ORIGINAL ARTICLE

Evolutionary Aspects of Selenium Binding Protein (SBP)

Irene Dervisi[1] · Chrysanthi Valassakis[1] · Aikaterini Koletti[2] · Vassilis N. Kouvelis[3] · Emmanouil Flemetakis[2] ·
Christos A. Ouzounis[4,5] · Andreas Roussis[1]

Abstract

Selenium-binding proteins represent a ubiquitous protein family and recently SBP1 was described as a new stress response regulator in plants. SBP1 has been characterized as a methanethiol oxidase, however its exact role remains unclear. Moreover, in mammals, it is involved in the regulation of anti-carcinogenic growth and progression as well as reduction/oxidation modulation and detoxification. In this work, we delineate the functional potential of certain motifs of SBP in the context of evolutionary relationships. The phylogenetic profiling approach revealed the absence of SBP in the fungi phylum as well as in most non eukaryotic organisms. The phylogenetic tree also indicates the differentiation and evolution of characteristic SBP motifs. Main evolutionary events concern the CSSC motif for which Acidobacteria, Fungi and Archaea carry modifications. Moreover, the CC motif is harbored by some bacteria and remains conserved in Plants, while modified to CxxC in Animals. Thus, the characteristic sequence motifs of SBPs mainly appeared in Archaea and Bacteria and retained in Animals and Plants. Our results demonstrate the emergence of SBP from bacteria and most likely as a methanethiol oxidase.

Keywords Selenium · SBP · Phylogenetic profiling · Phylogenetic tree · CxxC motif · Motifs

Introduction

Selenium (Se) was discovered in 1817 by Berzelius yet it was not for another 140 years that the essentiality of this element was broadly realized (Schwarz and Foltz 1957). Selenium is an essential micronutrient for many organisms (Archaea, Bacteria, Protozoa, green algae and Animals) (Birringer et al. 2002). At low doses, selenium can promote growth in plants such as potato (Turakainen et al. 2006), ryegrass (Hartikainen et al. 2000), tea (Hu et al. 2003), rice (Liu et al. 2004) and soybean (Djanaguiraman et al. 2010).

As Sulfur and Selenium both belong to chalcogens they resemble each other chemically. In plants they share similar pathways for uptake and translocation and they are components of proteins as constituents of cysteine, methionine selenocysteine and selenomethionine (Jacob et al. 2003; Hasanuzzaman et al. 2020). However, Selenium is not able to create π bonds and the electrons of its outer valence are looser than those of sulfur (S) (Reich and Hondal 2016) allowing selenium to react with Reactive Oxygen Species (ROS) faster than sulfur. Nonetheless, the inability of creating a π-bond in Se-O allows its easier reduction compared with the respective S–O bonds (Reich and Hondal 2016). Due to this chemical property, Se participates in specific biological processes by replacing sulfur, such as in selenoproteins.

In organisms where selenium is essential, it is required for the biosynthesis of the selenoamino acid Se-Cys (SeC), used for the translation of selenoproteins. Selenoproteins have been identified in several organisms

Handling editor: **Ananias Escalante**.

✉ Andreas Roussis
aroussis@biol.uoa.gr

1 Section of Botany, Department of Biology, National & Kapodistrian University of Athens, 15784 Athens, Greece

2 Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, 11855 Athens, Greece

3 Section of Genetics and Biotechnology, Department of Biology, National & Kapodistrian University of Athens, 15784 Athens, Greece

4 Biological Computation & Process Laboratory, Centre for Research & Technology Hellas, Chemical Process & Energy Resources Institute, 54124 Thessaloníki, Greece

5 Biological Computation & Computational Biology Group, AIIA Lab, School of Informatics, Aristotle University of Thessalonica, 57001 Thessaloníki, Greece

such as mammals, bacteria, archaea and viruses, but not in plants and yeasts (Stadtman 1990). Selenoproteins perform various critical functions in redox reactions, free radical scavenging and hormone regulation (Gladyshev and Kryukov 2001; Kryukov et al. 2003; Driscoll and Copeland 2003; Kryukov and Gladyshev 2004). In land plants, where selenium is not essential, homologues of selenoproteins do not incorporate selenium and therefore have a Cys residue instead of a Se-Cys.

The Selenium Binding Protein (SBP), which does not contain Se-Cys, is most probably involved in selenium metabolism. SBP was initially isolated from mouse liver, as a cytosolic protein and named SBP56 (Bansal et al. 1989, 1990). However, recent studies revealed that SBP56 is a misannotated methanethiol oxidase (Eyice et al. 2018; Pol et al. 2018). The degree of similarity across SBPs is high among mammals (mouse, human) and plants (Arabidopsis) reaching ~ 70%. Moreover, SBP1 has been proposed as a candidate of the selenium delivery system to selenophosphate synthetase (SPS), an enzyme important for selenophosphate production and thus the formation of selenoproteins (Tobe and Mihara 2018).

Human SBP protein (*hs*SBP, SELENBP) plays an important role in key processes such as the regulation of anti-carcinogenic growth and progression, reduction/oxidation modulation, detoxification, intra Golgi protein transport (Chen et al. 2004) and proteasome degradation pathway in a Se-dependent manner. Decreased levels of human SBP1 are associated with various cancers such as ovary, lung, esophagus, colon, stomach, liver and in uterine leiomyoma (Chen et al. 2004; Zhang et al. 2010a, b; Di Stasio et al. 2011; Xia et al. 2011; Huang et al. 2012; Ansong et al. 2015; Udawela et al. 2015). Moreover, *hs*SBP1 has been associated with Behcet's disease and schizophrenia and proposed as a serological marker in immunoscreening methods (Chen et al. 2004) and a biomarker for neurological diseases associated with schizophrenia (Chen et al. 2004; Kanazawa et al. 2009; Amar et al. 2010).

Recent studies have shown that in *Arabidopsis* SBP1 participates in a protein interaction network consisting at least of SBP, a NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and a fructose-bisphosphate aldolase (FBA) (Agalou et al. 2006). Moreover, it has been shown *in planta* that SBP1 also interacts with glutaredoxins (GRXS14, GRXS16) (Valassakis et al. 2019), a phospholipase (DALL3) (Dervisi et al. 2020) and a papain-like protease (RD19c) (Dervisi et al. 2022). A major step towards understanding SBP1 function in plants and its involvement in selenium metabolism and detoxification mechanisms was the identification of the Se-binding site and the involvement of two Cys residues in *At*SBP1, as well as the function of selenite reduction (Schild et al. 2014).

The bacterial SBP56 is a Cu-dependent methanethiol oxidase (MTO), a widely distributed protein (Eyice et al. 2018). MTOs degrade methanethiol to formaldehyde, hydrogen sulfide (H_2S) and hydrogen peroxide (H_2O_2), an enzyme activity not found in humans, where H_2S and H_2O_2 are known cellular signal molecules. Low levels of SELENBP1 and subsequently reduced MTO activity may cause malodor syndrome and other diseases (Pol et al. 2018).

