515Fsp/806Rsp). The poribacterial-specific internal primer sets were adapted from the studies of (Fieseler et al., 2004; Podell et al., 2019)A schematic illustration of the amplification strategy is shown in (Figure M2).

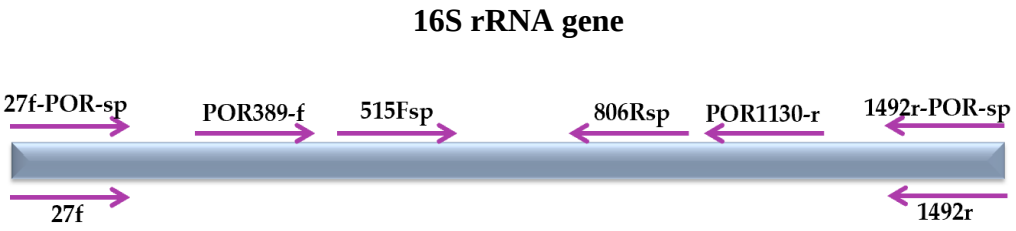


Figure M2. Schematic representation of the amplification of the 16S rRNA gene sequence of the *Poribacteria* from the *A. aerophoba* sponge tissue using a nested PCR approach. Abbreviation: sp, specific; f, forward; r, reverse; POR, Poribacteria.

The PCR cycling conditions for the external and internal PCR reactions are depicted in (Figure M3.A) Note that metagenomic-DNA was pre-heated with the required amount of milli-Q water for 10min at 90°C prior to the PCR reaction. Phusion polymerase was used in this PCR reaction according to manufacturer protocol. As a negative control, 1µl milli-Q instead of DNA was added to the PCR reaction. The resulting PCR products were purified using a polyacrylamide gel electrophoresis (PAGE), electro-eluted, phenol:chloroform extracted, and ethanol precipitated.

These purified amplified fragments were then cloned into pJET1.2 and transformed into electro-competent *Escherichia coli* DH5α. Plasmid DNAs were isolated using the miniprep kit (Nucleospin plasmid kit-MN) and were sent for sequencing to https://eurofinsgenomics.eu/. All primers are listed in (Table 2) in the materials section.

2.4. Amplification of *SELENON* operon from metagenomes extraction.

Because the *ROG3500* gene is more conserved in sequence than *SELENON*, it was used as a marker to amplify the *SELENON*-related operon. *ROG3500* was found downstream of one version of *SELENON*, in one operon with an intergenic space of 21 nucleotides (see Results section). Degenerated primers were designed based on the alignment of ten *ROG3500* sequences, retrieved from BLAST on *Candidatus Poribacteria* (taxid:265317) whole-genome shotgun contigs (WGS) database. The most conserved nucleotidic sequences were selected from the alignment and used to design the degenerated primers to be used in a conventional PCR amplification of the central region of *ROG3500*. Then two sets of specific inverted primers from the obtained *ROG3500* amplicon sequence were designed to be used in an inverse PCR approach (Clark & Pazdernik, 2016) as we aimed to amplify the flanking regions of the *ROG3500* (start and end of ROG3500, the 21 calculated spacer, and the upstream *SELENON* gene), a schematic view of the PCR strategy is illustrated in (Figure M4). All primers are listed in (Table 3) in the materials section.

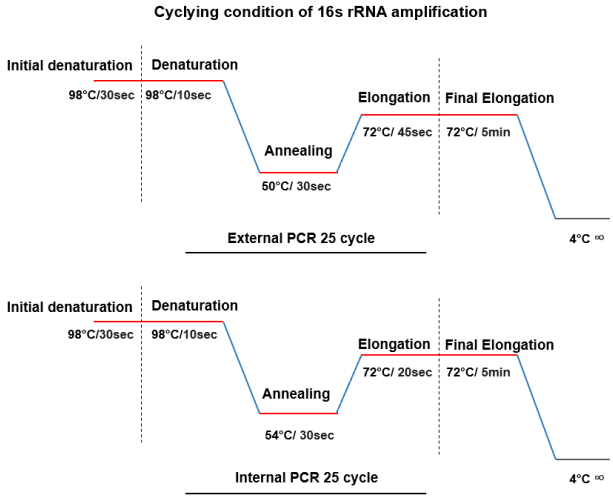
a) Amplification of the central *ROG3500* sequence: Conventional PCR using 20ng of genomic DNA as a template. The PCR cycling conditions for a touch-down PCR reaction are depicted in (Figure M3.B)

b) Amplification of the flanking regions of *ROG3500*:

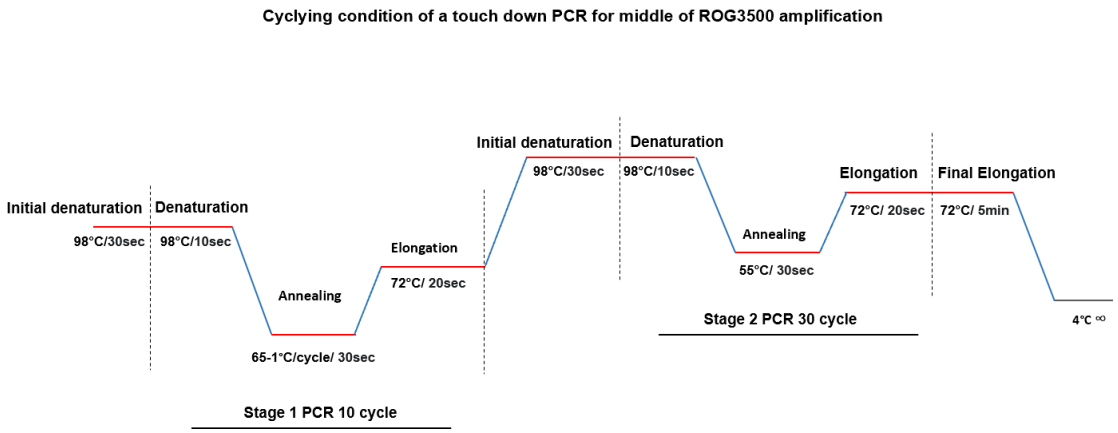
DNA preparation for the subsequent inverse PCR: the metagenomic DNA sample was digested with ten different restriction endonucleases separately. The selection of the enzymes was not arbitrary, but based on the alignment sequences of *SELENON* and *ROG3500* from BLAST search, to avoid the presence of restriction site(s) within the sequence to be amplified. The digestion enzymes used were: EcoR1;Xba1;BamH1;Xho1;BglII;Bfm1;Alu1;Mlu1;Pst1; and Nsi1. The digestion was followed by self-recircularization of the digested DNA fragments by T4-DNA ligase.

Nested-inverted PCR was performed on the digested-recircularized metagenomic DNA using 20ng for the external PCR, and 2µl from the product of the external PCR used in the nested-internal PCR. The PCR cycling conditions for the external and internal-nested touch-down inverted-PCR reaction are depicted in (Figure M3.C). PCR products were purified as mentioned previously in the 16S rRNA amplification. The purified fragments were then cloned into the pJET1.2 vector and transformed into MgCl2-competent *E. coli* DH5α. Plasmid DNAs were isolated as described before and sent for sequencing. Finally, the full *ROG3500* sequence was amplified with direct primers set ROG45f/ROG46r using a conventional PCR method. The primers set was designed based on the amplified upstream and downstream of the *ROG3500*, cycling conditions were identical to the one used to amplify the central sequence of *ROG3500*.

A)



B)



C)

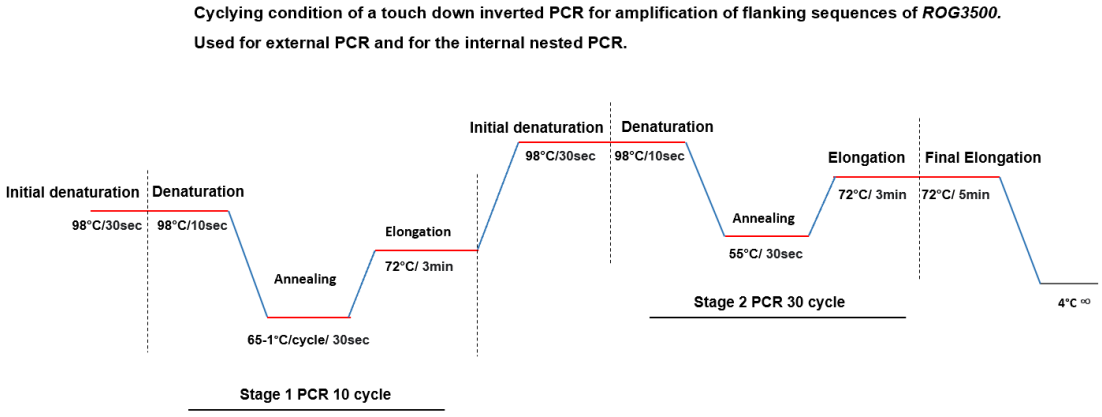


Figure M3. Cycling conditions for the PCR reactions for amplification of SELENON, ROG3500, and 16S rRNA.

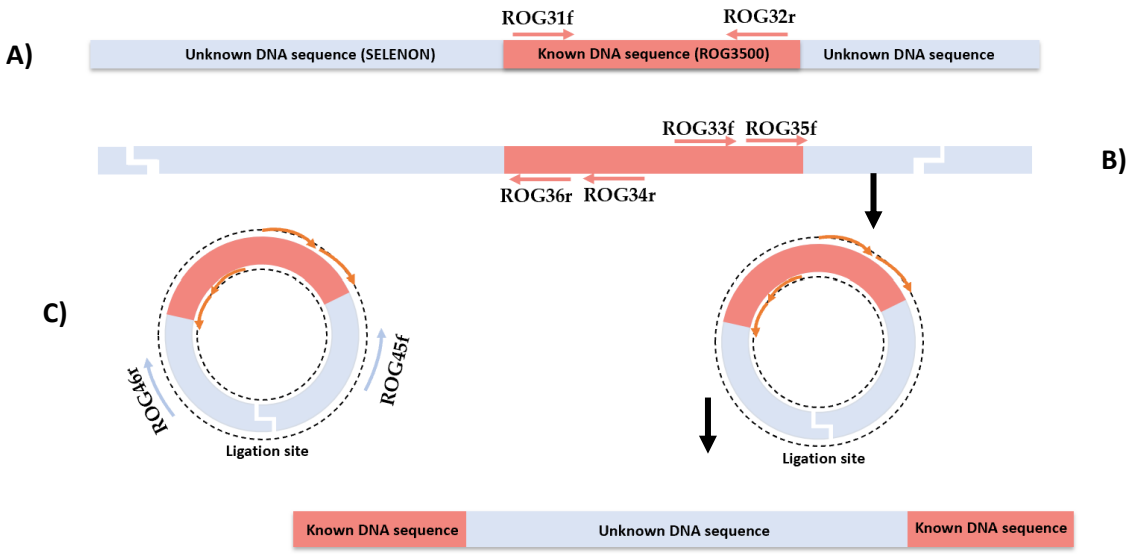


Figure M4. Schematic representation of *SELENON-ROG3500* operon amplification from the metagenome extraction from *A. aerophoba* tissue using an inverted PCR strategy. **A)** Conventional amplification of the central sequence of *ROG3500* using degenerated primers. **B)** Based on the amplified amplicon from the first step new specific primers were designed and used in the inverted PCR. The metagenomic DNA was digested with endonucleases and recircularized. Then, inverted PCR with a nested approach was used to amplify the flanking regions of *ROG3500* including the upstream *SELENON*, the intergenic spacer, and the sequence downstream of *ROG3500*. **C)** Conventional amplification of the complete *ROG3500* with ROG45f/ROG46r primers. The scheme was created by Microsoft PowerPoint.

2.5. Cloning and expression of bacterial Selenoprotein N

Synthesis of selenoproteins in classical *E coli* strains is limited by two factors: the inefficiency of the Sec translation machinery due to its low expression and competition between $tRNA^{Ser[Sec]}$ and release factor 1 (RF1) for the Sec-UGA codon recognition. To increase the expression level of selenocysteinylated proteins, we used the strategy developed by (Cheng & Arnér, 2017). The original paper describing the design of the cloning vector pABC2 is in (Arnér et al., 1999).

The pABC2 cloning vector was used to clone the *SELENON* gene. It is a streptomycin resistant plasmid that encodes genes for the three major factors of the Sec insertion machinery: *SelA*, *SelB*, and a modified *SelC2*, in which *SelC* is mutated to pair with UAG instead of the classical Sec-UGA codon. A map of the pABC2 vector is depicted in (Figure M5).

The C321ΔA expression strain is a genomically recoded organism (GRO) release factor 1 (RF1) depleted *E. coli* strain. It was engineered to recode the UAG amber stop codon as a newly-assigned codon. All the 321 endogenous UAG stop codons were replaced by UAA codons. This engineered strain alleviates the competition with translation termination RF1 on the UAG codon and allows the redefinition of the UAG codon for the introduction of natural or non-natural amino acids into the nascent polypeptide chain. An illustration of the engineered C321ΔA strain is shown in (Figure M6) (Lajoie et al., 2013). The use of the combination of these two tools permits a high expression level of recombinant selenoproteins.

Plasmid digestion: Sequential digestion of 3-5µg of pABC2 by Eag1(Eco52I)/Nde1 restriction endonucleases from Thermo-scientific was performed according to the supplier protocols.

SELENON* constructs*: *SELENON* in an operon with *ROG3500*, and as an individual gene, were constructed and fused to a 6x-histidine-SUMO tag (Small Ubiquitin-like Modifier) at the N-terminus. Alternatively, a GST tag was also added at N-terminus for another construct of *SELENON* alone. An individual *ROG3500* construct was fused to a Strep-tag at the C-terminus. Of note, the *SELENON* sequence amplified from the metagenome presents two sequential Met codons in the 5', therefore, *SELENON* constructs with either one or two ATG was constructed and fused with 6x-histidine tag at the N-terminus. For compatibility with the cloning and expression system, the Sec-TGA codon was reprogrammed as TAG by site-directed mutagenesis in all constructs. In addition, a Cys-TGT mutant was created in parallel for all constructs to

compare the expression with the Sec form and to use in activity tests later on. All bacterial *SELENON* constructs obtained are listed in (Table M2).

Table M2. Bacterial *SELENON* constructs. SCUG / SCCG refers to the amino acids of the catalytic center

Plasmid construct	Description
pABC2-**6His-SUMO**-SELENON-SCU/CG-ROG3500-**STREP**	SELENON-ROG3500 Operon
pABC2-**6His-SUMO**-SELENON-SCU/CG	SELENON Individual gene
pABC2-**GST**-SELENON-SCU/CG	SELENON Individual gene
pABC2-**6His**-SELENON-SCU/CG-(2ATG)	SELENON Individual gene
pABC2-**6His**-SELENON-SCU/CG-(ATG)	SELENON Individual gene
pABC2-ROG3500-**STREP**	ROG3500 Individual gene

All PCRs were performed using the Phusion High-Fidelity DNA polymerase and the dream Taq DNA polymerase for the PCR diagnostic screening.

To analyze PCR reactions, a mix of 5 µl reaction + 5 µl mQ + 2 µl DNA loading buffer was loaded on 1% or 0.8% agarose gels (according to fragments length) in TBE buffer and run at 80-100mV.

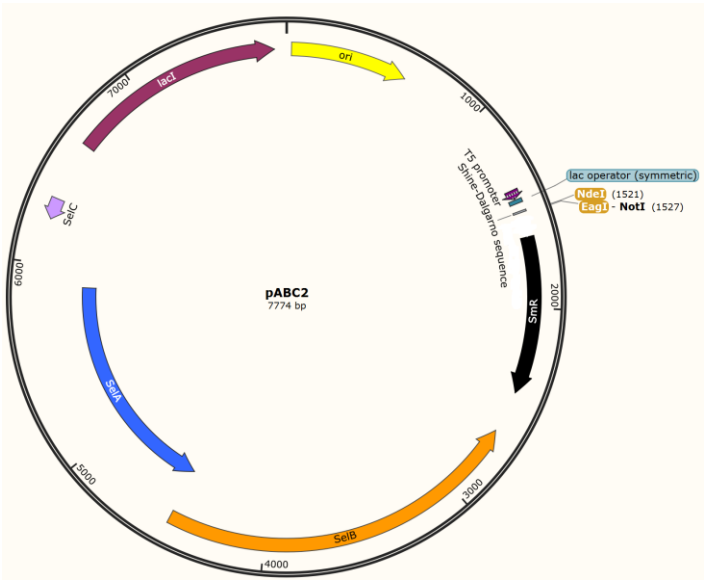


Figure M5. pABC2 selenoproteins expression vector. Created by Addgene application.

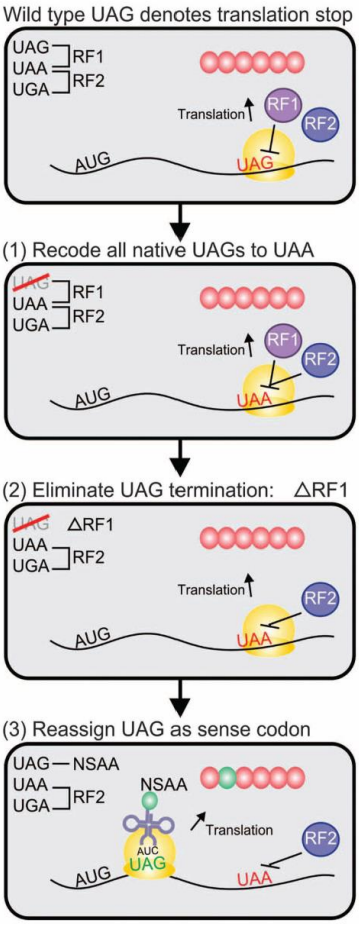


Figure M6. Engineering a genomically recoded organism (GRO) with a reassigned UAG codon. Wild-type E.coliMG1655 has 321 known UAG codons that are decoded as translation stops by RF1 (for UAG and UAA).

