Genomic and ecosystem evidence demonstrate the importance of selenium for the harmful alga, *Aureococcus anophagefferens*

Christopher J. Gobler1,*, Alexei V. Lobanov2, Ying-Zhong Tang1, Martina Doblin3, Anton A. Turanov2, Yan Zhang2, Gordon T. Taylor1, Sergio A. Sañudo-Wilhelmy4, Igor V. Grigoriev5, Vadim N. Gladyshev2*

1Stony Brook University, School of Marine and Atmospheric Sciences, Stony Brook, NY 11794-5000, USA
2Brigham and Women's Hospital and Harvard Medical School, Boston MA 02115, USA
3University of Technology Sydney, Climate Change Cluster, NSW, 2007, Sydney, Australia
4University of Southern California, Department of Biological Sciences and Department of Earth Sciences, Los Angeles, CA, 90089 USA
5US Department of Energy, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, USA 94598, USA

March 2013

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

Genomic and ecosystem evidence demonstrate the importance of selenium for the harmful alga, *Aureococcus anophagefferens*

Christopher J. Gobler[1,*], Alexei V. Lobanov[2], Ying-Zhong Tang[1], Martina Doblin[3], Anton A. Turanov[2], Yan Zhang[2], Gordon T. Taylor[1], Sergio A. Sañudo-Wilhelmy[4], Igor V. Grigoriev[5], Vadim N. Gladyshev[2*]

[1]Stony Brook University, School of Marine and Atmospheric Sciences, Stony Brook, NY 11794-5000, USA

[2]Brigham and Women's Hospital and Harvard Medical School, Boston MA 02115, USA

[3]University of Technology Sydney, Climate Change Cluster, NSW, 2007, Sydney, Australia

[4]University of Southern California, Department of Biological Sciences and Department of Earth Sciences, Los Angeles, CA, 90089 USA

[5]US Department of Energy, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, USA 94598, USA.

[*]Corresponding authors: Christopher.gobler@sunysb.edu; vgladyshev@rics.bwh.harvard.edu

Abstract

Selenium is required for the biosynthesis of selenocysteine, the 21st amino acid in the genetic code. Here, we examined the role of selenium in the biology and ecology of the harmful pelagophyte, *Aureococcus anophagefferens*, through cell culture, genome analysis, and ecosystem studies. This organism had the largest and the most diverse known selenoproteome that included all known eukaryotic selenoproteins, selenoproteins previously found only in bacteria, and novel selenoproteins. The *A. anophagefferens* selenoproteome was dominated by the thioredoxin fold and oxidoreductase functions could be assigned to the majority of selenoproteins. Insertion of selenocysteine in these proteins was supported by a unique selenocysteine sequence and selenoprotein diversity and abundance was supported by metabolic ^{75}Se labeling of this alga. Selenium was required for the growth of *A. anophagefferens*: cultures grew maximally at nanomolar selenium concentrations and displayed a half saturation constant of 0.27 nM. In field experiments, dissolved selenium concentrations were up to 1 nM before and after the *A. anophagefferens* blooms, but were reduced to 0.05 nM during the peak of blooms. Consistent with these findings, enrichment of seawater with selenite during field incubation experiments before and after a bloom did not affect the growth of *A. anophagefferens*, but additions during the peak of the blooms significantly increased population growth rates ($p<0.0001$). Collectively, these findings demonstrate that selenium inventories, which can be anthropogenically enriched, influence the temporal and spatial occurrences of harmful brown tides caused by *A. anophagefferens* through synthesis of a large arsenal of algal selenium-dependent oxidoreductases that control cellular redox homeostasis.

Introduction

Selenium (Se) is an essential element, which is required for the biosynthesis of selenocysteine (Sec), the twenty first naturally occurring amino acid in the genetic code (Stadtman 1996). Being a trace element, Se is the 66th most abundant element in the earth's crust, but it can be enriched in some alkaline sediments (Bodek et al, 1988). Se can also be anthropogenically enriched in aquatic environments by numerous agricultural, industrial and urban activities including fossil fuel (coal) combustion, refinery activity, irrigation of selenium-rich soils, fertilizer use, and sewage discharge (Bodek et al, 1988, Cutter 1989, Cappon 1991). High levels of Se can be toxic to a suite of aquatic organisms (Lemly 1985, Hoffman et al 1988; Hamilton et al. 1990).