We have recently shown that in *Chlamydomonas reinhardtii* lack of the SBP1 (SBD1) homologue affects the molecular and biochemical responses upon oxidative stress dramatically. Moreover, interactions of SBP1 with certain proteins in *Arabidopsis* are also conserved in *C. reinhardtii*. The same study revealed that SBP1 is involved in redox early sensing and triggering of the subsequent cellular responses via protein–protein interactions (Koletti et al. 2022).

The diversity of SBPs in many different organisms with differential utilization requires a deeper examination of the evolutionary history of this protein family. The aim of the present study is to understand the functional potential of certain motifs in protein coding sequence of *At*SBP1 among different taxa. We investigate the involvement of these motifs with the function of SBP, in the context of evolutionary relationships, towards a better understanding of their potential roles. In addition, we report a phylogenetic analysis for 120 species (from archaea to mammals) to obtain valuable insights into the evolution of SBP.

Material and Methods

Data Sources and Searches

The Reference protein database of NCBI (refseq_protein) was used to retrieve all sequences analyzed in this study. Their selection was based on similarity levels (> 80% query cover and 70% identity) with the *Arabidopsis thaliana* Selenium Binding Protein 1 (UniProt Acc. No. O23264), using NCBI BLAST with default parameters (Johnson et al. 2008; Boratyn et al. 2013). In total, 129 different SBP proteins identified as homologues to *Arabidopsis thaliana* SBP belonging to representative organisms from all Domains of Life were analyzed in detail (Table S1).

For phylogenetic profiling, the latest version (2022_01) of reference proteome collection (excluding viruses) was used (Chen et al. 2011; The UniProt Consortium 2019), comprising 10,273 individual proteomes and 59,653,876 protein sequences.

Domain Identification

The structure of all Selenium Binding Proteins used in the matrix of the phylogenetic analysis was examined based on

the information provided by previous studies (Flemetakis et al. 2002; Agalou et al. 2006; Martins Alves et al. 2019). All proteins were examined for the presence of the CC, *DEL, CxxC, HxD and HxxD motifs, which comprise the putative functional domains of the protein, documented elsewhere (Flemetakis et al. 2002; Agalou et al. 2006; Schild et al. 2014). The motifs and domains presented here were edited using Jalview (Waterhouse et al. 2009) and visualized with WebLogo 3 (http://weblogo.threeplusone.com/) (Schneider and Stephens 1990; Crooks et al. 2004).

3D Modeling

The X-ray structure of the hypothetical selenium-binding protein from *Sulfolobus tokodaii* ST0059 (PDB identifier: 2ece, (Yamada et al. 2008, unpublished) was obtained from PDB (Burley et al. 2021). Analysis and visualization were performed by UCSF Chimera (Pettersen et al. 2004). Topology diagrams of the 2ece coordinates were generated with PDBsum (Laskowski 2009). Annotation and highlighting of conserved residues were based on the master alignment (Data DS1, DS4).

Sequence Alignments and Phylogenetic Trees

For phylogenetic analysis, sequences were aligned with MAFFT (Kuraku et al. 2013; Katoh et al. 2019). Alignment parameters were set to default, and the result was verified manually. Manual editing was restricted only to the N-terminus of the protein, with the constraint to include the CC motif, where possible (Data DS1).

For Bayesian inference, PartitionFinder 2 (Lanfear et al. 2012, 2016) was used to determine a phylogenetic model that best fitted the data. The best model (LG + I + G) was determined based on AICc. Four independent MCMCMC searches were performed using different random starting points (1,000,000 number of generations), with sampling every 1,000 generations. Convergence was checked visually by plotting likelihood scores vs. generation for the runs. Based on this analysis, the burn-in was set to 25%. This analysis was performed in CIPRES Science Gateway V. 3.3, a public resource for inference of large phylogenetic trees (Miller et al. 2010), visualized using iTOL (Letunic and Bork 2019) and FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and edited with Inkscape (Oualline and Oualline 2018).

Results and Discussion

Phylogenetic Profiling

*At*SBP1 was screened for homologues in more than 10,000 Uniprot proteomes with known sequences. Remarkably, only in 1,156 proteomes a SBP1 homologue was present, with 1,078 of them belonging to eukaryotic organisms. Because of some redundancy and revision issues, SBP1 homologues eventually correspond to 1,055 eukaryotic proteins. On the other hand, of the 9,117 species that lack SBP1, 1,053 are eukaryotes (Data DS2) and most of them are fungi (762), followed by 51 birds, 47 apicomplexans, 29 oomycetes and 21 kinetoplastids. Regarding the phylum of plants there are just 13 species of which 4 flowering plants with no SBP1 homologue. The potential absence of SBP homologue in those might be attributed to genome assembly or annotation issues, and not to a biological trait.

Among the detected SBP1 homologues, there are 20 proteins not characterized as selenium-binding proteins or methanethiol oxidase homologues (Upload list Data DS3 to https://www.uniprot.org/uploadlists/). This list includes some intriguing proteins, a DNA_LIGASE_A3 domain-containing protein (A0A445CL38, potentially involved in sulfur metabolism https://string-db.org/network/3818.A0A445CL38), a Rapid Alkalinization Factor (A0A1R3G161, potentially involved in cysteine and methionine metabolism and sulfur metabolism https://string-db.org/network/210143.A0A1R3G161), two *Gossypium* Adenosylhomocysteinases (A0A7J8X274, A0A7J8LSY9), a DYW_deaminase domain-containing protein (D7MH79) and a very complex Unconventional myosin-Ie (A0A3N0XSR5, http://pfam.xfam.org/protein/A0A3N0XSR5).

Adenosylhomocysteinase is a hydrolase (EC: 3.3.1.1) participating in cysteine and methionine metabolism and thus playing a crucial role in sulfur metabolism. Previous reports have indicated that bifidobacteria are unable to assimilate inorganic sulfur and consequently these bacteria need cysteine as an organic sulfur source (Hassinen et al. 1951; Ueda et al. 1983; Schell et al. 2002; Lee and O'Sullivan 2010; Ferrario et al. 2015). This requirement can be overcome by replacing cysteine with methionine. Other studies suggest that bifidobacteria could have as sulfur source sulfur-containing metabolites such as hydrogen sulfide (H_2S), methanethiol or glutathione (Schöpping et al. 2021). Considering that SBPs are methanethiol oxidases and thus convert methanethiol to formaldehyde, H_2S and H_2O_2, a modification of SBP in *Gossypium* might have occurred to achieve a more effective sulfur metabolism.

The DYW deaminase-domains are known to act in editing enzymes and are important for RNA editing and cleavage. Moreover, this domain contains the zinc-binding motif HXE and the CXXC motif, also present in SBPs and thus might be slightly misannotated.