(1) All known UAG codons replaced by UAA, relieving dependence on RF1 for termination.

(2) Abolished UAG translational termination by deleting RF1, creating a blank codon.

(3) Expand the genetic code: introduced an orthogonal* aminoacyl–tRNA synthetase (aaRS) /tRNA pairs to reassign UAG as a dedicated sense codon capable of incorporating nonstandard amino acids (NSAAs) with new chemical properties (Lajoie et al., 2013).

*Orthogonality is the ability of a tRNA and its corresponding aaRS to interact exclusively with each other and avoid cross-reactions with additional types of tRNAs and aaRSs in a given organism.

3. Expression and Purification of bacterial Selenoprotein N

3.1. Bacterial Selenoprotein N expression

Culture and Expression: A pre-culture of 5ml 2YT medium containing Ampicillin 100µg /ml and Streptomycin 50µg/ml was inoculated with one selected colony and incubated overnight at 30°C to reach an OD_{600} of 3 to 5 units in the stationary phase. The next day, the culture is seeded with OD_{600} 0.1, by dilution of the pre-culture with the desired volume of 2YT and appropriate antibiotics, supplemented with 2% glucose to inhibit protein production. The culture is incubated at 25°C for 4 to 5 hours until the OD_{600} reaches 0.8 to 1. Glucose is removed by centrifugation at 4000 rpm for 15min at 4°C and a fresh 2YT medium containing antibiotics is added. Protein expression is induced with 0.5mM IPTG and the growth medium is incubated at 20°C for 16h. For SelenoN-Sec expression, the culture is supplemented with 5µM selenite before and at the time of induction. The next day, the culture is harvested and centrifuged at 4000 rpm for 15min at 4°C and the pellet is collected and stored at -20°C for later use.

Sample preparation: Harvested cells pellet is resuspended with 1x lysis buffer (5ml lysis buffer/mg of pellet size). To break open the cells, physical disruption by sonication was performed in a short burst of 10sec 4-5 times (micro-sonicator for small lysate), 30sec 4-5 times (medium-sonicator 120v ≈ 60Hz), with a 1min pause. The lysate is kept on ice throughout the process to avoid heat denaturation of the protein. Sonication frequency is adjusted according to culture size. After cells disruption, 50µl of the lysate was kept as a total protein extract (TE) and the rest of the crude suspension was centrifuged at 12000 rpm for 30min at 4°C, 50µl of the supernatant is kept as a soluble protein extract (SE), and the rest is kept at 4°C to be used later in purification as soon as possible. SDS-PAGE is prepared, usually as 10% gel or upon the desired percentage according to the size of the protein of interest. The expression of the recombinant SelenoN was verified by SDS-PAGE and western blot using the anti-6xHis antibodies, and anti-strep tag antibodies for Rog3500. All compositions of lysis buffer, western blot solutions, and SDS-PAGE preparation are available in (Tables 5, 6, and 8) in the materials section.

3.2. Purification of the recombinant bacterial Selenoprotein N by Immobilized Metal Affinity Chromatography (IMAC)

Native purification: Sample preparation as mentioned before. His60 Ni Superflow Resin Imino-diacetic acid (Ni-IDA) is used for 6X-His tagged protein purification. According to culture size, a certain amount of Ni-IDA slurry resin solution is loaded on the column. The column is equilibrated with 20-bed volumes of lysis buffer to remove ethanol from the slurry solution. The soluble extract was batched with the beads overnight at 4°C with rolling. Then, the flow through was collected by gravity flow. The batched protein retained on the column is washed with 20-bed volumes of the wash buffer. Followed by a chaperone removal wash, applied twice with 5mM ATP-$MgCl_2$ added to the lysis buffer (30 min at 37°C, and 15 min at 37°C/ with rolling). Two more washes with wash buffer are applied to remove excess ATP from the matrix. Then, the protein is eluted with 10-bed volumes of elution buffer. All purification buffer preparations are available in (Table 7) in the materials section.

Proteins Fractionation on SDS-PAGE:

Protein loads are prepared for total, soluble, and flow-through fractions by mixing 2.5 µl sample + 6.5 µl milli-Q + 3µl 4X protein loading dye (supplied with 20µg of DTT). The sample is then preheated at 95°C for 5min and loaded on an SDS-Gel electrophoresis system 0.75mm thick glassware. The gel is run at 100 V. Next, the gel is rinsed a couple of times with tap water and stained with Coomassie blue for 1-2h, then de-stained with 10% acetic acid till the gel is clear for imaging by GelDoc imaging system. For wash and elution fractions, protein load is prepared by mixing 5µl sample + 5µl milli-Q + 3 µl protein loading dye.

The protein concentration is determined by measuring absorption at 280nm by spectrometric reading. For purified protein BCA or Bradford protocol is used to measure the concentration of the protein. The following website is used to find the pI, Extinction coefficient of a protein of interest https://www.expasy.org/resources/protparam.

3.3. Peptide Antibodies design

Antibodies directed against the bacterial SelenoN fused to the thioredoxin-like domain identified by the BLAST search tool were already available in the lab. However, since the sequences obtained from the high-throughput sequencing from metagenomic DNA are not fully reliable, we have prepared new antibodies targeting the new amplified SelenoN3, corresponding to the first natural bacterial SelenoN sequence. Moreover, since the *SELENON* gene exists in multicopy number in *Poribacteria*, new peptide antibodies targeting the catalytic site and the most conservative 20 amino acids upstream and downstream were designed based on the alignment and consensus amino acids sequence of 158 SelenoN copies. The purpose of these new anti-peptide antibodies is to detect all SelenoN proteins expressed from multiple *Poribacteria* strains. However, these antibodies were recently developed and they are still under validation. A logoplot of the consensus of the conserved amino acid residues of the 158 aligned SelenoNs was developed using https://weblogo.berkeley.edu/logo.cgi and is shown in (Figure M7).

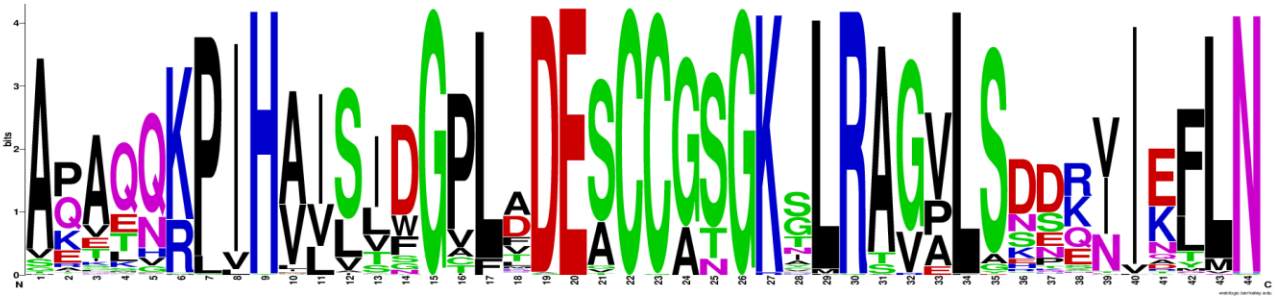


Figure M7. Sequence consensus of 158 alignments of poribacterial SelenoN catalytic site.

4. Fluorescence *in situ* hybridization (FISH)

Sponge sample collection and processing were explained previously in the sample collection section. The abundance of *Poribacteria* in the *A. aerophoba* sponge sample, in normal and stress conditions, was examined using 16s rRNA oligonucleotide HRP-labeled probes specific to *Poribacteria* in Catalyzed Reporter Deposition Fluorescence *in situ* Hybridization (CARD-FISH). We also adapted the method of helper probes (unlabeled oligonucleotides) from (Fuchs et al., 2000)**.** The helper probes method aid in the accessibility of oligonucleotide probes in FISH, and strengthens the hybridization signals. The probes were designed from the adjacent nucleotidic residues of the target site of the HRP-labeled probes. All CARD-FISH preparations and probes sequence are available in the materials section under the CARD-FISH subsection (Table 15-20). Step by step FISH protocol is available in (Figure M8)

Classical FISH, and CARD-FISH

A. Sample fixation and preparation

A. aerophoba sponge tissue was preserved in 4% PFA buffered paraformaldehyde in 1x PBS pH 7.5 and kept at 4°C till use. The objective of tissue fixation is to preserve the natural morphological structure and biochemical characterization. After fixation series of washes and dehydration were performed to minimize tissue distortion before proceeding with embedding in paraffin molds. Once the molds were dried, sectioning of tissues by Microtome (MICROM HM 340E Thermo-SCIENTIFIC) of 10-12µm thickness was made. The sections were placed in a water bath of 40°C supplemented with $1/10^{th}$ volume Tissue Section Adhesive (Thermo-Scientific REF 86014) to relax for a few seconds then pasted on slides. Next, Slides air dried for 1h then placed into a 37°C chamber for at least 3 hours, then stored in a dust-moist-free box. The deparaffinization step is carried out right before proceeding with CARD-FISH procedures.

B. Treatment of tissue sections (for CARD-FISH)

Inactivation of endogenous peroxidases: The CARD reaction is initiated by the oxidation of H_2O_2 by peroxidase activity; therefore endogenous peroxidases have to be inactivated prior to the CARD reaction to eliminate false-positive signals and reduce background. therefore, slides were incubated in 0.1 M HCl for 2min at room temperature, followed by incubation in 20mM Tris.HCL for 10min at room temperature.

Permeabilization: HRP is a large molecule of molecular weight of approximately 40 kDa. HRP-labeled probes penetrate cells less efficiently than fluorescent-labeled probes (molecular weights of fluorophore range 500–1,000 Da); therefore, pre-treatment of tissues to improve the accessibility of HRP-labeled probes is important for CARD-FISH. Below we adapted the optimized permeabilization protocol from (Croué et al., 2013).

Permeabilization step 1: tissue sections covered up with proteinase K (0.5µg/ml) and incubated for 5min in a humid chamber at 37°C, followed by a wash in 20mM Tris.HCL for 10min then air dried.

Permeabilization step 2: tissue sections covered up with lysozyme (10mg/ml) in lysozyme buffer and incubated for 1h at 37°C. The two permeabilization steps are separated by a quick dip of slides in 0.1% low melting point agarose to avoid detachment of the tissues of the slides. These permeabilization steps are important to facilitate the penetration of the HRP probes inside the cells due to their large size. After hybridization, a quick wash in milli-Q followed by dehydration in 96% ethanol then air-dried before the hybridization step.

C. Quantification of HRP-Labelled Probes

HRP is absorbed at 260nm and 404nm, while DNA and RNA are absorbed at 260nm, therefore, we need to make a correction of the HRP concentration to obtain the exact concentration of the oligonucleotide probes. Purified probe stocks were delivered lyophilized. To determine the exact concentration the probes were resuspended in 50 µl sterilized milli-Q water. A preparation of $1/10^{th}$ dilution of each probe stock was measured at 260nm by (DeNovix spectrophotometer/Fluorometer DS-11 FX+), see (Table M3) for absorption values and

calculations. Then probe working aliquots of 50 ng/µl of 125 µl volume were prepared and stored at −20°C, Once thawed stored at 4°C and are valid for one month. However, the helper probes (ssDNA) were resuspended into 100 µl sterilized milli-Q water.

Table M3. HRP probes absorbance calculations and helper probes.

Probes	OD260	OD404	Concentration ng/µl
POR600-HRP	1.4276	0.5778	418.44
POR1130-HRP	1.1002	0.4060	326.04
EUB338-HRP	2.28	1.38	626.7
NON338-HRP	2.45	1.44	677.3
582-H	166.68	-----	1667
619-H	169.26	-----	1693
1112-H	75.49	-----	755
1131-H	62.63	-----	626.3

- Correction equation for HRP conjugated probes OD260 = [measured OD260- (measured OD404 x cf)] x 33 x L x df (Protocols for Nucleic Acid Analysis by Nonradioactive Probes edited by Elena Hilario, John F. MacKay page 158 book)
- Correction equation for ssDNA without HRP probes = measured OD260 x df
  - HRP= horse radish peroxidase, H= helper, CY3= cyanidine, L = path length 1cm, df= dilution factor is 10, abs = absorbance, cf= correction factor = 0.276.

D. FISH Hybridization

Formamide lowers the melting temperature of hybrids and therefore increases the stringency of the probe binding to the target and minimizes nonspecific hybridization. To test the stringency of the hybridization, probes were diluted to a final concentration of 2.5ng/µl in a hybridization buffer of 10% and 30% formamide concentration. Tissue sections were covered up with hybridization buffer-probe mix, and slides were placed in a humid chamber (the bottom is wet with a 10%/30% formamide-water solution) and incubated at 46°C for overnight*. Slides were placed in a wash buffer preheated to 48°C two degrees higher than the hybridization temperature and incubated for 15 min. Slides were then equilibrated in 1X PBS buffer for 15 min at room temp. Excess PBS was removed by tapping the slides quickly on a tissue without fully drying, otherwise, the background fluorescence will increase., In

classical FISH, the slides were washed in milli-Q for 1 min, then completely air dried before counterstaining and mounting. For CARD-FISH we continued to CARD-FISH reaction right after the PBS equilibration.

*HRP was found to be active and thermally stable above 37°C in CARD-FISH, with optimization of hybridization stringency and incubation time (Hoshino et al., 2008; Tujula et al., 2006).

E. CARD-FISH reaction

CARD substrate preparation is shown in (Table 19). Slides were covered up with CARD-substrate and incubated in a water-humid chamber box at 37°C for up to 20 min in the dark. After the CARD reaction, the excess CARD substrate is tapped gently onto a piece of tissue and the slides were placed in 1X PBS at room temperature for 15 min. Slides were then washed in milli-Q for 1 min, then completely air dried before counterstaining and mounting.

F. Staining for microscopy observation

Slides were counterstained and mounted by adding a drop of the counterstaining mounting medium mixed DAPI solution (Citifluor:Vectashield, 4:1 ratio with DAPI [2.5 µg/ml]), quickly with cautious lay the coverslip with a slight inclination over the slides.

G. Imaging

Imaging with Widefield UPRIGHT FLUORESCENCE microscope (OLYMPUS-PROVIS) equipped with DAPI and FITC filter sets and associated with CMOS Camera. Plankton diversity, microbial community counting and observation – DP72 : CCD 12.8 Mega pixels (4140 x 3096) color 12 bit – Qimaging Exi Blue : CCD (1392 x 1040) monochrome 14 bit Binning 8x8.

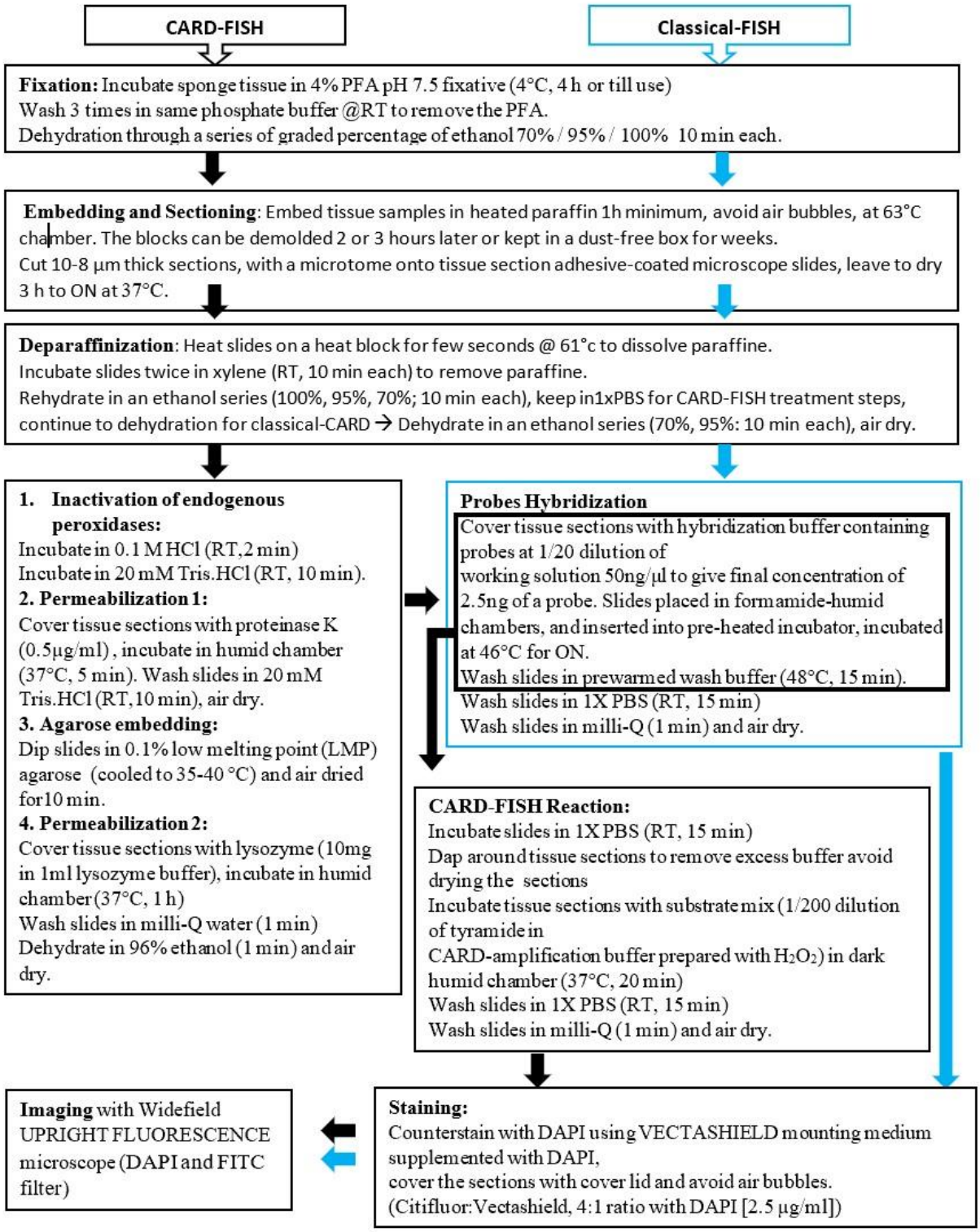


Figure M8. Classical FISH, and CARD-FISH Protocol.

H. Degenerated oligonucleotide probes of SCUG and SECIS Sequence.

A consensus sequence of the catalytic site SCUG in addition to the downstream SECIS from 166 *SELENON* genes was created from a BLAST search tool (Figure M9). This consensus sequence was used to design three degenerated probes that will be used in CARD-FISH to examine the expression level of all SelenoN mRNA in *A. aerophoba* tissue. To test the efficiency of the probes. they were used in a northern blot on an RNA extraction from *A. aerophoba* tissue.

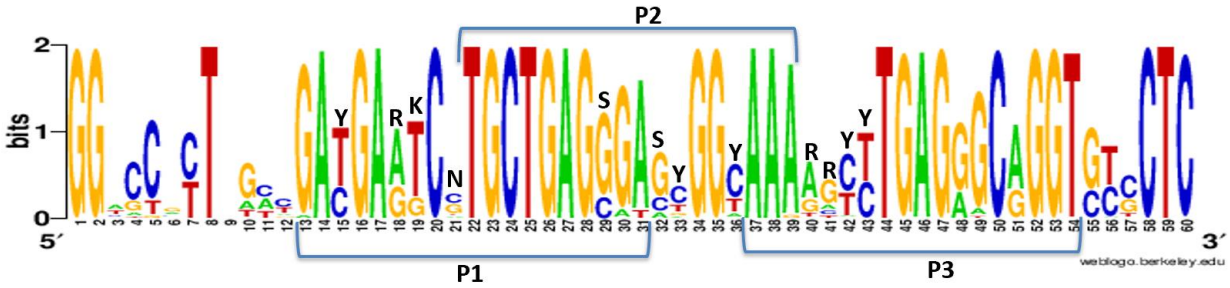


Figure M9. Sequence consensus of SCUG and SECIS in 166 *Poribacteria SELENON*.

5. Electron Microscope Transmission electron microscopy & Immunogold-labeling

The morphology and distribution of *A. aerophoba* symbionts were examined using transmission electron microscopy (TEM). From a previous study on *Poribacteria* morphology by (Jahn et al., 2016), we could anticipate the localization of *Poribacteria* within the sponge mesohyl. *In* addition, the expression level of SelenoN within the sponge tissue sections was examined using lab-made antibody developed against one form of bacterial selenoN to use in Immunogold-labeling.

A) Sample preparation for Transmission Electron Microscope TEM:

Samples were preserved in glutaraldehyde and kept at 4°C with frequent changes of the fixative (every two days if possible) till further use. Glutaraldehyde is efficient in preserving the outer membrane and ultracellular structures. Then, a series of washes with sodium cacodylate buffer were applied at room temperature, followed by Osmium fixation and a last wash with sodium cacodylate buffer at room temperature. After that dehydration in a series of ethanol preceded the impregnation in a series of concentrations with the impregnation medium. If the experiment will be done a few days later, keep the sample at 4°C and change the medium every two days if possible. After successful impregnation, The sectioning of the samples was done by an ultramicrotome Leica Ultracut R, and grids of tissue sections were prepared. Then the grids were stained with Uranyl acetate aqueous solution and lead citrate aqueous solution. Step by step protocol is depicted in (Figure M10)

All electron microscope solution preparations are available in the materials section under the Electron Microscope subsection (Table 21-22).

B) Sample preparation for Immunohistochemistry (IHC) EM.

Samples were preserved in a mixed fixative of (3% paraformaldehyde/ 0.5% glutaraldehyde)* and kept at 4°C with frequent changes of the fixative (every two days if possible) until further use. Same washes and dehydration steps as before but without the osmium fixation. After that impregnation of the sample in LR WHITE resin. The sectioning of the samples was done by an ultramicrotome Leica Ultracut R, and grids of tissue sections were prepared.

Immunogold-labeling.

To avoid non-specific binding to the antibodies, the grids were incubated with glycine to inactivate free aldehyde groups. Then the grids were covered with AURION BLOCKING SOLUTION followed by incubation in an incubation buffer. After blocking, grids were incubated at 4°C overnight with primary antibodies diluted in an incubation buffer. Washes with incubation buffer were performed, followed by incubation with secondary antibodies gold of two concentrations (GAR 10nm/GAR 25nm). Several washes were performed before the final step. Staining with Uranyl acetate. Step by step protocol is depicted in (Figure M10)

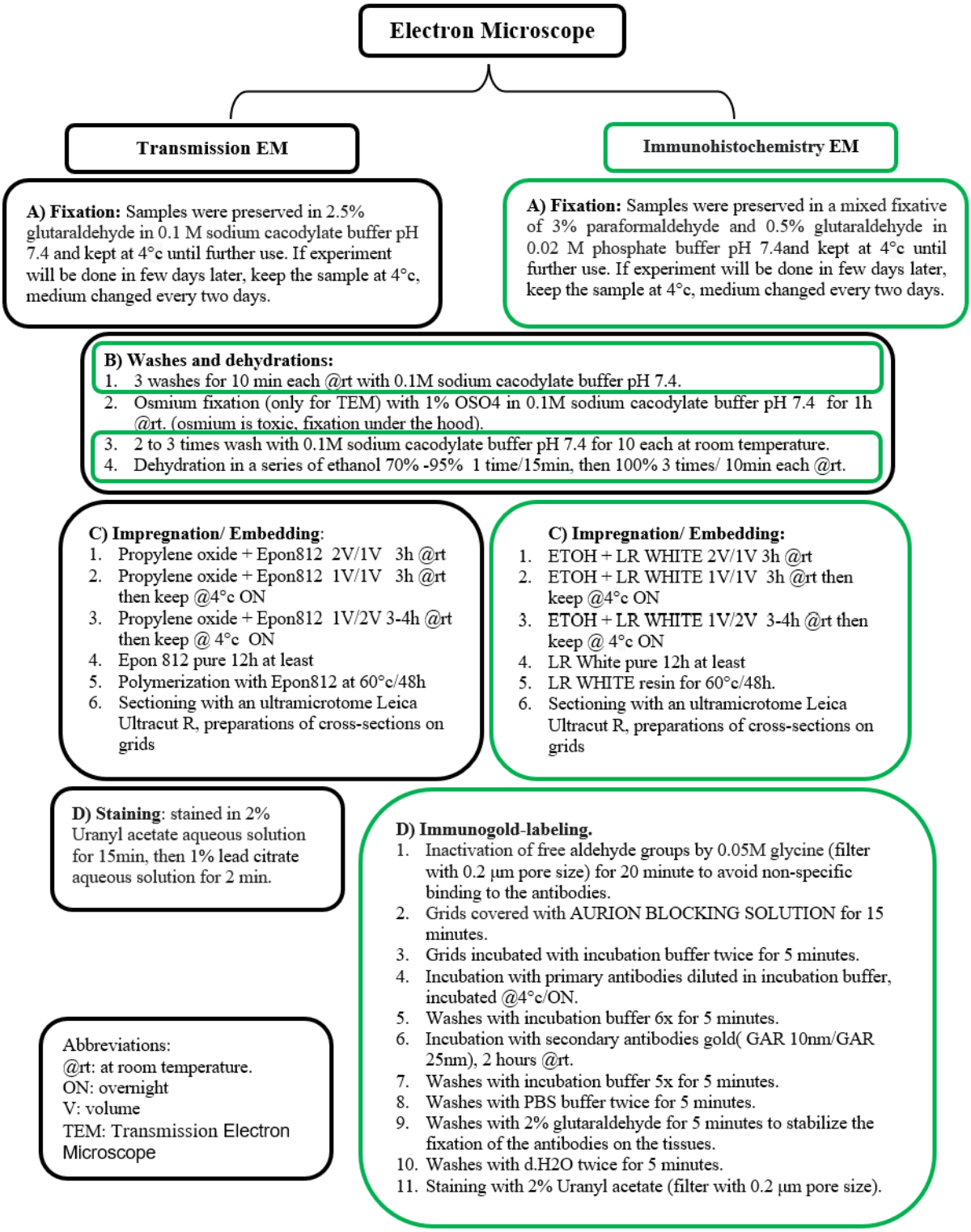


Figure M10. TEM and IHC protocol.

BIBLIOGRAPHY

A

Aachmann, F. L., Fomenko, D. E., Soragni, A., Gladyshev, V. N., & Dikiy, A. (2007). Solution Structure of Selenoprotein W and NMR Analysis of Its Interaction with 14-3-3 Proteins *. *Journal of Biological Chemistry*, *282*(51), 37036–37044. https://doi.org/10.1074/jbc.M705410200

Abid, H., Cartier, D., Hamieh, F. A., françois-bellan, anne-marie, Bucharles, C., Pothion, H., Manecka, D.-L., Leprince, J., Adriouch, S., Boyer, O., Anouar, Y., & Lihrmann, I. (2019). AMPK Activation of PGC-1α/NRF-1-Dependent SELENOT Gene Transcription Promotes PACAP-Induced Neuroendocrine Cell Differentiation Through Tolerance to Oxidative Stress. *Molecular Neurobiology*, *56*. https://doi.org/10.1007/s12035-018-1352-x

Afanasieva, E., Chaudhuri, I., Martin, J., Hertle, E., Ursinus, A., Alva, V., Hartmann, M. D., & Lupas, A. N. (2019). Structural diversity of oligomeric β-propellers with different numbers of identical blades. *ELife*, *8*, e49853. https://doi.org/10.7554/eLife.49853

Allamand, V., Richard, P., Lescure, A., Ledeuil, C., Desjardin, D., Petit, N., Gartioux, C., Ferreiro, A., Krol, A., Pellegrini, N., Urtizberea, J. A., & Guicheney, P. (2006). A single homozygous point mutation in a 3′untranslated region motif of selenoprotein N mRNA causes SEPN1-related myopathy. *EMBO Reports*, *7*(4), 450–454. https://doi.org/10.1038/sj.embor.7400648

Allmang, C., & Krol, A. (2006). Selenoprotein synthesis: UGA does not end the story. *Biochimie*, *88*(11), 1561–1571. https://doi.org/10.1016/j.biochi.2006.04.015

Allmang, C., Carbon, P., & Krol, A. (2002). The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: Identification of SBP2 amino acids important to SECIS RNA binding. *RNA*, *8*(10), 1308–1318. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1370339/

Allmang, C., Wurth, L., & Krol, A. (2009). The selenium to selenoprotein pathway in eukaryotes: More molecular partners than anticipated. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1790*(11), 1415–1423. https://doi.org/10.1016/j.bbagen.2009.03.003

Almanza, A., Carlesso, A., Chintha, C., Creedican, S., Doultsinos, D., Leuzzi, B., Luís, A., McCarthy, N., Montibeller, L., More, S., Papaioannou, A., Püschel, F., Sassano, M. L., Skoko, J., Agostinis, P.,

de Belleroche, J., Eriksson, L. A., Fulda, S., Gorman, A. M., Samali, A. (2019). Endoplasmic reticulum stress signalling – from basic mechanisms to clinical applications. *The FEBS Journal*, *286*(2), 241–278. https://doi.org/10.1111/febs.14608

Alpdagtas, S., & Binay, B. (2021). Nadp+-dependent formate dehydrogenase: A review. *Biocatalysis and Biotransformation*, *39*(4), 260–268. https://doi.org/10.1080/10242422.2020.1865933

Alquezar-Planas, D. E., Löber, U., Cui, P., Quedenau, C., Chen, W., & Greenwood, A. D. (2021). DNA sonication inverse PCR for genome scale analysis of uncharacterized flanking sequences. *Methods in Ecology and Evolution*, *12*(1), 182–195. https://doi.org/10.1111/2041-210X.13497

Alvarez, B., Frings, P. J., Clymans, W., Fontorbe, G., & Conley, D. J. (2017). Assessing the Potential of Sponges (Porifera) as Indicators of Ocean Dissolved Si Concentrations. *Frontiers in Marine Science*, *4*. https://www.frontiersin.org/articles/10.3389/fmars.2017.00373

Ambudkar, I. S., de Souza, L. B., & Ong, H. L. (2017). TRPC1, Orai1, and STIM1 in SOCE: Friends in tight spaces. *Cell Calcium*, *63*, 33–39. https://doi.org/10.1016/j.ceca.2016.12.009

An, B. C., Choi, Y.-D., Oh, I.-J., Kim, J. H., Park, J.-I., & Lee, S. (2018). GPx3-mediated redox signaling arrests the cell cycle and acts as a tumor suppressor in lung cancer cell lines. *PLOS ONE*, *13*(9), e0204170. https://doi.org/10.1371/journal.pone.0204170

Andersson, D. C., Betzenhauser, M. J., Reiken, S., Meli, A. C., Umanskaya, A., Xie, W., Shiomi, T., Zalk, R., Lacampagne, A., & Marks, A. R. (2011). Ryanodine Receptor Oxidation Causes Intracellular Calcium Leak and Muscle Weakness in Aging. *Cell Metabolism*, *14*(2), 196–207. https://doi.org/10.1016/j.cmet.2011.05.014

Arbogast, S., & Ferreiro, A. (2010). Selenoproteins and Protection against Oxidative Stress: Selenoprotein N as a Novel Player at the Crossroads of Redox Signaling and Calcium Homeostasis. *Antioxidants & Redox Signaling*, *12*(7), 893–904. https://doi.org/10.1089/ars.2009.2890

Arbogast, S., Beuvin, M., Fraysse, B., Zhou, H., Muntoni, F., & Ferreiro, A. (2009). Oxidative stress in SEPN1-related myopathy: From pathophysiology to treatment. *Annals of Neurology*, *65*(6), 677–686. https://doi.org/10.1002/ana.21644

Arnér, E. S. J. (2010). Selenoproteins—What unique properties can arise with selenocysteine in place of cysteine? *Experimental Cell Research*, *316*(8), 1296–1303. https://doi.org/10.1016/j.yexcr.2010.02.032

Arnér, E. S. J., & Holmgren, A. (2000). Physiological functions of thioredoxin and thioredoxin reductase. *European Journal of Biochemistry*, *267*(20), 6102–6109. https://doi.org/10.1046/j.1432-1327.2000.01701.x

Arnér, E. S. J., Sarioglu, H., Lottspeich, F., Holmgren, A., & Böck, A. (1999). High-level expression in Escherichia coli of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the selA , selB and selC genes 1 1Edited by M. Gottesman. *Journal of Molecular Biology*, *292*(5), 1003–1016. https://doi.org/10.1006/jmbi.1999.3085

Arthur, J. R. (2001). The glutathione peroxidases. *Cellular and Molecular Life Sciences CMLS*, *57*(13), 1825–1835. https://doi.org/10.1007/PL00000664

B

Baclaocos, J., Santesmasses, D., Mariotti, M., Bierła, K., Vetick, M. B., Lynch, S., McAllen, R., Mackrill, J. J., Loughran, G., Guigó, R., Szpunar, J., Copeland, P. R., Gladyshev, V. N., & Atkins, J. F. (2019). Processive recoding and metazoan evolution of Selenoprotein P: Up to 132 UGAs in molluscs. *Journal of Molecular Biology*, *431*(22), 4381–4407. https://doi.org/10.1016/j.jmb.2019.08.007

Bayer, K., Schmitt, S., & Hentschel, U. (2008). Physiology, phylogeny and in situ evidence for bacterial and archaeal nitrifiers in the marine sponge Aplysina aerophoba. *Environmental Microbiology*, *10*(11), 2942–2955. https://doi.org/10.1111/j.1462-2920.2008.01582.x

Bibliography

Berry, M. J., Banu, L., Chen, Y., Mandel, S. J., Kieffer, J. D., Harney, J. W., & Larsen, P. R. (1991). Recognition of UGA as a selenocysteine codon in Type I deiodinase requires sequences in the 3′ untranslated region. *Nature*, *353*(6341), Article 6341. https://doi.org/10.1038/353273a0

Betzenhauser, M. J., & Marks, A. R. (2010). Ryanodine receptor channelopathies. *Pflugers Archiv : European Journal of Physiology*, *460*(2), 467–480. https://doi.org/10.1007/s00424-010-0794-4

Bianco, A. C., Salvatore, D., Gereben, B., Berry, M. J., & Larsen, P. R. (2002). Biochemistry, Cellular and Molecular Biology, and Physiological Roles of the Iodothyronine Selenodeiodinases. *Endocrine Reviews*, *23*(1), 38–89. https://doi.org/10.1210/edrv.23.1.0455

Björnstedt, M., Xue, J., Huang, W., Akesson, B., & Holmgren, A. (1994). The thioredoxin and glutaredoxin systems are efficient electron donors to human plasma glutathione peroxidase. *The Journal of Biological Chemistry*, *269*(47), 29382–29384. https://doi.org/10.1016/S0021-9258(18)43889-6

Boukhzar, L., Hamieh, A., Cartier, D., Tanguy, Y., Alsharif, I., Castex, M., Arabo, A., Hajji, S. E., Bonnet, J.-J., Errami, M., Falluel-Morel, A., Chagraoui, A., Lihrmann, I., & Anouar, Y. (2016). Selenoprotein T Exerts an Essential Oxidoreductase Activity That Protects Dopaminergic Neurons in Mouse Models of Parkinson's Disease. *Antioxidants & Redox Signaling*, *24*(11), 557–574. https://doi.org/10.1089/ars.2015.6478

Brigelius-Flohé, R., Müller, M., Lippmann, D., & Kipp, A. P. (2012). The Yin and Yang of Nrf2-Regulated Selenoproteins in Carcinogenesis. *International Journal of Cell Biology*, *2012*, 486147. https://doi.org/10.1155/2012/486147