Structural and Functional Context of Conserved Motifs

Alignment and comparison of the SBP amino acid sequences from all organisms examined in the present study, revealed the presence of conserved motifs such as CC/CXXC (Schild et al. 2014), KDEL, CSSC, HXD and HXXHC (Flemetakis et al. 2002; Agalou et al. 2006) (Fig. 1, 2). Particularly, the Se-binding motif CC in SBP1 of *Arabidopsis thaliana* (Schild et al. 2014), contains two Cys residues (Cys 21 and Cys 22) which could bind a single atom of selenium to form a R-S-Se (II)-S-R-type complex (Schild et al. 2014). The CC motif (Fig. 1, positions 18–19, Fig. 2) was found in the Plantae except for two out of three sequences of *Glycine max*, as well as in cyanobacteria and a-Proteobacteria species. Moreover, this motif was present in one from the six representatives of Anthozoa, in the representative of Porifera and Leptocardii, in one from the two Nematoda and in two out of the seven Planctomycetes (Fig. 2). This observation could be due to the fact that the CC motif has been characterized as a motif able to bind Se in photosynthetic organisms (Schild et al. 2014). In contrast, the CC motif has been substituted with CxxC (Fig. 1, 18–52, Fig. 2) in most Chordata, where it is predicted as a possible candidate for Se binding (Schild et al. 2014). On

the other hand, neither the CC motif nor the CXXC motif are present in Bacteria, Archaea and Fungi (Fig. 3, S1).

The motif [G/K/R/D/N]DEL appears as *DEL (Fig. 1, positions 130–132, Fig. 2) in all examined species except *Psyscommitrella* (the only one in Bryophyta) and *Amphideon* (the only one in Sponges) which have a substitution (*DEI), while some Fungi contain *SEV (Fig. 3, Table S2). This motif is an endoplasmic reticulum (ER) retention signal of many soluble proteins contained in the cisternal lumen in eukaryotic cells (Stornaiuolo et al. 2003).

The CxxC (CSSC) motif (Fig. 1, positions 140–144, Fig. 2) appears conserved in all species except *Pelodiscus sinensis* (the only one in Testudines) which has a CSSS substitution and *Galdieria sulphuraria* (the only one in Rhodophyta) CCSS (Table S2). In the clade of Archaea (Halobacteria) it is present as CSSSC (Table S2). The CXXC motif with cysteine residues, is employed by many redox proteins for the formation, isomerization, and reduction of disulfide bonds or other redox functions. This sequence is referred to as thioredoxin-like domain (Edman et al. 1985; Kimura et al. 2004). Recent studies revealed the presence of many natural homologues of CxxC-containing proteins, in which the C-terminal Cys in the CxxC motif, is replaced with serine (the CxxS motif) (Fomenko and Gladyshev 2002). Some of these enzymes are biochemically characterized revealing an expanded
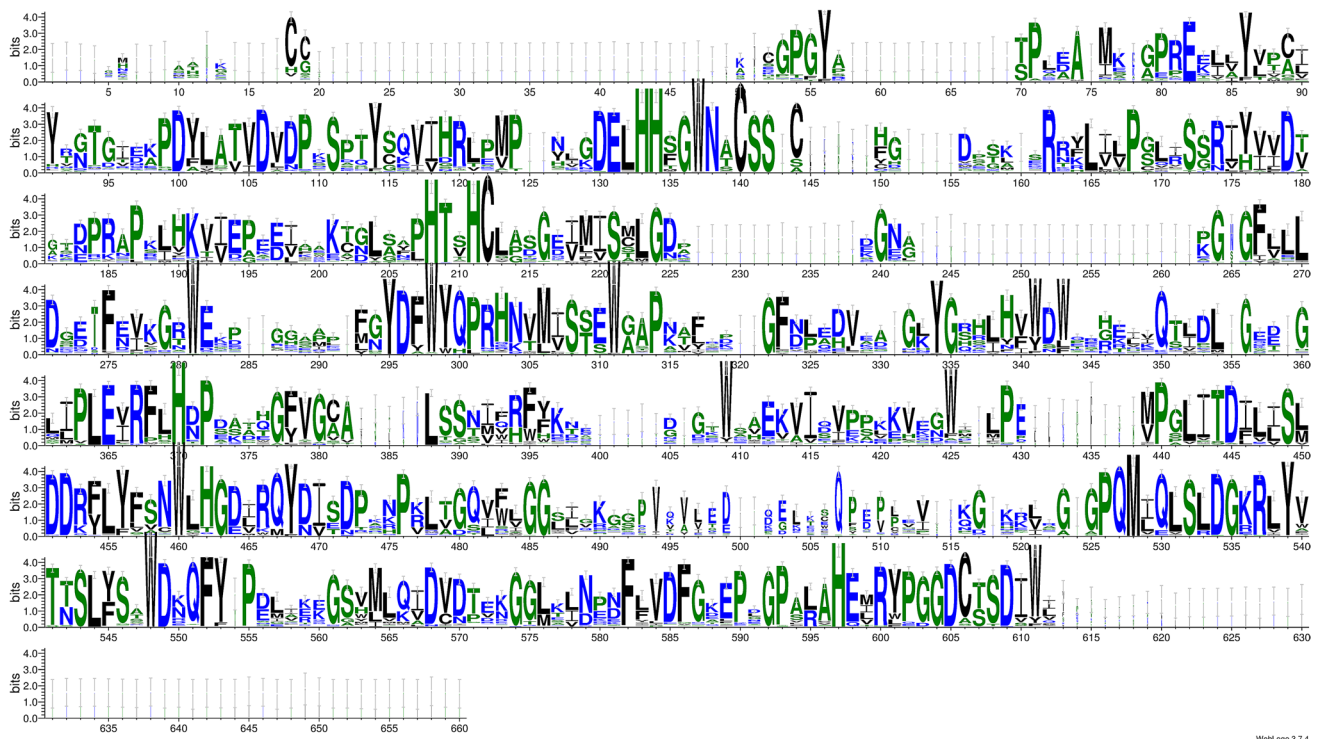

Fig. 1 Alignment Logo. Multiple sequence alignment was done using MAFFT and after editing was upload to the online tool WebLogo (http://weblogo.threeplusone.com/create.cgi)

repertoire of redox functions, such as the role of CxxS-containing proteins in methionine sulfoxide reduction (Kumar et al. 2002) and protein retention through transient formation of intermolecular disulfide bonds (Anelli et al. 2002). Genomic analyses determined that the CxxS motifs are highly conserved and present in structurally distinct proteins, suggesting that CxxS is a new fold-independent redox motif (Fomenko and Gladyshev 2002, 2003). Recently, it was shown that the CXXC motif of the branched-chain aminotransferase (BCAT1)–a protein which participates in myeloid leukaemia development–is crucial for the redox homeostasis and redox-mediated cellular processes having a novel antioxidant role. Particularly, CXXC motif can reduce H_2O_2 and utilize reducing equivalents from NADPH in this process (Hillier et al. 2022). These characteristics are consistent with the hypothesis that SBP participates in antioxidant processes and is connected with several diseases. Moreover, in silico analysis of *Theobroma cacao* SBP homologue showed interaction between CSSC motif and selenite, whereas no interaction was found with selenate (Martins Alves et al. 2019).

In addition, the putative metal binding motif with histidine residues HxD appears in two different parts of the SBP protein sequence (Fig. 1, positions 150–156 and 462–464, Fig. 2), while altered in some representative species. Particularly, the HXD in positions 150–156 of the alignment was observed only in α-Proteobacteria, Cyanobacteria, Chloroflexi, Aves and Chondrichthyes and most Plants, whereas the HXD in positions 462–464 is more conserved and absent in one Chondrichthyes, some Anthozoa, Fungi, Acidobacteria and modified in Archaea (HXE, HXN) and *Chlamydomonas reinhardtii* (RXD) (Table S2). The HxD motif is one of the highly conserved structural components of the catalytic core of protein kinases which are critical for redox signaling (Zhang et al. 2015). Furthermore, HXD motif is known as a metal-binding motif and in Tet-eleven translocation (Tet) family can bind Fe (II) and act as cofactor, as well as Ni (II) and Cd (II), with higher affinity to Ni (II) affecting the activity of Tet-mediated DNA hydroxymethylation (Yin et al. 2017).