Brinkmann, C. M., Marker, A., & Kurtböke, D. İ. (2017). An Overview on Marine Sponge-Symbiotic Bacteria as Unexhausted Sources for Natural Product Discovery. *Diversity*, *9*(4), Article 4. https://doi.org/10.3390/d9040040

Budiman, M. E., Bubenik, J. L., Miniard, A. C., Middleton, L. M., Gerber, C. A., Cash, A., & Driscoll, D. M. (2009). Eukaryotic Initiation Factor 4a3 is a Selenium-regulated RNA-binding Protein that

Selectively Inhibits Selenocysteine Incorporation. *Molecular Cell*, *35*(4), 479–489. https://doi.org/10.1016/j.molcel.2009.06.026

Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., & Zemskaya, T. I. (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Scientific Data*, *6*(1), Article 1. https://doi.org/10.1038/sdata.2019.7

Burk, R. F., & Hill, K. E. (2005). SELENOPROTEIN P: An Extracellular Protein with Unique Physical Characteristics and a Role in Selenium Homeostasis. *Annual Review of Nutrition*, *25*(1), 215–235. https://doi.org/10.1146/annurev.nutr.24.012003.132120

Burk, R. F., & Hill, K. E. (2010). 4.13—Glutathione Peroxidases. In C. A. McQueen (Ed.), *Comprehensive Toxicology (Second Edition)* (pp. 229–242). Elsevier. https://doi.org/10.1016/B978-0-08-046884-6.00413-9

C

Carrier, T. J., Maldonado, M., Schmittmann, L., Pita, L., Bosch, T. C. G., & Hentschel, U. (2022). Symbiont transmission in marine sponges: Reproduction, development, and metamorphosis. *BMC Biology*, *20*(1), Article 1. https://doi.org/10.1186/s12915-022-01291-6

Castets, P., Bertrand, A. T., Beuvin, M., Ferry, A., Le Grand, F., Castets, M., Chazot, G., Rederstorff, M., Krol, A., Lescure, A., Romero, N. B., Guicheney, P., & Allamand, V. (2011). Satellite cell loss and impaired muscle regeneration in selenoprotein N deficiency. *Human Molecular Genetics*, *20*(4), 694–704. https://doi.org/10.1093/hmg/ddq515

Castets, P., Lescure, A., Guicheney, P., & Allamand, V. (2012). Selenoprotein N in skeletal muscle: From diseases to function. *Journal of Molecular Medicine*, *90*(10), 1095–1107. https://doi.org/10.1007/s00109-012-0896-x

Castets, P., Maugenre, S., Gartioux, C., Rederstorff, M., Krol, A., Lescure, A., Tajbakhsh, S., Allamand, V., & Guicheney, P. (2009). Selenoprotein N is dynamically expressed during mouse

development and detected early in muscle precursors. *BMC Developmental Biology*, *9*, 46. https://doi.org/10.1186/1471-213X-9-46

Castex, M., Arabo, A., Benard, M., Roy, V., Le Joncour, V., Prévost, G., Bonnet, J.-J., Anouar, Y., & Falluel-Morel, A. (2016). Selenoprotein T Deficiency Leads to Neurodevelopmental Abnormalities and Hyperactive Behavior in Mice. *Molecular Neurobiology*, *53*. https://doi.org/10.1007/s12035-015-9505-7

Chapple, C. E., & Guigó, R. (2008). Relaxation of Selective Constraints Causes Independent Selenoprotein Extinction in Insect Genomes. *PLOS ONE*, *3*(8), e2968. https://doi.org/10.1371/journal.pone.0002968

Chavatte, L., Brown, B. A., & Driscoll, D. M. (2005). Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nature Structural & Molecular Biology*, *12*(5), Article 5. https://doi.org/10.1038/nsmb922

Chen, P., Wang, R.-R., Ma, X.-J., Liu, Q., & Ni, J.-Z. (2013). Different Forms of Selenoprotein M Differentially Affect Aβ Aggregation and ROS Generation. *International Journal of Molecular Sciences*, *14*(3), 4385–4399. https://doi.org/10.3390/ijms14034385

Cheng, Q., & Arnér, E. S. J. (2017). Selenocysteine Insertion at a Predefined UAG Codon in a Release Factor 1 (RF1)-depleted Escherichia coli Host Strain Bypasses Species Barriers in Recombinant Selenoprotein Translation. *Journal of Biological Chemistry*, *292*(13), 5476–5487. https://doi.org/10.1074/jbc.M117.776310

Chernorudskiy, A., Varone, E., Colombo, S. F., Fumagalli, S., Cagnotto, A., Cattaneo, A., Briens, M., Baltzinger, M., Kuhn, L., Bachi, A., Berardi, A., Salmona, M., Musco, G., Borgese, N., Lescure, A., & Zito, E. (2020). Selenoprotein N is an endoplasmic reticulum calcium sensor that links luminal calcium levels to a redox activity. *Proceedings of the National Academy of Sciences*, *117*(35), 21288–21298. https://doi.org/10.1073/pnas.2003847117

Clark, D. P., & Pazdernik, N. J. (2016). DNA Synthesis In Vivo and In Vitro. In *Biotechnology* (pp. 97–130). Elsevier. https://doi.org/10.1016/B978-0-12-385015-7.00004-1

Clarke, N. F., Kidson, W., Quijano-Roy, S., Estournet, B., Ferreiro, A., Guicheney, P., Manson, J. I., Kornberg, A. J., Shield, L. K., & North, K. N. (2006). SEPN1: Associated with congenital fiber-type disproportion and insulin resistance. *Annals of Neurology*, *59*(3), 546–552. https://doi.org/10.1002/ana.20761

Croué, J., West, N. J., Escande, M.-L., Intertaglia, L., Lebaron, P., & Suzuki, M. T. (2013). A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge Crambe crambe. *Scientific Reports*, *3*, 2583. https://doi.org/10.1038/srep02583

Cyr, D. M., & Hebert, D. N. (2009). Protein quality control—Linking the unfolded protein response to disease. *EMBO Reports*, *10*(11), 1206–1210. https://doi.org/10.1038/embor.2009.224

D

Dat, T. T. H., Steinert, G., Cuc, N. T. K., Smidt, H., & Sipkema, D. (2021). Bacteria Cultivated From Sponges and Bacteria Not Yet Cultivated From Sponges—A Review. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.737925

De Jesus, L. A., Hoffmann, P. R., Michaud, T., Forry, E. P., Small-Howard, A., Stillwell, R. J., Morozova, N., Harney, J. W., & Berry, M. J. (2006). Nuclear Assembly of UGA Decoding Complexes on Selenoprotein mRNAs: A Mechanism for Eluding Nonsense-Mediated Decay? *Molecular and Cellular Biology*, *26*(5), 1795–1805. https://doi.org/10.1128/MCB.26.5.1795-1805.2006

Deniziak, M., Thisse, C., Rederstorff, M., Hindelang, C., Thisse, B., & Lescure, A. (2007). Loss of selenoprotein N function causes disruption of muscle architecture in the zebrafish embryo. *Experimental Cell Research*, *313*(1), 156–167. https://doi.org/10.1016/j.yexcr.2006.10.005

Dentice, M., Ambrosio, R., Damiano, V., Sibilio, A., Luongo, C., Guardiola, O., Yennek, S., Zordan, P., Minchiotti, G., Colao, A., Marsili, A., Brunelli, S., Del Vecchio, L., Larsen, P. R., Tajbakhsh, S., & Salvatore, D. (2014). Intracellular Inactivation of Thyroid Hormone Is a Survival Mechanism

for Muscle Stem Cell Proliferation and Lineage Progression. *Cell Metabolism*, *20*(6), 1038. https://doi.org/10.1016/j.cmet.2014.10.009

Dentice, M., Marsili, A., Ambrosio, R., Guardiola, O., Sibilio, A., Paik, J.-H., Minchiotti, G., DePinho, R. A., Fenzi, G., Larsen, P. R., & Salvatore, D. (2010). The FoxO3/type 2 deiodinase pathway is required for normal mouse myogenesis and muscle regeneration. *The Journal of Clinical Investigation*, *120*(11), 4021–4030. https://doi.org/10.1172/JCI43670

Dentice, M., Marsili, A., Zavacki, A., Larsen, P. R., & Salvatore, D. (2013). The deiodinases and the control of intracellular thyroid hormone signaling during cellular differentiation. *Biochimica et Biophysica Acta*, *1830*(7), 3937–3945. https://doi.org/10.1016/j.bbagen.2012.05.007

Diamond, A. M., Choi, I. S., Crain, P. F., Hashizume, T., Pomerantz, S. C., Cruz, R., Steer, C. J., Hill, K. E., Burk, R. F., & McCloskey, J. A. (1993). Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA([Ser]Sec). *Journal of Biological Chemistry*, *268*(19), 14215–14223. https://doi.org/10.1016/S0021-9258(19)85229-8

Dikiy, A., Novoselov, S. V., Fomenko, D. E., Sengupta, A., Carlson, B. A., Cerny, R. L., Ginalski, K., Grishin, N. V., Hatfield, D. L., & Gladyshev, V. N. (2007). SelT, SelW, SelH, and Rdx12: Genomics and Molecular Insights into the Functions of Selenoproteins of a Novel Thioredoxin-like Family. *Biochemistry*, *46*(23), 6871–6882. https://doi.org/10.1021/bi602462q

Du, X., Li, H., Wang, Z., Qiu, S., Liu, Q., & Ni, J. (2013). Selenoprotein P and selenoprotein M block Zn2+-mediated Aβ42 aggregation and toxicity†. *Metallomics*, *5*(7), 861–870. https://doi.org/10.1039/c3mt20282h

E

Eklund, H., Ingelman, M., Söderberg, B.-O., Uhlin, T., Nordlund, P., Nikkola, M., Sonnerstam, U., Joelson, T., & Petratos, K. (1992). Structure of oxidized bacteriophage T4 glutaredoxin (thioredoxin): Refinement of native and mutant proteins. *Journal of Molecular Biology*, *228*(2), 596–618. https://doi.org/10.1016/0022-2836(92)90844-A

Enoch, H. G., & Lester, R. L. (1972). Effects of Molybdate, Tungstate, and Selenium Compounds on Formate Dehydrogenase and Other Enzyme Systems in Escherichia coli1. *Journal of Bacteriology*, *110*(3), 1032–1040. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC247525/

F

Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P., & Krol, A. (2000). Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *The EMBO Journal*, *19*(17), 4796–4805. https://doi.org/10.1093/emboj/19.17.4796

Fan, Y., Xu, Z., Li, X., Gao, F., Guo, E., Chang, X., Wei, C., Zhang, C., Yu, Q., Que, C., Xiao, J., Yan, C., Wang, Z., Yuan, Y., & Xiong, H. (2022). Novel SEPN1 Mutations in Exon 1 Are Common in Rigid Spine With Muscular Dystrophy Type 1 in Chinese Patients. *Frontiers in Genetics*, *13*. https://www.frontiersin.org/articles/10.3389/fgene.2022.825793

Ferguson, A., Labunskyy, V., Fomenko, D., Araç, D., Chelliah, Y., Amezcua, C., Rizo, J., Gladyshev, V., & Deisenhofer, J. (2006). NMR Structures of the Selenoproteins Sep15 and SelM Reveal Redox Activity of a New Thioredoxin-like Family. *The Journal of Biological Chemistry*, *281*, 3536–3543. https://doi.org/10.1074/jbc.M511386200

Ferreiro, A., Ceuterick-de Groote, C., Marks, J. J., Goemans, N., Schreiber, G., Hanefeld, F., Fardeau, M., Martin, J.-J., Goebel, H. H., Richard, P., Guicheney, P., & Bönnemann, C. G. (2004). Desmin-related myopathy with mallory body–like inclusions is caused by mutations of the selenoprotein N gene. *Annals of Neurology*, *55*(5), 676–686. https://doi.org/10.1002/ana.20077

Bibliography

Ferreiro, A., Quijano-Roy, S., Pichereau, C., Moghadaszadeh, B., Goemans, N., Bönnemann, C., Jungbluth, H., Straub, V., Villanova, M., Leroy, J.-P., Romero, N. B., Martin, J.-J., Muntoni, F., Voit, T., Estournet, B., Richard, P., Fardeau, M., & Guicheney, P. (2002). Mutations of the Selenoprotein N Gene, Which Is Implicated in Rigid Spine Muscular Dystrophy, Cause the Classical Phenotype of Multiminicore Disease: Reassessing the Nosology of Early-Onset Myopathies. *American Journal of Human Genetics*, *71*(4), 739–749. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC378532/

Fieseler, L., Horn, M., Wagner, M., & Hentschel, U. (2004). Discovery of the Novel Candidate Phylum "Poribacteria" in Marine Sponges. *Applied and Environmental Microbiology*, *70*(6), 3724–3732. https://doi.org/10.1128/AEM.70.6.3724-3732.2004

Filipe, A., Chernorudskiy, A., Arbogast, S., Varone, E., Villar-Quiles, R.-N., Pozzer, D., Moulin, M., Fumagalli, S., Cabet, E., Dudhal, S., De Simoni, M.-G., Denis, R., Vadrot, N., Dill, C., Giovarelli, M., Szweda, L., De Palma, C., Pinton, P., Giorgi, C., … Ferreiro, A. (2021). Defective endoplasmic reticulum-mitochondria contacts and bioenergetics in SEPN1-related myopathy. *Cell Death and Differentiation*, *28*(1), 123–138. https://doi.org/10.1038/s41418-020-0587-z

Forchhammer, K., & Böck, A. (1991). Selenocysteine synthase from Escherichia coli. Analysis of the reaction sequence. *The Journal of Biological Chemistry*, *266*(10), 6324–6328. https://doi.org/10.1016/S0021-9258(18)38121-3

Franke, K. W. (1934). A New Toxicant Occurring Naturally in Certain Samples of Plant Foodstuffs: I. Results Obtained in Preliminary Feeding Trials: Eight Figures. *The Journal of Nutrition*, *8*(5), 597–608. https://doi.org/10.1093/jn/8.5.597

Fredericks, G. J., & Hoffmann, P. R. (2015). Selenoprotein K and Protein Palmitoylation. *Antioxidants & Redox Signaling*, *23*(10), 854–862. https://doi.org/10.1089/ars.2015.6375

Fredericks, G. J., Hoffmann, F. W., Hondal, R. J., Rozovsky, S., Urschitz, J., & Hoffmann, P. R. (2017). Selenoprotein K Increases Efficiency of DHHC6 Catalyzed Protein Palmitoylation by Stabilizing the Acyl-DHHC6 Intermediate. *Antioxidants*, *7*(1), 4. https://doi.org/10.3390/antiox7010004

Fuchs, B. M., Glöckner, F. O., Wulf, J., & Amann, R. (2000). Unlabeled Helper Oligonucleotides Increase the In Situ Accessibility to 16S rRNA of Fluorescently Labeled Oligonucleotide Probes. *Applied and Environmental Microbiology*, *66*(8), 3603–3607. https://doi.org/10.1128/AEM.66.8.3603-3607.2000

Fuchs, B. M., Wallner, G., Beisker, W., Schwippl, I., Ludwig, W., & Amann, R. (1998). Flow Cytometric Analysis of the In Situ Accessibility of *Escherichia coli* 16S rRNA for Fluorescently Labeled Oligonucleotide Probes. *Applied and Environmental Microbiology*, *64*(12), 4973–4982. https://doi.org/10.1128/AEM.64.12.4973-4982.1998

Fuerst, J. A., Webb, R. I., Garson, M. J., Hardy, L., & Reiswig, H. M. (1998). Membrane-bounded nucleoids in microbial symbionts of marine sponges. *FEMS Microbiology Letters*, *166*(1), 29–34. https://doi.org/10.1111/j.1574-6968.1998.tb13179.x

G

Garcin, E., Vernede, X., Hatchikian, E., Volbeda, A., Frey, M., & Fontecilla-Camps, J. (1999). The crystal structure of a reduced [NiFeSe] hydrogenase provides an image of the activated catalytic center. *Structure*, *7*(5), 557–566. https://doi.org/10.1016/S0969-2126(99)80072-0

Ge, K., Xue, A., Bai, J., & Wang, S. (1983). Keshan disease-an endemic cardiomyopathy in China. *Virchows Archiv A*, *401*(1), 1–15. https://doi.org/10.1007/BF00644785

Georgiou, G., & Valax, P. (1996). Expression of correctly folded proteins in Escherichia coli. *Current Opinion in Biotechnology*, *7*(2), 190–197. https://doi.org/10.1016/S0958-1669(96)80012-7

Gladyshev, V. N., & Kryukov, G. V. (2001). Evolution of selenocysteine-containing proteins: Significance of identification and functional characterization of selenoproteins. *BioFactors (Oxford, England)*, *14*(1–4), 87–92. https://doi.org/10.1002/biof.5520140112

Gloeckner, V., Wehrl, M., Moitinho-Silva, L., Gernert, C., Schupp, P., Pawlik, J. R., Lindquist, N. L., Erpenbeck, D., Wörheide, G., & Hentschel, U. (2014). The HMA-LMA Dichotomy Revisited:

An Electron Microscopical Survey of 56 Sponge Species. *The Biological Bulletin*, *227*(1), 78–88. https://doi.org/10.1086/BBLv227n1p78

Gobler, C. J., Lobanov, A. V., Tang, Y.-Z., Turanov, A. A., Zhang, Y., Doblin, M., Taylor, G. T., Sañudo-Wilhelmy, S. A., Grigoriev, I. V., & Gladyshev, V. N. (2013). The central role of selenium in the biochemistry and ecology of the harmful pelagophyte, Aureococcus anophagefferens. *The ISME Journal*, *7*(7), 1333–1343. https://doi.org/10.1038/ismej.2013.25

Godefroy, N., Le Goff, E., Martinand-Mari, C., Belkhir, K., Vacelet, J., & Baghdiguian, S. (2019). Sponge digestive system diversity and evolution: Filter feeding to carnivory. *Cell and Tissue Research*, *377*(3), 341–351. https://doi.org/10.1007/s00441-019-03032-8

Gong, T., Hashimoto, A. C., Sasuclark, A. R., Khadka, V. S., Gurary, A., & Pitts, M. W. (2021). Selenoprotein M Promotes Hypothalamic Leptin Signaling and Thioredoxin Antioxidant Activity. *Antioxidants & Redox Signaling*, *35*(10), 775–787. https://doi.org/10.1089/ars.2018.7594

Gromer, S., Johansson, L., Bauer, H., Arscott, L. D., Rauch, S., Ballou, D. P., Williams, C. H., Schirmer, R. H., & Arnér, E. S. J. (2003). Active sites of thioredoxin reductases: Why selenoproteins? *Proceedings of the National Academy of Sciences of the United States of America*, *100*(22), 12618–12623. https://doi.org/10.1073/pnas.2134510100

Gu, Q. P., Ream, W., & Whanger, P. D. (2002). Selenoprotein W gene regulation by selenium in L8 cells. *Biometals: An International Journal on the Role of Metal Ions in Biology, Biochemistry, and Medicine*, *15*(4), 411–420. https://doi.org/10.1023/a:1020241032336

Guimarães, M. J., Peterson, D., Vicari, A., Cocks, B. G., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Ferrick, D. A., Kastelein, R. A., Bazan, J. F., & Zlotnik, A. (1996). 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*, *93*(26), 15086–15091. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC26360/

Gursinsky, T., Jäger, J., Andreesen, J. R., & Söhling, B. (2000). A selDABC cluster for selenocysteine incorporation in Eubacterium acidaminophilum. *Archives of Microbiology*, *174*(3), 200–212. https://doi.org/10.1007/s002030000196

Gutleben, J., Loureiro, C., Ramírez Romero, L. A., Shetty, S., Wijffels, R. H., Smidt, H., & Sipkema, D. (2020). Cultivation of Bacteria From Aplysina aerophoba: Effects of Oxygen and Nutrient Gradients. *Frontiers in Microbiology*, *11*, 175. https://doi.org/10.3389/fmicb.2020.00175

H

Hamieh, A., Cartier, D., Abid, H., Calas, A., Burel, C., Bucharles, C., Jehan, C., Grumolato, L., Landry, M., Lerouge, P., Anouar, Y., & Lihrmann, I. (2017). Selenoprotein T is a novel OST subunit that regulates UPR signaling and hormone secretion. *EMBO Reports*, *18*(11), 1935–1946. https://doi.org/10.15252/embr.201643504

Hammadi, M., Oulidi, A., Gackière, F., Katsogiannou, M., Slomianny, C., Roudbaraki, M., Dewailly, E., Delcourt, P., Lepage, G., Lotteau, S., Ducreux, S., Prevarskaya, N., & Van Coppenolle, F. (2013). Modulation of ER stress and apoptosis by endoplasmic reticulum calcium leak via translocon during unfolded protein response: Involvement of GRP78. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *27*(4), 1600–1609. https://doi.org/10.1096/fj.12-218875

Hanschmann, E.-M., Godoy, J. R., Berndt, C., Hudemann, C., & Lillig, C. H. (2013). Thioredoxins, Glutaredoxins, and Peroxiredoxins—Molecular Mechanisms and Health Significance: From Cofactors to Antioxidants to Redox Signaling. *Antioxidants & Redox Signaling*, *19*(13), 1539–1605. https://doi.org/10.1089/ars.2012.4599

Hatfield, D. L., Choi, I. S., Ohama, T., Jung, J.-E., & Diamond, A. M. (1994). Selenocysteine tRNA[Ser]Sec Isoacceptors as Central Components in Selenoprotein Biosynthesis in Eukaryotes.

In R. F. Burk (Ed.), *Selenium in Biology and Human Health* (pp. 25–44). Springer. https://doi.org/10.1007/978-1-4612-2592-8_3

Hatfield, D. L., Schweizer, U., Tsuji, P. A., & Gladyshev, V. N. (Eds.). (2016). *Selenium: Its Molecular Biology and Role in Human Health*. Springer International Publishing. https://doi.org/10.1007/978-3-319-41283-2

Hatfield, D. L., Tsuji, P. A., Carlson, B. A., & Gladyshev, V. N. (2014). Selenium and selenocysteine: Roles in cancer, health and development. *Trends in Biochemical Sciences*, *39*(3), 112–120. https://doi.org/10.1016/j.tibs.2013.12.007

Hentschel, U., Piel, J., Degnan, S. M., & Taylor, M. W. (2012). Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology*, *10*(9), Article 9. https://doi.org/10.1038/nrmicro2839

Hentschel, U., Usher, K. M., & Taylor, M. W. (2006). Marine sponges as microbial fermenters. *FEMS Microbiology Ecology*, *55*(2), 167–177. https://doi.org/10.1111/j.1574-6941.2005.00046.x

Hilal, T., Killam, B. Y., Grozdanović, M., Dobosz-Bartoszek, M., Loerke, J., Bürger, J., Mielke, T., Copeland, P. R., Simonović, M., & Spahn, C. M. T. (2022). Structure of the mammalian ribosome as it decodes the selenocysteine UGA codon. *Science*, *376*(6599), 1338–1343. https://doi.org/10.1126/science.abg3875

Hoshino, T., Yilmaz, L. S., Noguera, D. R., Daims, H., & Wagner, M. (2008). Quantification of Target Molecules Needed To Detect Microorganisms by Fluorescence In Situ Hybridization (FISH) and Catalyzed Reporter Deposition-FISH. *Applied and Environmental Microbiology*, *74*(16), 5068–5077. https://doi.org/10.1128/AEM.00208-08

Howard, M. T., Gonzales-Flores, J. N., & Copeland, P. R. (2012). Molecular Mechanism of Eukaryotic Selenocysteine Incorporation. In D. L. Hatfield, M. J. Berry, & V. N. Gladyshev (Eds.), *Selenium: Its Molecular Biology and Role in Human Health* (pp. 33–46). Springer. https://doi.org/10.1007/978-1-4614-1025-6_3

Howard, M. T., Moyle, M. W., Aggarwal, G., Carlson, B. A., & Anderson, C. B. (2007). A recoding element that stimulates decoding of UGA codons by Sec tRNA[Ser]Sec. *RNA*, *13*(6), 912–920. https://doi.org/10.1261/rna.473907

I

Ishii, T. M., Kotlova, N., Tapsoba, F., & Steinberg, S. V. (2013). The Long D-stem of the Selenocysteine tRNA Provides Resilience at the Expense of Maximal Function. *The Journal of Biological Chemistry*, *288*(19), 13337–13344. https://doi.org/10.1074/jbc.M112.434704

Itoh, Y., Chiba, S., Sekine, S., & Yokoyama, S. (2009). Crystal structure of human selenocysteine tRNA. *Nucleic Acids Research*, *37*(18), 6259–6268. https://doi.org/10.1093/nar/gkp648

J

Jahn, M. T., Markert, S. M., Ryu, T., Ravasi, T., Stigloher, C., Hentschel, U., & Moitinho-Silva, L. (2016). Shedding light on cell compartmentation in the candidate phylum Poribacteria by high resolution visualisation and transcriptional profiling. *Scientific Reports*, *6*(1), 35860. https://doi.org/10.1038/srep35860

Jeon, Y. H., Park, Y. H., Lee, J. H., Hong, J.-H., & Kim, I. Y. (2014). Selenoprotein W enhances skeletal muscle differentiation by inhibiting TAZ binding to 14-3-3 protein. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1843*(7), 1356–1364. https://doi.org/10.1016/j.bbamcr.2014.04.002

Jiang, L., Lu, Y., Zheng, L., Li, G., Chen, L., Zhang, M., Ni, J., Liu, Q., & Zhang, Y. (2020). The algal selenoproteomes. *BMC Genomics*, *21*, 699. https://doi.org/10.1186/s12864-020-07101-z

Jones, D. P., & Sies, H. (2015). The Redox Code. *Antioxidants & Redox Signaling*, *23*(9), 734–746. https://doi.org/10.1089/ars.2015.6247

Jousset, H., Frieden, M., & Demaurex, N. (2007). STIM1 knockdown reveals that store-operated Ca2+ channels located close to sarco/endoplasmic Ca2+ ATPases (SERCA) pumps silently refill the

endoplasmic reticulum. *The Journal of Biological Chemistry*, *282*(15), 11456–11464. https://doi.org/10.1074/jbc.M609551200

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jurynec, M. J., Xia, R., Mackrill, J. J., Gunther, D., Crawford, T., Flanigan, K. M., Abramson, J. J., Howard, M. T., & Grunwald, D. J. (2008). Selenoprotein N is required for ryanodine receptor calcium release channel activity in human and zebrafish muscle. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(34), 12485–12490. https://doi.org/10.1073/pnas.0806015105

K

Kamke, J., Rinke, C., Schwientek, P., Mavromatis, K., Ivanova, N., Sczyrba, A., Woyke, T., & Hentschel, U. (2014). The Candidate Phylum Poribacteria by Single-Cell Genomics: New Insights into Phylogeny, Cell-Compartmentation, Eukaryote-Like Repeat Proteins, and Other Genomic Features. *PLoS ONE*, *9*(1), e87353. https://doi.org/10.1371/journal.pone.0087353

Kamke, J., Sczyrba, A., Ivanova, N., Schwientek, P., Rinke, C., Mavromatis, K., Woyke, T., & Hentschel, U. (2013). Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *The ISME Journal*, *7*(12), 2287–2300. https://doi.org/10.1038/ismej.2013.111

Kang, H., Jeon, Y. H., Ham, M., Ko, K., & Kim, I. Y. (2021). Compensatory Protection of Thioredoxin-Deficient Cells from Etoposide-Induced Cell Death by Selenoprotein W via Interaction with 14-3-3. *International Journal of Molecular Sciences*, *22*(19), 10338. https://doi.org/10.3390/ijms221910338

Bibliography

Karimi, E., Keller-Costa, T., Slaby, B. M., Cox, C. J., da Rocha, U. N., Hentschel, U., & Costa, R. (2019). Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable Alphaproteobacteria. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-38737-x

Kayrouz, C. M., Huang, J., Hauser, N., & Seyedsayamdost, M. R. (2022). Biosynthesis of selenium-containing small molecules in diverse microorganisms. *Nature*, *610*(7930), Article 7930. https://doi.org/10.1038/s41586-022-05174-2

Khangulov, S. V., Gladyshev, V. N., Dismukes, G. C., & Stadtman, T. C. (1998). Selenium-containing formate dehydrogenase H from Escherichia coli: A molybdopterin enzyme that catalyzes formate oxidation without oxygen transfer. *Biochemistry*, *37*(10), 3518–3528. https://doi.org/10.1021/bi972177k

Kim, H.-Y., Fomenko, D. E., Yoon, Y.-E., & Gladyshev, V. N. (2006). CATALYTIC ADVANTAGES PROVIDED BY SELENOCYSTEINE IN METHIONINE-S-SULFOXIDE REDUCTASES. *Biochemistry*, *45*(46), 13697–13704. https://doi.org/10.1021/bi0611614

Kim, I. Y., & Stadtman, T. C. (1995). Selenophosphate synthetase: Detection in extracts of rat tissues by immunoblot assay and partial purification of the enzyme from the archaean Methanococcus vannielii. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(17), 7710–7713. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC41215/

Kirst, H., & Kerfeld, C. A. (2019). Bacterial microcompartments: Catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering. *BMC Biology*, *17*(1), 79. https://doi.org/10.1186/s12915-019-0691-z

Kryukov, G. V., Kryukov, V. M., & Gladyshev, V. N. (1999). New mammalian selenocysteine-containing proteins identified with an algorithm that searches for selenocysteine insertion sequence elements. *The Journal of Biological Chemistry*, *274*(48), 33888–33897. https://doi.org/10.1074/jbc.274.48.33888

Kubota, K. (2013). CARD-FISH for Environmental Microorganisms: Technical Advancement and Future Applications. *Microbes and Environments*, *28*(1), 3–12. https://doi.org/10.1264/jsme2.ME12107

Kulik, K., Sadowska, K., Wielgus, E., Pacholczyk-Sienicka, B., Sochacka, E., & Nawrot, B. (2022). 2-Selenouridine, a Modified Nucleoside of Bacterial tRNAs, Its Reactivity in the Presence of Oxidizing and Reducing Reagents. *International Journal of Molecular Sciences*, *23*(14), 7973. https://doi.org/10.3390/ijms23147973

L

Labunskyy, V. M., Ferguson, A. D., Fomenko, D. E., Chelliah, Y., Hatfield, D. L., & Gladyshev, V. N. (2005). A novel cysteine-rich domain of Sep15 mediates the interaction with UDP-glucose:glycoprotein glucosyltransferase. *The Journal of Biological Chemistry*, *280*(45), 37839–37845. https://doi.org/10.1074/jbc.M508685200

Labunskyy, V. M., Hatfield, D. L., & Gladyshev, V. N. (2014). Selenoproteins: Molecular Pathways and Physiological Roles. *Physiological Reviews*, *94*(3), 739–777. https://doi.org/10.1152/physrev.00039.2013

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M., & Isaacs, F. J. (2013). Genomically Recoded Organisms Expand Biological Functions. *Science (New York, N.Y.)*, *342*(6156), 357–360. https://doi.org/10.1126/science.1241459

Lasken, R., Raghunathan, A., Kvist, T., Ishøy, T., Westermann, P., Ahring, B., & Boissy, R. (2005). *Multiple displacement amplification from single bacterial cells* (pp. 119–146). https://doi.org/10.13140/2.1.4645.7281

Lee, B. C., Peterfi, Z., Hoffmann, F. W., Moore, R. E., Kaya, A., Avanesov, A., Tarrago, L., Zhou, Y., Weerapana, E., Fomenko, D. E., Hoffmann, P. R., & Gladyshev, V. N. (2013). MsrB1 and

MICALs regulate actin assembly and macrophage function via reversible stereoselective methionine oxidation. *Molecular Cell*, *51*(3), 397–404. https://doi.org/10.1016/j.molcel.2013.06.019

Lee, B. J., Worland, P. J., Davis, J. N., Stadtman, T. C., & Hatfield, D. L. (1989). Identification of a selenocysteyl-tRNA(Ser) in mammalian cells that recognizes the nonsense codon, UGA. *The Journal of Biological Chemistry*, *264*(17), 9724–9727. DOI:https://doi.org/10.1016/S0021-9258(18)81714-8

Leiter, O., Zhuo, Z., Rust, R., Wasielewska, J. M., Grönnert, L., Kowal, S., Overall, R. W., Adusumilli, V. S., Blackmore, D. G., Southon, A., Ganio, K., McDevitt, C. A., Rund, N., Brici, D., Mudiyan, I. A., Sykes, A. M., Rünker, A. E., Zocher, S., Ayton, S., … Walker, T. L. (2022). Selenium mediates exercise-induced adult neurogenesis and reverses learning deficits induced by hippocampal injury and aging. *Cell Metabolism*, *34*(3), 408-423.e8. https://doi.org/10.1016/j.cmet.2022.01.005

Lescure, A., Baltzinger, M., & Zito, E. (2018). Uncovering the Importance of Selenium in Muscle Disease. In *Selenium* (pp. 345–362). Springer International Publishing. https://doi.org/10.1007/978-3-319-95390-8_18

Lescure, A., Briens, M., & Ferreiro, A. (2016). What Do We Know About Selenium Contributions to Muscle Physiology? In D. L. Hatfield, U. Schweizer, P. A. Tsuji, & V. N. Gladyshev (Eds.), *Selenium* (pp. 475–486). Springer International Publishing. https://doi.org/10.1007/978-3-319-41283-2_40

Lescure, A., Gautheret, D., Carbon, P., & Krol, A. (1999). Novel Selenoproteins Identified in Silico andin Vivo by Using a Conserved RNA Structural Motif*. *Journal of Biological Chemistry*, *274*(53), 38147–38154. https://doi.org/10.1074/jbc.274.53.38147

Li, K., Feng, T., Liu, L., Liu, H., Huang, K., & Zhou, J. (2021). Hepatic Proteomic Analysis of Selenoprotein T Knockout Mice by TMT: Implications for the Role of Selenoprotein T in

Glucose and Lipid Metabolism. *International Journal of Molecular Sciences*, *22*(16), 8515. https://doi.org/10.3390/ijms22168515

Li, Y., & Camacho, P. (2004). Ca2+-dependent redox modulation of SERCA 2b by ERp57. *The Journal of Cell Biology*, *164*(1), 35–46. https://doi.org/10.1083/jcb.200307010

Lindsay, M. R., Webb, R. I., & Fuerst, J. A. (1997). Pirellulosomes: A new type of membrane-bounded cell compartment in planctomycete bacteria of the genus Pirellula. *Microbiology*, *143*(3), 739–748. https://doi.org/10.1099/00221287-143-3-739

Lobanov, A. V., Fomenko, D. E., Zhang, Y., Sengupta, A., Hatfield, D. L., & Gladyshev, V. N. (2007). Evolutionary dynamics of eukaryotic selenoproteomes: Large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome Biology*, *8*(9), R198. https://doi.org/10.1186/gb-2007-8-9-r198

Lobanov, A. V., Hatfield, D. L., & Gladyshev, V. N. (2008). Reduced reliance on the trace element selenium during evolution of mammals. *Genome Biology*, *9*(3), R62. https://doi.org/10.1186/gb-2008-9-3-r62

Lobanov, A. V., Hatfield, D. L., & Gladyshev, V. N. (2009). Eukaryotic selenoproteins and selenoproteomes. *Biochimica et Biophysica Acta*, *1790*(11), 1424. https://doi.org/10.1016/j.bbagen.2009.05.014