Another highly conserved metal ion coordination motif, is the HXXH which is followed by a Cys residue creating the HXXHC motif (Fig. 1, positions 218–222, Fig. 2). This motif is present in all protein sequences used in this analysis with the exception of the outgroup (Fig. 2, Table S2). In some cases, such as the Characterising N-acetylglucosaminylphosphatidylinositol de-N-acetylase (CaGpi12) of *Candida albicans* and in the 3′-Phosphoadenosine-5′-phosphosulfate (PAPS) of human, it has been considered as important motif for the function of these enzymes by achieving binding of the substrates

(Yadav et al. 2018; Zhang et al. 2022). Therefore, we hypothesize the plausible participation of this motif in the binding of SBP substrates.

Another intriguing protein trait is the presence of two clathrin-binding boxes in positions 446–451 and 582–587 (Fig. 1). The clathrin binding box consists of pLφpφp, where φ is a bulky hydrophobic amino acid and p any amino polar residue. This motif is present in clathrin binding proteins such as AP-2, AP180, amphiphysin, epsin and arrestin2 (Kang et al. 2009). The pLφpφp motif of SBP sequences may correlate with a previous observation of SBP linked to membrane trafficking functions (Porat et al. 2000; Agalou et al. 2006).

The above motifs that have been defined for SBP1 can be further understood within the structural context of the hypothetical selenium-binding protein of *S. tokodaii* (Fig. 2). The most conserved motifs are clustered away from the core of the molecule which is defined by a beta-propeller topology (Fig. 2). The KDEL motif is located at the end of beta strand three (residues 59–66) followed by a number of invariant positions that lead into beta strand four (positions 77–79) (Fig. 2, bottom panel). The CSSC/HxD motifs are all located prior to strand five (positions 96–102), thus not participating in the structural core of the beta propeller. Remarkably, the HxxHC motif is the only conserved motif that is located on strand eight (positions 137–146) and possesses two of the six invariant histidines (Data DS4) five of which are in close spatial proximity (positions 133, 134, 208, 211, 597 and 74, 75, 141, 144, 445, in Fig. 1 and 2ece Fig. 2, respectively). The sixth invariant histidine is located at position 370 (Fig. 1, position 260 in Fig. 2). The functional significance of this arrangement is not understood at present. In contrast to the conserved motifs mentioned above, the remaining invariant residues appear to contribute to the structural stability of the molecule, most likely performing helix-breaking (positions 214, 226) or strand-forming (positions 365–401) roles (Fig. 2, bottom panel).

Vertical and Potential Horizontal Inheritance Across the Tree of Life

To explore the evolutionary relationship among SBP proteins, a multiple sequence alignment of protein sequences from different species was performed using MAFFT and a phylogenetic tree was generated with MrBayes software using as an outgroup the methanethiol oxidase of *Hyphomicrobium sp* (Fig. 3, Fig. S1). The identity levels between the query sequence from *Arabidopsis thaliana* and all reported homologs as well as their lengths in amino acid residues are shown as frequency distribution diagrams (Fig. S2), with average values for sequence identity 58.2% and length of 479 residues. All the members of the SBP family were divided into four major clades Archaea (Halobacteria),
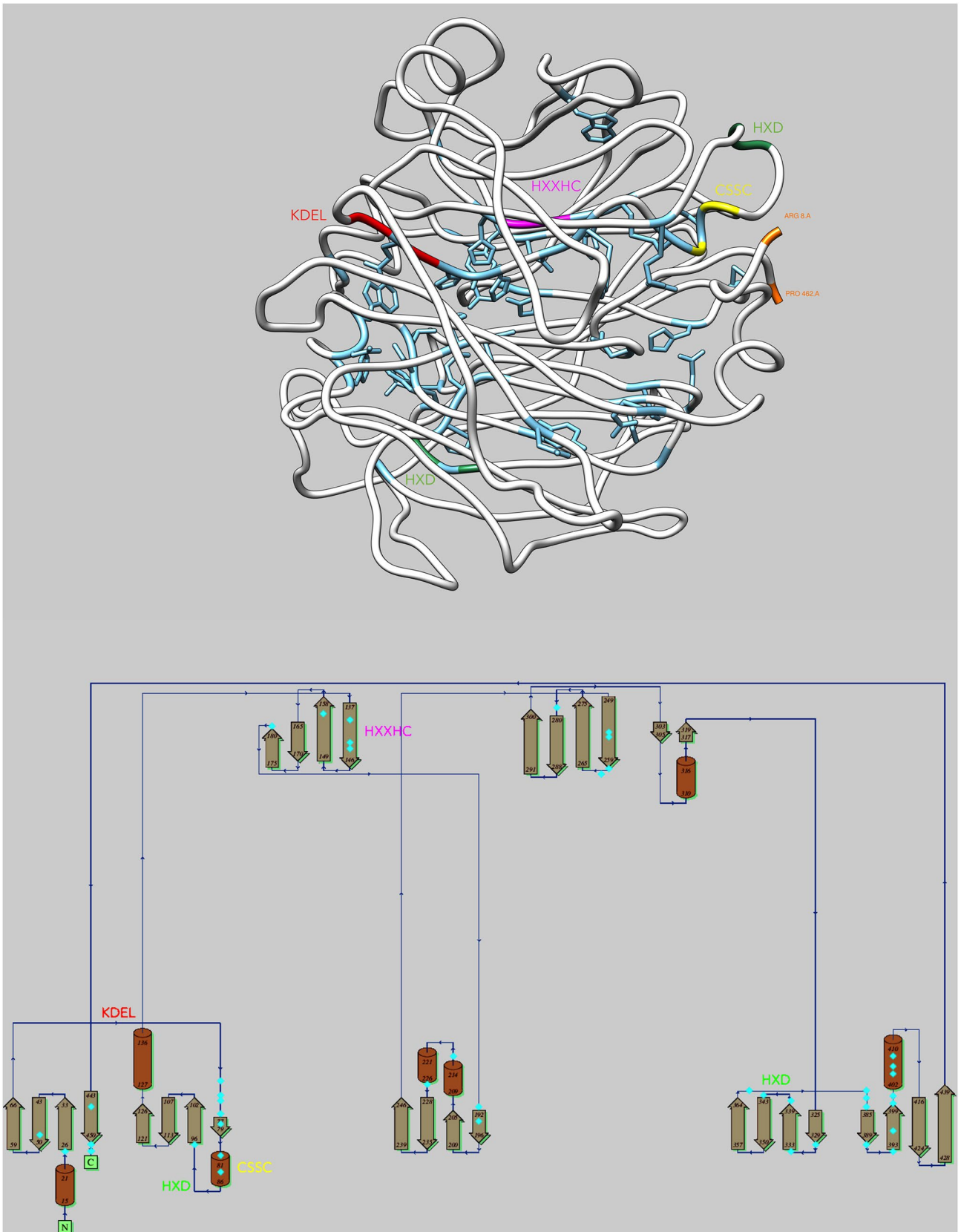
◀**Fig. 2** Structure of *S. tokodaii* SBP homolog (PDB cite: 2ece). Top panel: three-dimensional model for 2ece (ribbon, in gray) with key motifs shown in their structural context (KDEL in red, CSSC in yellow, HxD in green and HxxHC in magenta). The N- and C-terminal residues are also shown (top right, in orange). From the alignment of 129 representative sequences (see Methods), 43 invariant (100% identical) positions are selected, out of 454 residues available in the SBP structural model, or ∼10% of total. Side chains of the invariant residues are also shown (main-chain and side-chain in cyan). Model generated by Chimera. Bottom panel: two-dimensional topology diagram for 2ece (strands in olive green, helices in brown). The key motifs are also shown as in the top panel. Invariant residues are approximately located (bright cyan spots) to indicate their relative position within the secondary structure elements. Residue numbers are also provided. Model generated by PDBsum