Low, S. C., Grundner-Culemann, E., Harney, J. W., & Berry, M. J. (2000). SECIS–SBP2 interactions dictate selenocysteine incorporation efficiency and selenoprotein hierarchy. *The EMBO Journal*, *19*(24), 6882–6890. https://doi.org/10.1093/emboj/19.24.6882

M

Maia, L. B., Moura, J. J. G., & Moura, I. (2015). Molybdenum and tungsten-dependent formate dehydrogenases. *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry*, *20*(2), 287–309. https://doi.org/10.1007/s00775-014-1218-2

Maiti, B., Arbogast, S., Allamand, V., Moyle, M., Anderson, C., Richard, P., Guicheney, P., Ferreiro, A., Flanigan, K., & Howard, M. (2009). A mutation in the SEPN1 SRE reduces selenocysteine incorporation and leads to SEPN1-related myopathy. *Human Mutation*, *30*(3), 411–416. https://doi.org/10.1002/humu.20879

Mandal, M. K., Chanu, Ng. K., & Chaurasia, N. (2020). Cyanobacterial pigments and their fluorescence characteristics: Applications in research and industry. In *Advances in Cyanobacterial Biology* (pp. 55–72). Elsevier. https://doi.org/10.1016/B978-0-12-819311-2.00005-X

Mangiapane, E., Pessione, A., & Pessione, E. (2014). Selenium and Selenoproteins: An Overview on Different Biological Systems. *Current Protein & Peptide Science*, *15*(6), 598–607. https://doi.org/10.2174/1389203715666140608151134

Manta, B., Makarova, N. E., & Mariotti, M. (2022). The selenophosphate synthetase family: A review. *Free Radical Biology and Medicine*, *192*, 63–76. https://doi.org/10.1016/j.freeradbiomed.2022.09.007

Marino, M., Stoilova, T., Giorgi, C., Bachi, A., Cattaneo, A., Auricchio, A., Pinton, P., & Zito, E. (2015). SEPN1, an endoplasmic reticulum-localized selenoprotein linked to skeletal muscle pathology, counteracts hyperoxidation by means of redox-regulating SERCA2 pump activity. *Human Molecular Genetics*, *24*(7), 1843–1855. https://doi.org/10.1093/hmg/ddu602

Mariotti, M. (2016). Selenocysteine Extinctions in Insects. In C. Raman, M. R. Goldsmith, & T. A. Agunbiade (Eds.), *Short Views on Insect Genomics and Proteomics: Insect Proteomics, Vol.2* (pp. 113–140). Springer International Publishing. https://doi.org/10.1007/978-3-319-24244-6_5

Bibliography

Mariotti, M., Lobanov, A. V., Manta, B., Santesmasses, D., Bofill, A., Guigó, R., Gabaldón, T., & Gladyshev, V. N. (2016). Lokiarchaeota Marks the Transition between the Archaeal and Eukaryotic Selenocysteine Encoding Systems. *Molecular Biology and Evolution*, *33*(9), 2441–2453. https://doi.org/10.1093/molbev/msw122

Mariotti, M., Ridge, P. G., Zhang, Y., Lobanov, A. V., Pringle, T. H., Guigo, R., Hatfield, D. L., & Gladyshev, V. N. (2012). Composition and Evolution of the Vertebrate and Mammalian Selenoproteomes. *PLOS ONE*, *7*(3), e33066. https://doi.org/10.1371/journal.pone.0033066

Mariotti, M., Salinas, G., Gabaldón, T., & Gladyshev, V. N. (2019). Utilization of selenocysteine in early-branching fungal phyla. *Nature Microbiology*, *4*(5), 759–765. https://doi.org/10.1038/s41564-018-0354-9

Martinez-Porchas, M., Villalpando-Canchola, E., Ortiz Suarez, L. E., & Vargas-Albores, F. (2017). How conserved are the conserved 16S-rRNA regions? *PeerJ*, *5*, e3036. https://doi.org/10.7717/peerj.3036

McClung, J. P., Roneker, C. A., Mu, W., Lisk, D. J., Langlais, P., Liu, F., & Lei, X. G. (2004). Development of insulin resistance and obesity in mice overexpressing cellular glutathione peroxidase. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(24), 8852–8857. https://doi.org/10.1073/pnas.0308096101

Meng, K., Chung, C. Z., Söll, D., & Krahn, N. (2022). Unconventional genetic code systems in archaea. *Frontiers in Microbiology*, *13*. https://www.frontiersin.org/articles/10.3389/fmicb.2022.1007832

Moghadaszadeh, B., Petit, N., Jaillard, C., Brockington, M., Roy, S. Q., Merlini, L., Romero, N., Estournet, B., Desguerre, I., Chaigne, D., Muntoni, F., Topaloglu, H., & Guicheney, P. (2001). Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome. *Nature Genetics*, *29*(1), Article 1. https://doi.org/10.1038/ng713

Moghadaszadeh, B., Rider, B. E., Lawlor, M. W., Childers, M. K., Grange, R. W., Gupta, K., Boukedes, S. S., Owen, C. A., & Beggs, A. H. (2013). Selenoprotein N deficiency in mice is associated with

abnormal lung development. *The FASEB Journal*, *27*(4), 1585–1599. https://doi.org/10.1096/fj.12-212688

Morra, S. (2022). Fantastic [FeFe]-Hydrogenases and Where to Find Them. *Frontiers in Microbiology*, *13*. https://www.frontiersin.org/articles/10.3389/fmicb.2022.853626

N

Nguyen, M. T. H. D., Liu, M., & Thomas, T. (2014). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Molecular Ecology*, *23*(6), 1635–1645. https://doi.org/10.1111/mec.12384

Nicacio, K. J., Ióca, L. P., Fróes, A. M., Leomil, L., Appolinario, L. R., Thompson, C. C., Thompson, F. L., Ferreira, A. G., Williams, D. E., Andersen, R. J., Eustaquio, A. S., & Berlinck, R. G. S. (2017, February 13). *Cultures of the Marine Bacterium Pseudovibrio denitrificans Ab134 Produce Bromotyrosine-Derived Alkaloids Previously Only Isolated from Marine Sponges* (world) [Rapid-communication]. ACS Publications; American Chemical Society and American Society of Pharmacognosy. https://doi.org/10.1021/acs.jnatprod.6b00838

Noda, Y., Okada, S., & Suzuki, T. (2022). Regulation of A-to-I RNA editing and stop codon recoding to control selenoprotein expression during skeletal myogenesis. *Nature Communications*, *13*, 2503. https://doi.org/10.1038/s41467-022-30181-2

O

Orfanoudaki, M., Hartmann, A., Alilou, M., Mehic, N., Kwiatkowski, M., Jöhrer, K., Nguyen Ngoc, H., Hensel, A., Greil, R., & Ganzera, M. (2021). Cytotoxic Compounds of Two Demosponges (Aplysina aerophoba and Spongia sp.) from the Aegean Sea. *Biomolecules*, *11*(5), 723. https://doi.org/10.3390/biom11050723

P

Pan, J. L., & Bardwell, J. C. A. (2006). The origami of thioredoxin-like folds. *Protein Science*, *15*(10), 2217–2227. https://doi.org/10.1110/ps.062268106

Peng, T., Lin, J., Xu, Y.-Z., & Zhang, Y. (2016). Comparative genomics reveals new evolutionary and ecological patterns of selenium utilization in bacteria. *The ISME Journal*, *10*(8), 2048–2059. https://doi.org/10.1038/ismej.2015.246

Perrone, M., Caroccia, N., Genovese, I., Missiroli, S., Modesti, L., Pedriali, G., Vezzani, B., Vitto, V. A. M., Antenori, M., Lebiedzinska-Arciszewska, M., Wieckowski, M. R., Giorgi, C., & Pinton, P. (2020). Chapter Four—The role of mitochondria-associated membranes in cellular homeostasis and diseases. In O. Kepp & L. Galluzzi (Eds.), *International Review of Cell and Molecular Biology* (Vol. 350, pp. 119–196). Academic Press. https://doi.org/10.1016/bs.ircmb.2019.11.002

Peters, K., Galinn, S., & Tsuji, P. (2016). Selenium: Dietary Sources, Human Nutritional Requirements and Intake Across Populations. In *Selenium: Its Molecular Biology and Role in Human Health: Fourth Edition* (pp. 295–305). https://doi.org/10.1007/978-3-319-41283-2_25

Petit, N. (2003). Selenoprotein N: An endoplasmic reticulum glycoprotein with an early developmental expression pattern. *Human Molecular Genetics*, *12*(9), 1045–1053. https://doi.org/10.1093/hmg/ddg115

Pita, L., Rix, L., Slaby, B. M., Franke, A., & Hentschel, U. (2018). The sponge holobiont in a changing ocean: From microbes to ecosystems. *Microbiome*, *6*(1), 46. https://doi.org/10.1186/s40168-018-0428-1

Pitts, M. W., & Hoffmann, P. R. (2018). Endoplasmic reticulum-resident selenoproteins as regulators of calcium signaling and homeostasis. *Cell Calcium*, *70*, 76–86. https://doi.org/10.1016/j.ceca.2017.05.001

Pitts, M. W., Reeves, M. A., Hashimoto, A. C., Ogawa, A., Kremer, P., Seale, L. A., & Berry, M. J. (2013). Deletion of Selenoprotein M Leads to Obesity without Cognitive Deficits. *The Journal of Biological Chemistry*, *288*(36), 26121–26134. https://doi.org/10.1074/jbc.M113.471235

Podell, S., Blanton, J. M., Neu, A., Agarwal, V., Biggs, J. S., Moore, B. S., & Allen, E. E. (2019). Pangenomic comparison of globally distributed Poribacteria associated with sponge hosts and marine particles. *The ISME Journal*, *13*(2), 468–481. https://doi.org/10.1038/s41396-018-0292-9

Pothion, H., Jehan, C., Tostivint, H., Cartier, D., Bucharles, C., Falluel-Morel, A., Boukhzar, L., Anouar, Y., & Lihrmann, I. (2020). Selenoprotein T: An Essential Oxidoreductase Serving as a Guardian of Endoplasmic Reticulum Homeostasis. *Antioxidants & Redox Signaling*, *33*(17), 1257–1275. https://doi.org/10.1089/ars.2019.7931

Prevost, G., Arabo, A., Jian, L., Quelennec, E., Cartier, D., Hassan, S., Falluel-Morel, A., Tanguy, Y., Gargani, S., Lihrmann, I., Kerr-Conte, J., Lefebvre, H., Pattou, F., & Anouar, Y. (2013). The PACAP-Regulated Gene Selenoprotein T Is Abundantly Expressed in Mouse and Human β-Cells and Its Targeted Inactivation Impairs Glucose Tolerance. *Endocrinology*, *154*(10), 3796–3806. https://doi.org/10.1210/en.2013-1167

R

Rederstorff, M., Castets, P., Arbogast, S., Lainé, J., Vassilopoulos, S., Beuvin, M., Dubourg, O., Vignaud, A., Ferry, A., Krol, A., Allamand, V., Guicheney, P., Ferreiro, A., & Lescure, A. (2011). Increased Muscle Stress-Sensitivity Induced by Selenoprotein N Inactivation in Mouse: A Mammalian Model for SEPN1-Related Myopathy. *PLoS ONE*, *6*(8), e23094. https://doi.org/10.1371/journal.pone.0023094

Reeves, M. A., & Hoffmann, P. R. (2009). The human selenoproteome: Recent insights into functions and regulation. *Cellular and Molecular Life Sciences : CMLS*, *66*(15), 2457–2478. https://doi.org/10.1007/s00018-009-0032-4

Reeves, M. A., Bellinger, F. P., & Berry, M. J. (2010). The Neuroprotective Functions of Selenoprotein M and its Role in Cytosolic Calcium Regulation. *Antioxidants & Redox Signaling*, *12*(7), 809–818. https://doi.org/10.1089/ars.2009.2883

Reich, H. J., & Hondal, R. J. (2016). Why Nature Chose Selenium. *ACS Chemical Biology*, *11*(4), 821–841. https://doi.org/10.1021/acschembio.6b00031

Robbins, S. J., Song, W., Engelberts, J. P., Glasl, B., Slaby, B. M., Boyd, J., Marangon, E., Botté, E. S., Laffy, P., Thomas, T., & Webster, N. S. (2021). A genomic view of the microbiome of coral reef demosponges. *The ISME Journal*, *15*(6), Article 6. https://doi.org/10.1038/s41396-020-00876-9

Rodriguez-Ruiz, A., Braun, D., Pflug, S., Brol, A., Sylvester, M., Steegborn, C., & Schweizer, U. (2022). Insights into the Mechanism of Human Deiodinase 1. *International Journal of Molecular Sciences*, *23*(10), Article 10. https://doi.org/10.3390/ijms23105361

Romero, H., Zhang, Y., Gladyshev, V. N., & Salinas, G. (2005). Evolution of selenium utilization traits. *Genome Biology*, *6*(8), R66. https://doi.org/10.1186/gb-2005-6-8-r66

S

Santesmasses, D., Mariotti, M., & Gladyshev, V. N. (2020). Bioinformatics of Selenoproteins. *Antioxidants & Redox Signaling*, *33*(7), 525–536. https://doi.org/10.1089/ars.2020.8044

Schomburg, L. (2022). Selenoprotein P – Selenium transport protein, enzyme and biomarker of selenium status. *Free Radical Biology and Medicine*, *191*, 150–163. https://doi.org/10.1016/j.freeradbiomed.2022.08.022

Schwarz, K., & Foltz, C. M. (1958). Factor 3 Activity of Selenium Compounds. *Journal of Biological Chemistry*, *233*(1), 245–251. https://doi.org/10.1016/S0021-9258(19)68065-8

Schweizer, U., Schlicker, C., Braun, D., Köhrle, J., & Steegborn, C. (2014). Crystal structure of mammalian selenocysteine-dependent iodothyronine deiodinase suggests a peroxiredoxin-like catalytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(29), 10526–10531. https://doi.org/10.1073/pnas.1323873111

Schweizer, U., Wirth, E. K., Klopstock, T., Hölter, S. M., Becker, L., Moskovitz, J., Grune, T., Fuchs, H., Gailus-Durner, V., Hrabe de Angelis, M., Köhrle, J., & Schomburg, L. (2022). Seizures, ataxia and parvalbumin-expressing interneurons respond to selenium supply in Selenop-deficient mice. *Redox Biology*, *57*, 102490. https://doi.org/10.1016/j.redox.2022.102490

Seiler, A., Schneider, M., Förster, H., Roth, S., Wirth, E. K., Culmsee, C., Plesnila, N., Kremmer, E., Rådmark, O., Wurst, W., Bornkamm, G. W., Schweizer, U., & Conrad, M. (2008). Glutathione Peroxidase 4 Senses and Translates Oxidative Stress into 12/15-Lipoxygenase Dependent- and AIF-Mediated Cell Death. *Cell Metabolism*, *8*(3), 237–248. https://doi.org/10.1016/j.cmet.2008.07.005

Shakoori, A. R. (2017). Fluorescence In Situ Hybridization (FISH) and Its Applications. *Chromosome Structure and Aberrations*, 343–367. https://doi.org/10.1007/978-81-322-3673-3_16

Shchedrina, V. A., Everley, R. A., Zhang, Y., Gygi, S. P., Hatfield, D. L., & Gladyshev, V. N. (2011). Selenoprotein K Binds Multiprotein Complexes and Is Involved in the Regulation of Endoplasmic Reticulum Homeostasis. *The Journal of Biological Chemistry*, *286*(50), 42937–42948. https://doi.org/10.1074/jbc.M111.310920

Shmakov, S. A., Faure, G., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2019). Systematic prediction of functionally linked genes in bacterial and archaeal genomes. *Nature Protocols*, *14*(10), 3013–3031. https://doi.org/10.1038/s41596-019-0211-1

Siegl, A., & Hentschel, U. (2010). PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environmental Microbiology Reports*, *2*(4), 507–513. https://doi.org/10.1111/j.1758-2229.2009.00057.x

Siegl, A., Bayer, K., Kozytska, S., Hentschel, U., & Schmitt, S. (2008). SPONGES AND MICROBES – NEW FRONTIERS IN AN ANCIENT SYMBIOSIS. *Vie Milieu*, 11.

Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., Dandekar, T., & Hentschel, U. (2011). Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum

symbiotically associated with marine sponges. *The ISME Journal*, *5*(1), 61–70. https://doi.org/10.1038/ismej.2010.95

Silva, J. E. (1995). Thyroid Hormone Control of Thermogenesis and Energy Balance. *Thyroid*, *5*(6), 481–492. https://doi.org/10.1089/thy.1995.5.481

Silva, J. E., & Larsen, P. R. (1985). Potential of brown adipose tissue type II thyroxine 5'-deiodinase as a local and systemic source of triiodothyronine in rats. *Journal of Clinical Investigation*, *76*(6), 2296–2305. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC424353/

Silwal, A., Sarkozy, A., Scoto, M., Ridout, D., Schmidt, A., Laverty, A., Henriques, M., D'Argenzio, L., Main, M., Mein, R., Manzur, A. Y., Abel, F., Al-Ghamdi, F., Genetti, C. A., Ardicli, D., Haliloglu, G., Topaloglu, H., Beggs, A. H., & Muntoni, F. (2020). Selenoprotein N-related myopathy: A retrospective natural history study to guide clinical trials. *Annals of Clinical and Translational Neurology*, *7*(11), 2288–2296. https://doi.org/10.1002/acn3.51218

Small-Howard, A., Morozova, N., Stoytcheva, Z., Forry, E. P., Mansell, J. B., Harney, J. W., Carlson, B. A., Xu, X., Hatfield, D. L., & Berry, M. J. (2006). Supramolecular Complexes Mediate Selenocysteine Incorporation In Vivo. *Molecular and Cellular Biology*, *26*(6), 2337–2346. https://doi.org/10.1128/MCB.26.6.2337-2346.2006

Stockhausen, H. B. V. (1988). Selenium in total parenteral nutrition. *Biological Trace Element Research*, 1988 Jan-Apr;15:147-55. doi: 10.1007/BF02990133. PMID: 2484513.

Sturchler, C., Westhof, E., Carbon, P., & Krol, A. (1993). Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). *Nucleic Acids Research*, *21*(5), 1073–1079. https://doi.org/10.1093/nar/21.5.1073

T

Taboada, B., Estrada, K., Ciria, R., & Merino, E. (2018). Operon-mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, *34*(23), 4118–4120. https://doi.org/10.1093/bioinformatics/bty496

Takeuchi, A., Schmitt, D., Chapple, C., Babaylova, E., Karpova, G., Guigo, R., Krol, A., & Allmang, C. (2009). A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins. *Nucleic Acids Research*, *37*(7), 2126–2141. https://doi.org/10.1093/nar/gkp078

Tanguy, Y., Falluel-Morel, A., Arthaud, S., Boukhzar, L., Manecka, D.-L., Chagraoui, A., Prevost, G., Elias, S., Dorval-Coiffec, I., Lesage, J., Vieau, D., Lihrmann, I., Jégou, B., & Anouar, Y. (2011). The PACAP-regulated gene selenoprotein T is highly induced in nervous, endocrine, and metabolic tissues during ontogenetic and regenerative processes. *Endocrinology*, *152*(11), 4322. https://doi.org/10.1210/en.2011-1246

Tepedelen, B. E., Kirmizibayrak, P. B., Tepedelen, B. E., & Kirmizibayrak, P. B. (2019). Endoplasmic Reticulum-Associated Degradation (ERAD). In *Endoplasmic Reticulum*. IntechOpen. https://doi.org/10.5772/intechopen.82043

Tian, R.-M., Wang, Y., Bougouffa, S., Gao, Z.-M., Cai, L., Bajic, V., & Qian, P.-Y. (2014). Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge. *Environmental Microbiology*, *16*(11), 3548–3561. https://doi.org/10.1111/1462-2920.12586

Tsuji, P. A., Santesmasses, D., Lee, B. J., Gladyshev, V. N., & Hatfield, D. L. (2022). Historical Roles of Selenium and Selenoproteins in Health and Development: The Good, the Bad and the Ugly. *International Journal of Molecular Sciences*, *23*(1), Article 1. https://doi.org/10.3390/ijms23010005

Tujebajeva, R. M., Copeland, P. R., Xu, X.-M., Carlson, B. A., Harney, J. W., Driscoll, D. M., Hatfield, D. L., & Berry, M. J. (2000). Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Reports*, *1*(2), 158–163. https://doi.org/10.1093/embo-reports/kvd033

Tujula, N. A., Holmström, C., Mussmann, M., Amann, R., Kjelleberg, S., & Crocetti, G. R. (2006). A CARD-FISH protocol for the identification and enumeration of epiphytic bacteria on marine algae. *Journal of Microbiological Methods*, *65*(3), 604–607. https://doi.org/10.1016/j.mimet.2005.09.006

Turon, X., Becerro, M. A., & Uriz, M. J. (2000). Distribution of brominated compounds within the sponge Aplysina aerophoba: Coupling of X-ray microanalysis with cryofixation techniques. *Cell and Tissue Research*, *301*(2), 311–322. https://doi.org/10.1007/s004410000233

U

Unson, M. D., Holland, N. D., & Faulkner, D. J. (1994). A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Marine Biology*, *119*(1), 1–11. https://doi.org/10.1007/BF00350100

Ursini, F., & Maiorino, M. (2020). Lipid peroxidation and ferroptosis: The role of GSH and GPx4. *Free Radical Biology and Medicine*, *152*, 175–185. https://doi.org/10.1016/j.freeradbiomed.2020.02.027

V

Versluis, D., McPherson, K., van Passel, M. W. J., Smidt, H., & Sipkema, D. (2017). Recovery of Previously Uncultured Bacterial Genera from Three Mediterranean Sponges. *Marine Biotechnology (New York, N.y.)*, *19*(5), 454–468. https://doi.org/10.1007/s10126-017-9766-4

Vignais, P. M., & Billoud, B. (2007). Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chemical Reviews*, *107*(10), 4206–4272. https://doi.org/10.1021/cr050196r

Villalobos, C., Sobradillo, D., Hernández-Morales, M., & Núñez, L. (2016). Remodeling of Calcium Entry Pathways in Cancer. *Advances in Experimental Medicine and Biology*, *898*, 449–466. https://doi.org/10.1007/978-3-319-26974-0_19

Villar-Quiles, R. N., von der Hagen, M., Métay, C., Gonzalez, V., Donkervoort, S., Bertini, E., Castiglioni, C., Chaigne, D., Colomer, J., Cuadrado, M. L., de Visser, M., Desguerre, I., Eymard, B., Goemans, N., Kaindl, A., Lagrue, E., Lütschg, J., Malfatti, E., Mayer, M., … Ferreiro, A. (2020). The clinical, histologic, and genotypic spectrum of SEPN1-related myopathy. *Neurology*, *95*(11), e1512–e1527. https://doi.org/10.1212/WNL.0000000000010327

Vindry, C., Ohlmann, T., & Chavatte, L. (2018). Translation regulation of mammalian selenoproteins. *Biochimica Et Biophysica Acta. General Subjects*, *1862*(11), 2480–2492. https://doi.org/10.1016/j.bbagen.2018.05.010

W

Walter, P., & Ron, D. (2011). The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. *Science*, *334*(6059), 1081–1086. https://doi.org/10.1126/science.1209038

Webster, N. S., & Thomas, T. (2016). The Sponge Hologenome. *MBio*, *7*(2), e00135-16. https://doi.org/10.1128/mBio.00135-16

Weekley, C. M., & Harris, H. H. (2013). Which form is that? The importance of selenium speciation and metabolism in the prevention and treatment of disease. *Chemical Society Reviews*, *42*(23), 8870–8894. https://doi.org/10.1039/c3cs60272a

Wehrl, M., Steinert, M., & Hentschel, U. (2007). Bacterial Uptake by the Marine Sponge Aplysina aerophoba. *Microbial Ecology*, *53*(2), 355–365. https://doi.org/10.1007/s00248-006-9090-4

Wesolowski, L. T., Semanchik, P. L., & White-Springer, S. H. (2022). Beyond antioxidants: Selenium and skeletal muscle mitochondria. *Frontiers in Veterinary Science*, *9*. https://www.frontiersin.org/articles/10.3389/fvets.2022.1011159

Wu, H. H., Brennan, C., & Ashworth, R. (2011). Ryanodine receptors, a family of intracellular calcium ion channels, are expressed throughout early vertebrate development. *BMC Research Notes*, *4*, 541. https://doi.org/10.1186/1756-0500-4-541

Wu, R., Shen, Q., & Newburger, P. E. (2000). Recognition and binding of the human selenocysteine insertion sequence by nucleolin. *Journal of Cellular Biochemistry*, *77*(3), 507–516. https://doi.org/10.1002/(SICI)1097-4644(20000601)77:3<507::AID-JCB15>3.0.CO;2-P

X

Xu, H., & Van Remmen, H. (2021). The SarcoEndoplasmic Reticulum Calcium ATPase (SERCA) pump: A potential target for intervention in aging and skeletal muscle pathologies. *Skeletal Muscle*, *11*(1), 25. https://doi.org/10.1186/s13395-021-00280-7

Xu, X.-M., Carlson, B. A., Irons, R., Mix, H., Zhong, N., Gladyshev, V. N., & Hatfield, D. L. (2007). Selenophosphate synthetase 2 is essential for selenoprotein biosynthesis. *Biochemical Journal*, *404*(Pt 1), 115–120. https://doi.org/10.1042/BJ20070165

Xu, X.-M., Carlson, B. A., Mix, H., Zhang, Y., Saira, K., Glass, R. S., Berry, M. J., Gladyshev, V. N., & Hatfield, D. L. (2007). Biosynthesis of Selenocysteine on Its tRNA in Eukaryotes. *PLoS Biology*, *5*(1), e4. https://doi.org/10.1371/journal.pbio.0050004

Y

Yamashita, M., Yamashita, Y.,. (2010). Identification of a Novel Selenium-containing Compound, Selenoneine, as the Predominant Chemical Form of Organic Selenium in the Blood of Bluefin Tuna. *The Journal of Biological Chemistry*, *285*(24), 18134–18138. https://doi.org/10.1074/jbc.C110.106377

Yamashita, M., Yamashita, Y., Suzuki, T., Kani, Y., Mizusawa, N., Imamura, S., Takemoto, K., Hara, T., Hossain, M. A., Yabu, T., & Touhata, K. (2013). Selenoneine, a novel selenium-containing compound, mediates detoxification mechanisms against methylmercury accumulation and toxicity in zebrafish embryo. *Marine Biotechnology (New York, N.Y.)*, *15*(5), 559–570. https://doi.org/10.1007/s10126-013-9508-1

Yamazaki, S. (1982). A selenium-containing hydrogenase from Methanococcus vannielii. Identification of the selenium moiety as a selenocysteine residue. *The Journal of Biological Chemistry*, *257*(14), 7926–7929. DOI:https://doi.org/10.1016/S0021-9258(18)34271-6

Yang, B., Wang, Y., & Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, *17*(1), 135. https://doi.org/10.1186/s12859-016-0992-y

Yang, J., Lee, S. H., Ryu, J.-Y., Lee, H. S., & Kang, S. G. (2022). A Novel NADP-Dependent Formate Dehydrogenase From the Hyperthermophilic Archaeon Thermococcus onnurineus NA1. *Frontiers in Microbiology*, *13*. https://www.frontiersin.org/articles/10.3389/fmicb.2022.844735

Yant, L. J., Ran, Q., Rao, L., Van Remmen, H., Shibatani, T., Belter, J. G., Motta, L., Richardson, A., & Prolla, T. A. (2003). The selenoprotein GPX4 is essential for mouse development and protects from radiation and oxidative damage insults. *Free Radical Biology & Medicine*, *34*(4), 496–502. https://doi.org/10.1016/s0891-5849(02)01360-6

Ye, Y., Shibata, Y., Yun, C., Ron, D., & Rapoport, T. A. (2004). A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. *Nature*, *429*(6994), 841–847. https://doi.org/10.1038/nature02656

Yeganeh, B., Jäger, R., Gorman, A. M., Samali, A., & Ghavami, S. (2015). Chapter 5 - Induction of Autophagy: Role of Endoplasmic Reticulum Stress and Unfolded Protein Response. In M. A. Hayat (Ed.), *Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging* (pp. 91–101). Academic Press. https://doi.org/10.1016/B978-0-12-801043-3.00005-4

Yim, S. Y., Chae, K. R., Shim, S. B., Hong, J. T., Park, J. Y., Lee, C. Y., Son, H. J., Sheen, Y. Y., & Hwang, D. Y. (2009). ERK activation induced by selenium treatment significantly downregulates β/γ-secretase activity and Tau phosphorylation in the transgenic rat overexpressing human selenoprotein M. *International Journal of Molecular Medicine*, *24*(1), 91–96. https://doi.org/10.3892/ijmm_00000211

Yu, F., Qi, X., Shang, Y., Ping, Z., & Guo, X. (2019). Prevention and control strategies for children Kashin–Beck disease in China: A systematic review and meta-analysis. *Medicine*, *98*(36), e16823. https://doi.org/10.1097/MD.0000000000016823

Z

Zhang, F., Blasiak, L. C., Karolin, J. O., Powell, R. J., Geddes, C. D., & Hill, R. T. (2015). Phosphorus sequestration in the form of polyphosphate by microbial symbionts in marine sponges. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(14), 4381–4386. https://doi.org/10.1073/pnas.1423768112

Zhang, Y., Jin, J., Huang, B., Ying, H., He, J., & Jiang, L. (2022). Selenium Metabolism and Selenoproteins in Prokaryotes: A Bioinformatics Perspective. *Biomolecules*, *12*(7), 917. https://doi.org/10.3390/biom12070917

Zhang, Y., Romero, H., Salinas, G., & Gladyshev, V. N. (2006). Dynamic evolution of selenocysteine utilization in bacteria: A balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome Biology*, *7*(10), R94. https://doi.org/10.1186/gb-2006-7-10-r94

Zhang, Y., Zhou, Y., Schweizer, U., Savaskan, N. E., Hua, D., Kipnis, J., Hatfield, D. L., & Gladyshev, V. N. (2008). Comparative Analysis of Selenocysteine Machinery and Selenoproteome Gene Expression in Mouse Brain Identifies Neurons as Key Functional Sites of Selenium in Mammals *. *Journal of Biological Chemistry*, *283*(4), 2427–2438. https://doi.org/10.1074/jbc.M707951200

Zheng, J., & Conrad, M. (2020). The Metabolic Underpinnings of Ferroptosis. *Cell Metabolism*, *32*(6), 920–937. https://doi.org/10.1016/j.cmet.2020.10.011

Zhu, J.-H., & Lei, X. G. (2006). Double null of selenium-glutathione peroxidase-1 and copper, zinc-superoxide dismutase enhances resistance of mouse primary hepatocytes to acetaminophen toxicity. *Experimental Biology and Medicine (Maywood, N.J.)*, *231*(5), 545–552. https://doi.org/10.1177/153537020623100508

Zito, E., & Ferreiro, A. (2021). Calcium and Redox Liaison: A Key Role of Selenoprotein N in Skeletal Muscle. *Cells*, *10*(5), 1116. https://doi.org/10.3390/cells10051116

Le résumé de la thèse en français

Le résumé de la thèse en français

Sujet de Thèse : **Caractérisation génomique et biochimique d'un homologue bactérien de la Sélénoprotéine N, une protéine à sélénium impliquée dans différentes dystrophies musculaires congénitales**

1. Introduction

L'objectif principal du laboratoire est l'identification de la fonction de la sélénoprotéine N (SelenoN), une protéine contenant du sélénium, dont le gène est lié à différentes formes de maladies musculaires héréditaires chez l'homme, également appelées myopathies liées à SELENON (Lescure et al., 2018). Les sélénoprotéines contiennent l'oligoélément essentiel sélénium sous sa forme biologique majoritaire, la sélénocystéine (Sec). Ce 21ème acide aminé est inséré de manière co-traductionnelle dans la chaîne polypeptidique à l'aide d'un mécanisme de traduction spécialisé impliquant à la fois une structure de l'ARNm et des facteurs de traduction dédiés (Allmang et al., 2006). Vingt-cinq sélénoprotéines ont été identifiées dans le génome humain ; pour environ la moitié d'entre elles, la fonction reste inconnue (Tsuji et al., 2022). La présence dans les sélénoprotéines de Sec à la place de son homologue soufré cystéine (Cys) confère une réactivité accrue et une meilleure résistance à l'oxydation (Reich & Hondal, 2016).

La sélénoN est une glycoprotéine transmembranaire de type II, l'une des cinq sélénoprotéines résidentes du réticulum endoplasmique (RE), et elle est exprimée de manière ubiquitaire dans tous les tissus. Le site catalytique prédit sur la base de la localisation du résidu Sec présente des similitudes avec celui d'une autre séléno-enzyme, la thiorédoxine réductase, suggérant une activité redox, fonction commune à la plupart des sélénoprotéines. De plus, SelenoN contient un domaine de liaison au calcium, conférant une activité redox sensible à la concentration en calcium (Chernorudskiy et al., 2020). Au niveau cellulaire, il a été démontré que SelenoN joue un rôle dans le contrôle du stress oxydatif dans le RE, le maintien de l'homéostasie du calcium dans la cellule, ainsi qu'un lien avec l'activité mitochondriale (Zito & Ferreiro, 2021). De plus, des modèles animaux ont démontré que la fonction de SelenoN est importante pour le développement et le maintien des structures musculaires (Lescure et al., 2018). Cependant, aucune activité catalytique n'a été attribuée à SelenoN jusqu'à présent, aucune fonction n'a pu être prédite sur la base de l'homologie de séquence avec une protéine connue, et à l'heure actuelle, aucun traitement pour la maladie musculaire n'est disponible.

Le phénotype clinique lié à la perte de fonction de SelenoN a initialement étiqueté cette protéine comme ayant une activité spécifique du muscle et a donc orienté la recherche spécifiquement

vers ce tissu. Cependant, l'expression de SelenoN n'est pas limitée au muscle ; elle est exprimée dans tous les tissus analysés. De plus, ce gène est conservé au cours de l'évolution dans le règne animal, y compris chez les eucaryotes inférieurs comme les éponges, qui sont dépourvus de structure mésodermique. Ceci suggère que SelenoN possède un spectre de fonction plus ancien et plus large, qui n'est pas uniquement limitée à l'activité musculaire. De façon intéressante, nous avons identifié un homologue du gène *SELENON* dans un groupe de bactéries encore non cultivées, *Candidatus Poribacteria,* une bactérie qui vit en symbiose avec des éponges marines, comme *Aplysina aerophoba*. *C. Poribacteria* appartient à un groupe de bactéries non classifiées, caractérisées par la présence de membranes intracellulaires (Fieseler et al., 2004 ; Kamke et al., 2014). En outre, le gène *SELENON* n'est pas détecté dans un sous-groupe de *C. Poribacteria* vivant libre en suspension dans l'eau de mer (Podell et al., 2019) ; ceci suggère que l'activité SelenoN contribue aux interactions symbiotiques entre la bactérie et l'éponge. Compte tenu de sa distribution phylogénique limitée, il est probable que le gène SelenoN bactérien (b-SelenoN) ait été acquis à partir de l'hôte animal par un transfert horizontal de gènes, de la même manière que nombre d'autres gènes eucaryotes (Kamke et al., 2014).