Bacteria (Proteobacteria, Chloroflexi, Cyanobacteria, Planctomycetes), Plants, Animalia (Mammalia, Reptilia, Amphibia, Aves, Telesostei, Leptocardii, Chondrichthyes, Porifera, Nematoda, Anthozoa). Interestingly, SBP proteins with similar domain architecture were clustered in the same clades. Moreover, the architecture of the SBP tree seems to follow the evolutionary pathway from Archaea to Protista and Bacteria, then to Plants and finally to Animalia. In more detail, there is a clade at the base of the tree with Fungi and Acidobacteria and a member of Archaea (*Sulfurisphaera tokodai*).

The atypical cluster of Fungi and acidobacteria might represent a case of lateral gene transfer (LGT) between these two phyla. Fungal SBP homologs were found only in Ascomycota and in close association with acidobacteria exhibiting high similarity (identity ∼55%). Apart from the species of Ascomycota presented in the phylogenetic tree (Fig. 3), SBP homologs were detected only in *Cladophialophora* sp., *Neofusicoccum parvum* and *Fusarium decemcellulare*, also members of Ascomycota. Acidobacteria and Ascomycota are known to form microbial communities in the soil and participate in carbon and nitrogen cycles (Challacombe et al. 2019; Kalam et al. 2020). Moreover, recent observations suggest that members of both taxa are abundant in contaminated acidic soils (Liu et al. 2022) suggesting a coordinated biochemistry. Thus, the co-occurrence of Acidobacteria and Ascomycota could have led to LGT that explains the presence of SBP homologs in the fungi.

The group of Archaea is well based and it is followed by a group of the three out of seven Planctomycetes as a base for all the other clades. The other four Planctomycetes are grouped with the other Bacteria and as a base to this clade which is comprised of α-Proteobacteria, Cyanobacteria, Chlroroflexi and Rhodophyta (Protista). The clade of Plants is following with *Chlamydomonas reinhardtii* as the base and the clade of Animals with Nematoda as base. All the Bayesian bootstrap values ranged from 54–100% (Fig. S1).

All the members of Archaea (TACK group and Halobacteriota) clustered reliably, except *Sulfurisphaera tokodaii* (*Sulfolobus tokodaii*) which is grouped with the mixed clade of fungi and acidobacteria. We note an absence of detectable SBP homologs in the DPANN group and well-established methanogenic clades of Archaea, such as Methanococcales and Methanosarcinales. The relative abundance of SBP homologs in certain clades of Archaea with the aforementioned exceptions implies a potential LGT or even a massive gene loss of SBP homologs, difficult to distinguish at such scale for deep phylogenies.

Protista appear between different groups of Bacteria (Chloroflexi and Planctomycetes) and are represented by one member of Rhodophyta (sp. *Galdieria*). Bacteria are divided into four well supported groups except *Thermobaculum terrenum* which was clustered away from the Bacteria lineage and Acidobacteria which is grouped with Fungi. The four groups are Chloroflexi with three members *Ktedonosporobacter rubrisoli*, *Dictyobacter formicarum* and *Reticulibacter mediterranei,* Cyanobacteria with five members, α-Proteobacteria with ten representatives and Planctomycetes with seven.

The largest clade of the phylogenetic tree was divided in two large groups, the one group included the representatives of Plants and the other group the members of Animalia. In Plants, *Chlamydomonas reinhardtii* (Chlorophyta) is located at the base of this clade followed by *Physcomitrium patens* (Bryophyta) and *Selaginella moellendorffii* (Lycopodiopsida). In Animalia six groups are formed: One group with Mammals, Aves and Reptilia, second group with Teleostei, Leptocardii and Chondrichthyes, third group comprised by Amphibia, fourth with Anthozoa, fifth with Porifera and sixth with Nematoda.

The phylogenetic tree also indicates the differentiation and evolution of the characteristic SBP motifs. In more detail, the HxxHC is conserved in all representative sequences used, whereas the *DEL motif is always present except for fungi. Archaea is in the base of Bacteria, Plantae and Animalia while Acidobacteria and Fungi are basal to Archaea. The second HxD metal motif is absent from Fungi and Acidobacteria and appears with modifications in Archaea and in Planctomycetes, so the formation of this motif took place in Archaea and Planctomycetes and was maintained in the other phylogenetic groups. Likewise, evolutionary events also concern the CSSC motif where the basal clades (Acidobacteria, Fungi and Archaea) carry modifications. Moreover, the CC motif is harbored by some bacteria (α-Proteobacteria and Cyanobacteria) and remains conserved in Plants, whereas modified to CxxC in Animals. Thus, the characteristic sequence motifs of SBPs mainly appeared in Archaea and Bacteria and retained in Animals and Plants.
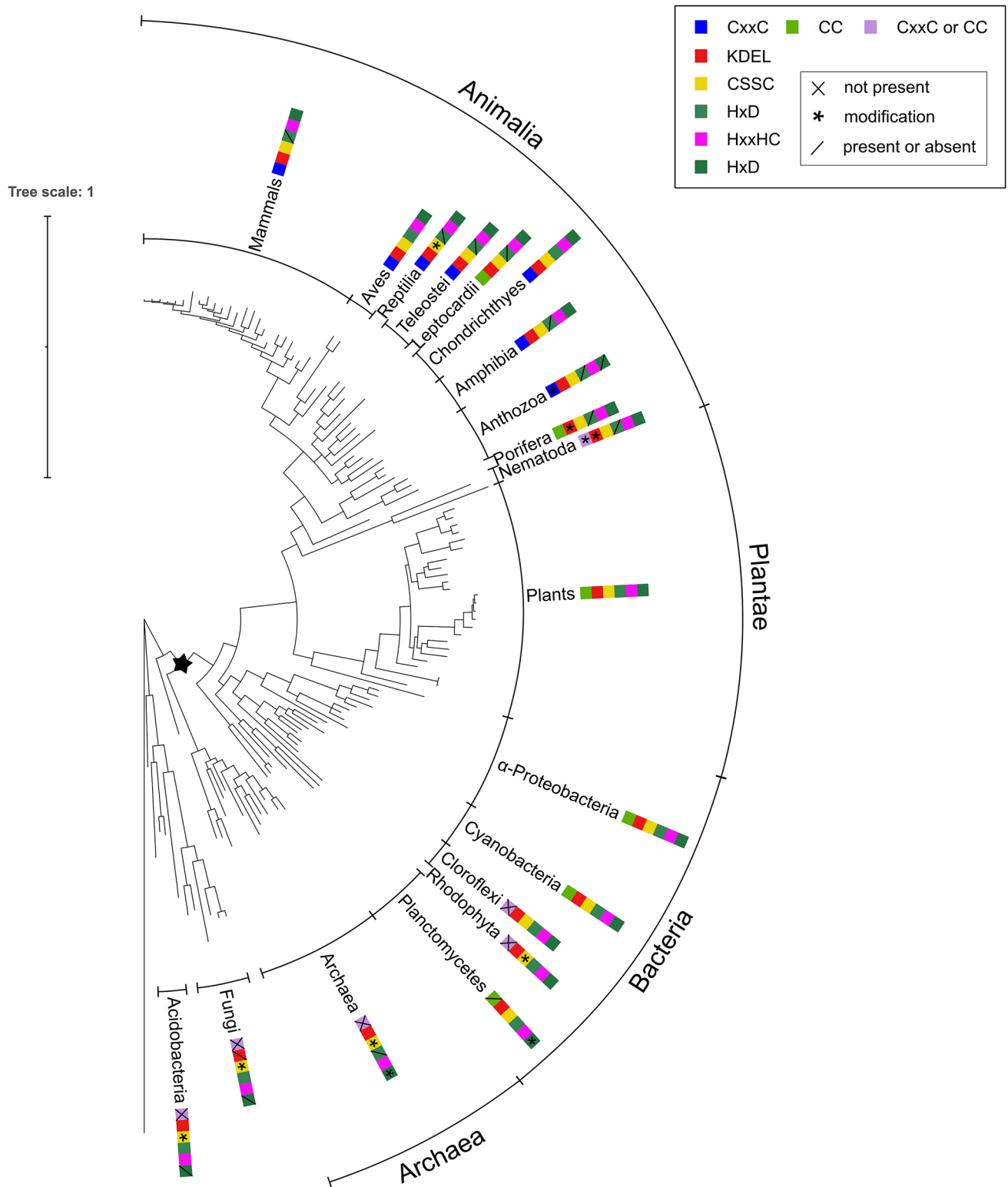
Fig. 3 Phylogenetic relationships of SBP proteins among different species. The phylogenetic tree was generated using MrBayes software and the bacterial MTO as outgroup. iTOL was used for presentation. Sequence motifs presence is pictured by different color boxes as motifs appeared in sequences. Asterisk depicts the consistent phylogeny of Bacteria, Plants, Animals with the limited number of archaeal SBP homologs, mostly from TACK group and Halobacteriota (Euryarchaeota), as well as the anomalous distribution of homologs in a narrow range of fungal species in close association with acidobacteria

It is worth pointing out that out of the four major clades reported herein, the two eukaryotic taxa of animals and plants as well as the bacterial domain exhibit a phylogenetically consistent distribution pattern for SBP homologs. In contrast, the fourth category that includes a section of Archaea alongside some Fungi and Acidobacteria indicates an anomalous phylogenetic history that might include lateral gene transfer events and/or losses. This pattern (Fig. 2) is a structural perspective based on sequence phylogenies and does not reflect any known functional diversification of SBPs, as a precise molecular function is limited to just a few instances, such as *A. thaliana*, *C. elegans* and *H. sapiens*. The *C. reinhardtii* SBP homolog was recently characterized as a novel stress sensor (Koletti et al. 2022), while the human homolog plays crucial role in cancer development (Bansal et al. 1990; Xia et al. 2011; Huang et al. 2012; Ansong et al. 2015). However, the only known function is based on bacterial SBP56 which is characterized as a Cu-dependent methanethiol oxidase (MTO) (Eyice et al. 2018).

In conclusion, the exact role of Selenium Binding Proteins remains an open question, yet our analysis demonstrates their emergence from bacteria and most likely as methanethiol oxidases. The similarity degree of SBPs is high and the level of conservation is comparable to other protein families like histones, actin, ubiquitin, γ-tubulin (49%), Heat Shock (44–75%), 14-3-3 proteins and eIF4E (Flemetakis et al. 2002; Agalou et al. 2006). Interestingly, the absence of SBPs in yeast from our phylogenetic profiling analysis correlates with a study where another redox related protein, the Quiescin Sulfhydryl Oxidase (QSOX), is also absent from fungi, indicating a common diversified evolutionary pathway and differentiation in time of redox response pathways in this taxon (Limor-Waisberg et al. 2013).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00239-023-10105-4.

Data Availability Sequence, structure and taxonomic information presented in this study (Data DS1-DS5) is available on FigShare at https://doi.org/10.6084/m9.figshare.22188193 to ensure reproducibility. Additional supplementary material is available on the journal website.

Declarations

Conflict of interest The authors declare no competing interests.

References

Agalou A, Spaink HP, Roussis A (2006) Novel interaction of selenium-binding protein with glyceraldehyde-3-phosphate dehydrogenase and fructose-bisphosphate aldolase of *Arabidopsis thaliana*. Funct Plant Biol 33:847–856. https://doi.org/10.1071/FP05312

Amar S, Ovadia O, Maier W et al (2010) Copy number variation of the SELENBP1 gene in schizophrenia. Behav Brain Funct 6:10

Anelli T, Alessio M, Mezghrani A et al (2002) ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. EMBO J 21:835–844. https://doi.org/10.1093/emboj/21.4.835

Ansong E, Ying Q, Ekoue DN et al (2015) Evidence that selenium binding protein 1 is a tumor suppressor in prostate cancer. PLoS ONE 10:e0127295. https://doi.org/10.1371/journal.pone.0127295

Bansal MP, Cook RG, Danielson KG, Medina D (1989) A 14-kilodalton selenium-binding protein in mouse liver is fatty acid-binding protein. J Biol Chem 254:13780–13784

Bansal MP, Mukhopadhyay T, Scott J et al (1990) DNA sequencing of a mouse liver protein that binds selenium: implications for selenium's mechanism of action in cancer prevention. Carcinogenesis 11:2071–2073

Birringer M, Pilawa S, Flohé L (2002) Trends in selenium biochemistry. Nat Prod Rep 19:693–718. https://doi.org/10.1039/B205802M

Boratyn GM, Camacho C, Cooper PS et al (2013) BLAST: a more efficient report with usability improvements. Nucleic Acids Res 41:W29–W33. https://doi.org/10.1093/nar/gkt282

Burley SK, Bhikadiya C, Bi C et al (2021) RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Res 49:D437–D451. https://doi.org/10.1093/nar/gkaa1038

Challacombe JF, Hesse CN, Bramer LM et al (2019) Genomes and secretomes of *Ascomycota* fungi reveal diverse functions in plant biomass decomposition and pathogenesis. BMC Genomics 20:976. https://doi.org/10.1186/s12864-019-6358-x