2. Résultats.

2.1. Analyses bioinformatiques des gènes SELENON identifiés dans le métagénome de *C. Poribacteria.*

Les génomes de *C. Poribacteria* sont accessibles dans les bases de données, dans le cadre d'analyses métagénomiques menées sur différentes communautés microbiennes symbiotiques d'éponges, également appelées holobionte. Une recherche BLAST sur les séquences 'shotgun' du génome (WGS) correspondant à 23 souches de *C. Poribacteria,* nous a permis d'identifier 166 séquences du gène *SELENON*. Les alignements multiples des séquences ont mis en évidence des résidus conservés correspondant au site catalytique, ainsi que la conservation de la séquence d'insertion de Sec, ou motif SECIS, une structure en forme de boucle de tige importante pour la reprogrammation du codon UGA codant pour Sec. En outre, nous avons constaté que les souches de *C. Poribacteria* contiennent plusieurs copies du gène *SELENON* (allant de 1 à 17 copies par souche) et parfois organisées en répétitions en tandem, ce qui suggère que le SelenoN est abondamment exprimée dans ces bactéries. L'inspection des séquences en acides aminés de SelenoN de *C. Poribacteria* a révélé que la plupart d'entre elles sont fusionnées à différents domaines additionnels N- ou C-terminaux, homologues à : un facteur de transcription sigma-70 spécifique du stress ; un domaine de type thiorédoxine (Trx)

; ou deux domaines de fonction inconnue (UND, DUF4878) présents soit seuls soit en combinaison.

En outre, les analyses génomiques de l'environnement des gènes *SELENON* ont montré que ces gènes sont souvent inclus dans des blocs conservés de gènes (synténies), parfois organisés en opérons. En particulier, le gène *SELENON* ne présentant pas de domaine supplémentaire a toujours été trouvé associé en opéron avec un gène annoté *ROG3500*. De même, *SELENON* fusionné à un domaine de type Trx est trouvé en opéron avec un gène codant pour une autre protéine de la famille des thiorédoxines. Malheureusement, les fonctions des deux protéines codées par ces gènes liés à *SELENON* sont inconnues. Cependant, puisque ces protéines sont co-exprimées dans un système d'opéron, nous avons émis l'hypothèse qu'elles sont susceptibles de constituer des partenaires de l'activité de SelenoN. Par conséquent, les informations provenant des données bioinformatiques ont été utilisées pour amplifier les deux opérons *SELENON* à partir de l'ADN total de l'éponge.

2.2. Amplification et clonage des gènes *SELENON*.

Pour confirmer la présence de *C. Poribacteria* dans des échantillons d'éponge *A. aerophoba* obtenus du laboratoire de Marcelino Suzuki à l'Observatoire Océanologique de Banyuls-sur-Mer (OOB), de l'ADN métagénomique a été préparé et le gène de l'ARNr16S de *C. Poribacteria* a été amplifié en utilisant des amorces spécifiques issues d'études antérieures (Fieseler et al., 2004 ; Podell et al., 2019). Ensuite, en utilisant les informations de l'analyse bioinformatique du génome de *C. Poribacteria*, une copie de *SELENON* en opéron avec *ROG3500* a été amplifiée, par une approche de PCR inverse. Les gènes amplifiés ont été clonés et exprimés de manière hétérologue dans un système *E. coli* spécifique pour l'expression des sélénoprotéines (Cheng & Arnér, 2017). Une version *SELENON*-Sec sauvage ou son mutant *SELENON*-Cys ont été construits soit en tant que gènes individuels, soit en opéron avec *ROG3500*. Les conditions d'expression de la b-SelenoN recombinante ont été optimisées, et l'expression de la protéine entière a été confirmée par SDS-PAGE et western blot. La protéine a été purifiée par chromatographie d'affinité sur nickel immobilisé (IMAC), et des expériences de spectrométrie de masse ont validé l'insertion correcte de Sec dans b-SelenoN recombinante. De plus, la b-SelenoN purifiée a été utilisée pour l'immunisation de lapins afin d'obtenir des anticorps qui seront utilisés pour détecter spécifiquement l'expression de cette protéine. En parallèle, sur la base de l'alignement multiple, deux peptides conservés correspondant au site catalytique et au motif SECIS ont été conçus et utilisés pour l'immunisation afin d'obtenir des

anticorps universels dirigés contre toutes les formes de b-SelenoN codées par les multiples copies de gènes dans les différentes souches de *C. Poribacteria*.

Dans une autre étude, l'association des deux protéines SelenoN et ROG3500 dans un complexe a été étudiée par co-expression et co-purification. Une analyse par spectrométrie de masse en solution a été effectuée, mais aucune quantité significative de ROG3500 n'a été co-purifiée (pull-down), ce qui indique que les deux protéines ne forment pas un complexe stable. Néanmoins, les protéines pourraient être liées fonctionnellement en contribuant à une même voie réactionnelle.

2.3. Oragnisation des Poribactéries dans l'éponge marine *A. aerophoba*.

Différents échantillons de *A. aerophoba* ont été fraîchement collectés en mer et des coupes ont été examinées pour rechercher la présence de *C. Poribacteria* en utilisant une sonde spécifique de l'ARN 16S. Afin d'augmenter la sensibilité de la détection et de réduire le bruit de fond d'autofluorescence, nous avons utilisé la technologie d'hybridation *in-situ* par fluorescence à dépôt catalysé de rapporteur (CARD-FISH). Les *C. Poribacteria* ont été détectées, et l'étude de leur distribution a montré un enrichissement dans des zones spécifiques de l'organisation de l'éponge : une densité plus élevée a été observée dans l'apex des digitations (ectosome) par rapport à la partie centrale et médiane (mésohyle). La morphologie des *C. Poribacteria* a également été examinée par microscopie électronique à transmission (TEM). Les résultats les plus significatifs obtenus jusqu'à présent ont montré l'accumulation de *C. Poribacteria* à proximité des cellules sphéruloïdes, cellules animales spécialisées dans la synthèse et le stockage de composés alcaloïdes bromés impliqués dans les mécanismes de défense des éponges contre les prédateurs et les infections bactériennes (Wu et al., 2020). La question de savoir si les alcaloïdes sont synthétisés par la bactérie ou la partie animale de l'éponge reste ouverte, bien que des données récentes montrent une contribution importante de l'holobionte (Gutleben et al., 2019). Dans une deuxième analyse, nous avons étudié la prolifération de *C. Poribacteria* en différentes situations de stress ; nous avons comparé la dynamique de la distribution de *C. Poribacteria* dans les tissus d'éponges exposés à des dommages mécaniques ou à un stress thermique aigu et chronique par rapport à un échantillon témoin. Pour cela, des éponges collectées en mer ont été maintenues en aquarium pendant plusieurs semaines et soumises à une augmentation de la température de l'eau pendant 24 heures ou deux semaines. En parallèle, d'autres individus ont été blessés à l'aide d'un scalpel. Les données préliminaires ne confirment pas l'hypothèse dans le cas du stress thermique, puisque

le nombre de *C. Poribacteria* était significativement réduit par rapport au contrôle. L'analyse du stress lié à la blessure doit encore être évaluée.

2.4. Observation en microscope électronique du niveau d'expression de la Sélénoprotéine N.

L'expression de SelenoN dans les *C. Poribacteria* a été détectée par immuno-marquage aux particules d'or en utilisant des anticorps polyclonaux dirigés contre la b-SelenoN et observée par microscopie électronique. Les résultats préliminaires montrent une forte densité de marquage au sein de bactéries symbiotiques qui ont été précédemment identifiées comme de possibles *C. Poribacteria* sur la base de caractéristiques morphologiques, telles que la présence de membranes internes. Ce marquage est en accord avec une expression abondante de b-SelenoN dans ces bactéries, en accord avec les analyses génomiques. L'immuno-marquage utilisant des anticorps nouvellement conçus dirigés contre les peptides conservés permettra de confirmer ces données.

3. Conclusion

L'identification d'un gène homologue à SelenoN dans une bactérie unique vivant en symbiose avec une éponge marine permet d'aborder l'activité de SelenoN dans un contexte biologique et écologique original, différent de celui du tissu musculaire étudié jusqu'à présent. Cette étude devrait fournir des informations clés qui aideront à découvrir la réaction catalysée par SelenoN, une étape essentielle vers le développement d'une thérapie ciblée pour la maladie musculaire liée à des mutations chez l'homme. L'étude bioinformatique a déjà fourni des informations précieuses concernant l'organisation génomique des gènes *SELENON* chez *C. Poribacteria* et sa connexion avec d'autres gènes fonctionnellement pertinents dans le contexte procaryote. En outre, les informations sur les séquences nucléotidiques ont été utilisées pour élaborer une stratégie permettant d'amplifier une copie complète correcte de SELENON ainsi que le gène qui lui est lié, excluant l'utilisation de séquences sujettes à erreur issues des bases de données métagénomiques. En outre, l'ingénierie d'anticorps dirigés contre b-SelenoN constitue un outil important pour aborder le rôle de SelenoN dans l'interaction éponge-Poribactéria, et en particulier dans le contexte de la relation symbiotique. Dans la deuxième partie de la thèse, la détection spécifique de *C. Poribacteria* par l'approche CARD-FISH a montré une amplification significative du signal, par rapport aux sondes fluorescentes classiques. Dans l'ensemble, nos observations de la distribution des *C. Poribacteria* conduisent

à deux indications remarquables : premièrement, les *C. Poribacteria* sont abondantes dans les tissus d'*A. aerophoba*, en particulier dans des zones sélectives de l'architecture des éponges, indiquant une interaction fonctionnelle bactérie-hôte ; deuxièmement, l'immuno-marquage et la microscopie électronique ont révélé une expression active des gènes *SELENON* dans les bactéries présentent dans l'environnement des cellules sphéruloïdes. En collaboration avec l'équipe de M. Suzuki à Banyuls, nous avons démontré la possibilité de maintenir des éponges *A. aerophoba* dans des conditions standardisées et contrôlées, et la possibilité d'exposer ces organismes à des situations de stress spécifiques. Nos expériences préliminaires n'ont pas montré la prolifération de *C. Poribacteria* dans les tissus d'éponges exposés à un stress thermique, mais le niveau d'expression des gènes SELENON reste à étudier ainsi que d'autres stress, comme la blessure, qui seront également testés dans le futur.

4. Remerciements.

Un grand merci à : Laurence Despons pour l'analyse bioinformatique, IBMC ; Marcelino Suzuki l'Observatoire Océanologique de Banyuls-sur-Mer (OOB), " Laboratoire Arago ", pour la fourniture d'échantillons d'éponge et l'accueil des expériences ; Marie-line Escande, OOB, pour les expériences au microscope électronique ; Laurence Besseau et Nyree West, OOB, pour la supervision des analyses FISH ; David Pecqueur, OOB, pour les images au microscope confocal.

Liste des communications orales and posters :

- Poster « **Unveiling the function of the selenium containing protein SelenoN by identification and characterization of a bacterial SelenoN homolog »;** Nedaa Y. Mahboub[1], Laurence Despons[1], Luc Thomès[1], Odile Lecompte[2], and Alain Lescure[1] ; ED 414 "IGBMC-FMI Graduate Student Symposium 2021' , 26/04/2021; Online.
- Poster participation; **"Identification of *SELENON* orthologous genes in the symbiotic bacteria *Candidatus Poribacteria"*** Nedaa Y. Mahboub[1], Laurence Despons[1], Luc Thomès[1], Odile Lecompte[2], Marcelino Suzuki[3] and Alain Lescure[1] ; 12th International Symposium on Selenium in Biology and Medicine; **February 16-20, 2022;** Hawaii, USA. Alain Lescure attended the conference and presented.
- Présentation orale; **« Identification of a gene with muscle specific activity *SELENON* in one unique bacteria, *Candidatus Poribacteria »;*** Nedaa Y. Mahboub[1], Laurence Despons[1], Luc Thomès[1], Odile Lecompte[2], Marcelino Suzuki[3] and Alain Lescure[1] ; Séminaire de Microbiologie de Strasbourg; March 24**, 2022;** Pôle API, ESBS, Illkirch.

Publication en cours de soumission.

« *Candidatus Poribacteria* genomic mapping reveals new insight into *SELENON* gene organization » ; Nedaa Y. Mahboub[1], Laurence Despons[1], Luc Thomès[1], Odile Lecompte[2], Marcelino Suzuki[3] and Alain Lescure[1] ; Biomolecules Journal Biomolecules | An Open Access Journal from MDPI.

Nedaa Yousef MAHBOUB

Caractérisation génomique et biochimique d'un homologue bactérien de la Sélénoprotiéne N, une protéine à sélénium impliquée dans différentes dystrophies musculaires congénitales

Résumé arabe مُلخّص

السيلينون هوعبارة عن بروتين غشائي يوجد في الشبكة الإندوبلازمية داخل الخلية الحيوانية. تم ربط الطفرات في الجين المنتج لبروتين السيلينون بأمراض عضلية وراثية مختلفة تظهر في الجنس البشري في الغالب، وقد ثبت أن بروتين السيلينون يدخل في نمو العضلات وتجديدها وفي غير ذلك من العمليات الحيوية في الخلية كدوره في حماية الخلية من الأكسدة التفاعلية الضارة وتنظيم مستوى الجلكوز في الدم وإنتاج الطاقة من المايتوكوندريا. ومع ذلك ، فإن الوظيفة التحفيزية للسيلينون مُبهمة للآن. في الحقيقة، حدّد بحث المعلوماتية الحيوية جيناً متماثلاً للسيلينون في مجموعة واحدة من البكتيريا الغير مصنفة ، الكانديداتوس بورييباكتيريا. هذه البكتيريا تعيش مع الإسفنج البحري في عيشة تكافلية، ولا يمكن زراعتها خارج موطنها الأصلي بعد. ركّز مشروع الأطروحة على دراسة التحليل الجيني لسلالات مختلفة من البورييباكتيريا. لقد قمتُ خلال رسالتي باستخراج نوع واحد من بروتين السيلينون والجين المرتبط معه، ثم قمت بالتعبير عن بروتين السيلينون وتنقيته لاستخدامه في تجارب تحيلية لإيجاد وظيفة السلينون في المستقبل. في الجزء الثاني من المشروع، قمتُ بدراسة وفرة وتوزيع البورييباكتيريا في هياكل الإسفنج المضيف، الأبليزينا إيروفوبا باستخدام تقنية التهجين الفلوري، والفحص المجهري الإلكتروني. أتاحت لنا هذه الدراسات بالحصول على معلومات حول تفاعل البكتيريا مع الجسم المُضيف والمساهمة المحتملة لبروتين السيلينون في البيئة التعايشية مع الإسفنج البحري.

الكلمات الرئيسية: بروتين السيلينون ، الكانديداتوس بورييباكتيريا ، الإسفنج البحري ، التعايش التكافلي.

Nedaa Yousef MAHBOUB

Caractérisation génomique et biochimique d'un homologue bactérien de la Sélénoprotiéne N, une protéine à sélénium impliquée dans différentes dystrophies musculaires congénitales

Résumé

La sélénoprotéine N (SelenoN) est une protéine transmembranaire du réticulum endoplasmique. Des mutations dans le gène *SELENON* ont été associées à différentes formes héréditaires de maladies musculaires chez l'homme et il a été démontré que la protéine SelenoN est impliquée dans le développement et la régénération musculaire. Cependant, la fonction catalytique de SelenoN reste inconnue. Une recherche bioinformatique a permis d'identifier un gène homologue à *SELENON* dans un seul groupe de bactéries non classifiées, *Candidatus Poribacteria*. Cette bactérie est un symbiote d'éponges marines, qui n'a pas été cultivé pour l'instant. Le projet de thèse s'est concentré sur l'étude de cet orthologue bactérien de *SELENON*. L'analyse génomique de différentes souches de *C. Poribacteria* a permis d'identifier plusieurs copies du gène *SELENON*, organisées en tandem ou en opéron avec d'autres gènes. J'ai amplifié une forme de *SELENON* et son gène partenaire associé en opéron, puis j'ai exprimé et purifié la protéine SelenoN qui sera utilisée pour des analyses fonctionnelles dans le futur. Dans un second projet, j'ai étudié l'abondance et la distribution des *Poribactéries* dans les structures de l'éponge hôte, *Aplysina aerophoba*, en utilisant une technique dédiée d'hybridation *in situ* par fluorescence, ainsi que des études de microscopie électronique par transmission et de marquage immunologique à l'or. Ces études nous ont permis d'obtenir des informations sur l'interaction éponge-bactérie et sur la contribution possible de SelenoN à la symbiose.

Mots clés : Sélénoprotéine N, Candidatus Poribacteria, éponges marines, interaction symbiotique.

Résumé en anglais

Selenoprotein N (SelenoN) is an endoplasmic reticulum resident transmembrane protein. Mutations in the *SELENON* gene were linked to different inherited forms of muscular diseases in Humans and the SelenoN protein was shown to be implicated in muscle development and regeneration. However, the catalytic function of SelenoN remains elusive. Bioinformatics search identified a gene homologous to *SELENON* in one single group of unclassified bacteria, *Candidatus Poribacteria.* This bacterium is a marine sponge symbiont, yet-to-be cultured. The thesis project has focused on the study of this bacterial ortholog of *SELENON*. Genomic analysis of different *Poribacteria* strains identified multicopy of *SELENON* gene, organized in tandem repeat or in operon with other genes. During my thesis I amplified one form of the *SELENON* gene and its operon-associated gene, and then expressed and purified the SelenoN protein to be used for functional analyses in the future. In the second part of the project, I studied the abundance and distribution of *Poribacteria* in the structures of the host sponge, *Aplysina aerophoba*, using fluorescence *in situ* hybridization technique, transmission electron microscopy, and immuno-gold labeling. These studies allowed us to get information about the sponge-bacteria interaction and the possible contribution of SelenoN to the symbiosis.

Key words: Selenoprotein N, *Candidatus Poribacteria,* marine sponge, symbiotic interaction.