Chen G, Wang H, Miller CT et al (2004) Reduced selenium-binding protein 1 expression is associated with poor outcome in lung adenocarcinomas. J Pathol 202:321–329. https://doi.org/10.1002/path.1524

Chen G, Snyder CL, Greer MS, Weselake RJ (2011) Biology and biochemistry of plant phospholipases. Crit Rev Plant Sci 30:239–258. https://doi.org/10.1080/07352689.2011.572033

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator: figure 1. Genome Res 14:1188–1190. https://doi.org/10.1101/gr.849004

Dervisi I, Valassakis C, Agalou A et al (2020) Investigation of the interaction of DAD1-LIKE LIPASE 3 (DALL3) with selenium

binding protein 1 (SBP1) in *Arabidopsis thaliana*. Plant Sci 291:110357. https://doi.org/10.1016/j.plantsci.2019.110357

Dervisi I, Haralampidis K, Roussis A (2022) Investigation of the interaction of a papain-like cysteine protease (RD19c) with selenium-binding protein 1 (SBP1) in *Arabidopsis thaliana*. Plant Sci 315:111157. https://doi.org/10.1016/j.plantsci.2021.111157

Di Stasio M, Volpe MG, Colonna G et al (2011) A possible predictive marker of progression for hepatocellular carcinoma. Oncol Lett 2:1247–1251. https://doi.org/10.3892/ol.2011.378

Djanaguiraman M, Sheeba JA, Devi DD et al (2010) Nitrophenolates spray can alter boll abscission rate in cotton through enhanced peroxidase activity and increased ascorbate and phenolics levels. J Plant Physiol 167:1–9. https://doi.org/10.1016/j.jplph.2009.05.018

Driscoll DM, Copeland PR (2003) Mechanism and regulation of selenoprotein synthesis. Annu Rev Nutr 23:17–40. https://doi.org/10.1146/annurev.nutr.23.011702.073318

Edman JC, Ellis L, Blacher RW et al (1985) Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin. Nature 317:267–270

Eyice Ö, Myronova N, Pol A et al (2018) Bacterial SBP56 identified as a Cu-dependent methanethiol oxidase widely distributed in the biosphere. ISME J 12:145–160. https://doi.org/10.1038/ismej.2017.148

Ferrario C, Duranti S, Milani C et al (2015) Exploring amino acid auxotrophy in bifidobacterium bifidum PRL2010. Front Microbiol. https://doi.org/10.3389/fmicb.2015.01331

Flemetakis E, Agalou A, Kavroulakis N et al (2002) *Lotus japonicus* gene Ljsbp is highly conserved among plants and animals and encodes a homologue to the mammalian Selenium-Binding Proteins. Mol Plant-Microbe Interact 15:313–322. https://doi.org/10.1094/MPMI.2002.15.4.313

Fomenko DE, Gladyshev VN (2002) CxxS: Fold-independent redox motif revealed by genome-wide searches for thiol/disulfide oxidoreductase function. Protein Sci 11:2285–2296. https://doi.org/10.1110/ps.0218302

Fomenko DE, Gladyshev VN (2003) Identity and Functions of CxxC-Derived Motifs. Biochemistry 42:11214–11225. https://doi.org/10.1021/bi034459s

Gladyshev VN, Kryukov GV (2001) Evolution of selenocysteine-containing proteins: Significance of identification and functional characterization of selenoproteins. BioFactors 14:87–92. https://doi.org/10.1002/biof.5520140112

Hartikainen H, Xue T, Piironen V (2000) Selenium as an anti-oxidant and pro-oxidant in ryegrass. Plant Soil 225:193–200

Hasanuzzaman M, Bhuyan MHMB, Raza A et al (2020) Selenium in plants: boon or bane? Environ Exp Bot 178:104170. https://doi.org/10.1016/j.envexpbot.2020.104170

Hassinen JB, Durbin GT, Tomarelli RM, Bernhart FW (1951) The minimal nutritional requirements of Lactobacillus bifidus. J Bacteriol 62:771–777. https://doi.org/10.1128/jb.62.6.771-777.1951

Hillier J, Allcott GJ, Guest LA et al (2022) The BCAT1 CXXC motif provides protection against ROS in acute myeloid leukaemia cells. Antioxidants 11:683. https://doi.org/10.3390/antiox11040683

Hu Q, Xu J, Pang G (2003) Effect of selenium on the yield and quality of green tea leaves harvested in early spring. J Agric Food Chem 51:3379–3381. https://doi.org/10.1021/jf0341417

Huang C, Ding G, Gu C et al (2012) Decreased selenium-binding protein 1 enhances glutathione peroxidase 1 activity and downregulates HIF-1α to promote hepatocellular carcinoma invasiveness. Clin Cancer Res 18:3042–3053

Jacob C, Giles GI, Giles NM, Sies H (2003) Sulfur and selenium: the role of oxidation state in protein structure and function. Angew Chem Int Ed 42:4742–4758. https://doi.org/10.1002/anie.200300573

Johnson M, Zaretskaya I, Raytselis Y et al (2008) NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5–W9. https://doi.org/10.1093/nar/gkn201

Kalam S, Basu A, Ahmad I et al (2020) Recent understanding of soil acidobacteria and their ecological significance: a critical review. Front Microbiol 11:580024. https://doi.org/10.3389/fmicb.2020.580024

Kanazawa T, Glatt SJ, Faraone SV et al (2009) Family-based association study of SELENBP1 in schizophrenia. Schizophr Res 113:268–272. https://doi.org/10.1016/j.schres.2009.06.011

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20:1160–1166. https://doi.org/10.1093/bib/bbx108

Kimura T, Nishida A, Ohara N et al (2004) Functional analysis of the CXXC motif using phage antibodies that cross-react with protein disulphide-isomerase family proteins. Biochem J 382:169–176. https://doi.org/10.1042/BJ20040116

Koletti A, Dervisi I, Kalloniati C et al (2022) Selenium-binding protein 1 (SBD1): a stress response regulator in *Chlamydomonas reinhardtii*. Plant Physiol. https://doi.org/10.1093/plphys/kiac230

Kryukov GV, Gladyshev VN (2004) The prokaryotic selenoproteome. EMBO Rep 5:538–543. https://doi.org/10.1038/sj.embor.7400126

Kryukov GV, Castellano S, Novoselov SV et al (2003) Characterization of mammalian selenoproteomes. Science 300:1439–1443. https://doi.org/10.1126/science.1083516

Kumar RA, Koc A, Cerny RL, Gladyshev VN (2002) Reaction mechanism, evolutionary analysis, and role of zinc in drosophila methionine-R-sulfoxide reductase. J Biol Chem 277:37527–37535. https://doi.org/10.1074/jbc.M203496200

Kuraku S, Zmasek CM, Nishimura O, Katoh K (2013) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res 41:W22–W28. https://doi.org/10.1093/nar/gkt389

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695–1701. https://doi.org/10.1093/molbev/mss020

Lanfear R, Frandsen PB, Wright AM et al (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol and Evol. https://doi.org/10.1093/molbev/msw260

Laskowski RA (2009) PDBsum new things. Nucleic Acids Res 37:D355–D359. https://doi.org/10.1093/nar/gkn860

Lee J-H, O'Sullivan DJ (2010) Genomic insights into Bifidobacteria. Microbiol Mol Biol Rev 74:378–416. https://doi.org/10.1128/MMBR.00004-10

Letunic I, Bork P (2019) Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239

Limor-Waisberg K, Ben-Dor S, Fass D (2013) Diversification of Quiescin sulfhydryl oxidase in a preserved framework for redox relay. BMC Evol Biol 13:70. https://doi.org/10.1186/1471-2148-13-70

Liu Q, Wang DJ, Jiang XJ, Cao ZH (2004) Effects of the interactions between selenium and phosphorus on the growth and selenium accumulation in rice (*Oryza Sativa*). Environ Geochem Health 26:325–330. https://doi.org/10.1023/B:EGAH.0000039597.75201.57

Liu H, Yang Y, Yang Y et al (2022) Dynamics of fungal and bacterial communities in different types of soil ageing with different dosages of cadmium. Ecotoxicol Environ Saf 242:113860. https://doi.org/10.1016/j.ecoenv.2022.113860

Martins Alves AM, Pereira Menezes S, Matos Lima E et al (2019) The selenium-binding protein of *Theobroma cacao*: a thermostable

protein involved in the witches' broom disease resistance. Plant Physiol Biochem 142:472–481. https://doi.org/10.1016/j.plaphy.2019.08.005

Oualline S, Oualline G (2018) Drawing with Inkscape. Practical Free Alternatives to Commercial Software. Apress, Berkeley, CA, pp 187–219

Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084

Pol A, Renkema GH, Tangerman A et al (2018) Mutations in SELENBP1, encoding a novel human methanethiol oxidase, cause extraoral halitosis. Nat Genet 50:120–129. https://doi.org/10.1038/s41588-017-0006-7

Porat A, Sagiv Y, Elazar Z (2000) A 56-kDa selenium-binding protein participates in intra-golgi protein transport. J Biol Chem 275:14457–14465. https://doi.org/10.1074/jbc.275.19.14457

Reich HJ, Hondal RJ (2016) Why nature chose selenium. ACS Chem Biol 11:821–841. https://doi.org/10.1021/acschembio.6b00031

Schell MA, Karmirantzou M, Snel B et al (2002) The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. Proc Natl Acad Sci USA 99:14422–14427. https://doi.org/10.1073/pnas.212527599

Schild F, Kieffer-Jaquinod S, Palencia A et al (2014) Biochemical and biophysical characterization of the selenium-binding and reducing site in *Arabidopsis thaliana* homologue to mammals selenium-binding protein 1. J Biol Chem 289:31765–31776. https://doi.org/10.1074/jbc.M114.571208

Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucl Acids Res 18:6097–6100. https://doi.org/10.1093/nar/18.20.6097

Schöpping M, Gaspar P, Neves AR et al (2021) Identifying the essential nutritional requirements of the probiotic bacteria Bifidobacterium animalis and Bifidobacterium longum through genome-scale modeling. Syst Biol Appl 7:47. https://doi.org/10.1038/s41540-021-00207-4

Schwarz K, Foltz CM (1957) Selenium as an integral part of factor 3 against dietary necrotic liver degradation. J Am Chem Soc 79:3292–3293. https://doi.org/10.1021/ja01569a087

Stadtman ER (1990) Metal ion-catalyzed oxidation of proteins: biochemical mechanism and biological consequences. Free Radic Biol Med 9:315–325. https://doi.org/10.1016/0891-5849(90)90006-5

Stornaiuolo M, Lotti LV, Borgese N et al (2003) KDEL and KKXX retrieval signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment, and golgi complex. Mol Biol Cell 14:889–902. https://doi.org/10.1091/mbc.e02-08-0468

The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049

Tobe R, Mihara H (2018) Delivery of selenium to selenophosphate synthetase for selenoprotein biosynthesis. Biochim Biophys Acta Gen Subj 1862:2433–2440

Turakainen M, Hartikainen H, Ekholm P, Seppänen MM (2006) Distribution of selenium in different biochemical fractions and raw darkening degree of potato (*Solanum tuberosum* L.) tubers supplemented with selenate. J Agric Food Chem 54:8617–8622. https://doi.org/10.1021/jf0613987

Udawela M, Money TT, Neo J et al (2015) SELENBP1 expression in the prefrontal cortex of subjects with schizophrenia. Transl Psychiatry 5:e615–e615. https://doi.org/10.1038/tp.2015.108

Ueda M, Nakamoto S, Nakai R, Takagi A (1983) Establishment of a defined minimal medium and isolation of auxotrophic mutants for *Bifidobacterium bifidum* ES 5. J Gen Appl Microbiol 29:103–114. https://doi.org/10.2323/jgam.29.103

Valassakis C, Dervisi I, Agalou A et al (2019) Novel interactions of selenium binding protein family with the PICOT containing proteins AtGRXS14 and AtGRXS16 in *Arabidopsis thaliana*. Plant Sci 281:102–112. https://doi.org/10.1016/j.plantsci.2019.01.021

Waterhouse AM, Procter JB, Martin DMA et al (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Xia Y-J, Ma Y-Y, He X-J et al (2011) Suppression of selenium-binding protein 1 in gastric cancer is associated with poor survival. Hum Pathol 42:1620–1628. https://doi.org/10.1016/j.humpath.2011.01.008

Yadav U, Rai TK, Sethi SC et al (2018) Characterising N-acetylglucosaminylphosphatidylinositol de-N-acetylase (CaGpi12), the enzyme that catalyses the second step of GPI biosynthesis in Candida albicans. FEMS Yeast Res. https://doi.org/10.1093/femsyr/foy067

Yamada M, Yoshida H, Kuramitsu S, Kamitori S (2008) X-ray structure of hypothetical selenium-binding protein from Sulfolobus tokodaii, ST0059. https://doi.org/10.2210/pdb2ECE/pdb

Yin R, Mo J, Dai J, Wang H (2017) Nickel(II) inhibits tet-mediated 5-methylcytosine oxidation by high affinity displacement of the cofactor Iron(II). ACS Chem Biol 12:1494–1498. https://doi.org/10.1021/acschembio.7b00261

Zhang C, Wang YE, Zhang P et al (2010a) Progressive loss of selenium-binding protein 1 expression correlates with increasing epithelial proliferation and papillary complexity in ovarian serous borderline tumor and low-grade serous carcinoma. Hum Pathol 41:255–261. https://doi.org/10.1016/j.humpath.2009.07.019

Zhang H, Zhang X, Clark E et al (2010b) TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. Cell Res 20:1390–1393. https://doi.org/10.1038/cr.2010.156

Zhang L, Wang J-C, Hou L et al (2015) Functional role of histidine in the conserved His-x-Asp motif in the catalytic core of protein kinases. Sci Rep 5:10115. https://doi.org/10.1038/srep10115

Zhang P, Zhang L, Hou Z et al (2022) Structural basis for the substrate recognition mechanism of ATP-sulfurylase domain of human PAPS synthase 2. Biochem Biophys Res Commun 586:1–7. https://doi.org/10.1016/j.bbrc.2021.11.062