The essential nature of Se is related to its occurrence in proteins in the form of Sec residues. Sec is inserted cotranslationally in response to UGA codon, which is recoded from the its normal stop codon function if a control element (called SECIS element) is encountered in the 3' UTR of eukaryotic and archaeal mRNAs, or immediately downstream of UGA in the coding regions of bacterial selenoprotein genes (Berry et al 1991, Stadtman 1996, Hatfield et al 2006). Selenoproteins can be a thousand times more effective as catalysts than their cysteine homologs (Kim et al 2006), a fact that likely accounts for the cellular investment in Se-dependent pathways and the specialized machinery used for Se insertion into proteins (Lobanov et al 2009). The size of eukaryotic selenoproteomes varies significantly, are generally larger in aquatic organisms than in terrestrial organisms, and are specifically large in microalgae which may have a more steady supply of Se in the seawater compared to terrestrial organisms (Lobanov et al 2009).

Se has been shown to be required for many, but not all phytoplankton. Examples of absolute Se requirements have been documented within Bacillariophyceae, Chlorophyceae, Dinophyceae, Phaeophyceae, Raphidophyceae, Rhodophyceae, Prymnesiophyeae, and Prasinophyceae (Harrison et al 1988, Fries 1982, Wheeler et al 1982, Keller et al 1984, Imai 1996, Doblin et al 1999). Se is also required by many dinoflagellates, particularly those known to form harmful algal blooms (HABs) including *Scrippsiella trochoidea* (Harrison et al 1988), *Katodinium rotundatum* (Harrison et al 1988), *Gymnodinium catenatum* (Doblin et al 2000 Band-Schmidt et al 2004), *Alexandrium minutum* (Doblin et al 1999), *Protoceratium reticulatum* (Mitrovic, et al 2004), *Karenia selliformis* (Mountfort,et al 2006), *Cochlodinium polykrikoides* (Lee 2008), *Prorocentrum minimum* (Baines and Fisher 2001), and *Peridinium gatunense* (Lindstrom 1991).

The pelagophyte *Aureococcus anophagefferens* is another alga that forms HABs sometime referred to as 'brown tides'. The recent sequencing and annotation of the *A. anophagefferens* genome revealed an unusually high number of Sec-containing proteins in this organism (Gobler et al 2011), i.e., at least twice the number present in *Ostreococcus lucimarinus*, which previously had the largest known selenoproteome (Palenik et al 2007; Lobanov et al., 2007). Despite this recent discovery, little is known regarding the specific selenoproteome organization, composition, and properties as well as regarding importance of Se in the ecology of *A. anophagefferens*.

The goal of the current study was to examine the role of Se and selenoproteins in the biology and ecology of the harmful alga, *A. anophagefferens*. Axenic cultures were grown under a range of Se concentrations to quantitatively establish the Se levels required for maximal growth. The selenoproteome, selenoprotein sequences, and the SECIS elements of this species

were characterized. The dynamics of naturally occurring blooms of this species and dissolved Se concentrations were quantified and, finally, Se-enrichment experiments were performed with field populations.

Materials and Methods

Culture experiments

Se concentrations required to achieve maximal growth rates in axenic *A. anophagefferens* clone CCMP1984 were determined using G-medium made from artificial seawater (Doblin et al 1999) supplemented with differing concentrations of Se added as sodium selenite. Cultures were grown at 21 °C in an incubator with a 12:12 h light:dark cycle, illuminated by a bank of fluorescent lights that provided a light intensity of ~100 µmol quanta m^{-2} s^{-1} to cultures. These conditions approximated temperature and light exposures found in Long Island estuaries during summer months when *A. anophagefferen*s blooms (Gobler and Sunda 2012). Cultures were maintained for a minimum of four transfers at each concentration prior to the collection of final growth rate data to ensure that cells were fully acclimated to treatment conditions and that the carryover of Se from the initial, full strength media was eliminated. Cellular growth rates were calculated for cultures based on cell densities (determined microscopically) in exponential growth phases, using the formula $\mu = \ln (B_t/B_0)/t$, where B_o and B_t are the initial and final biomass, and t is the incubation duration in days. Growth rates were averaged over the entire exponential phase, which typically persisted for 3 – 6 days, depending on the concentration of Se in the media. Se-limitation of cultures was confirmed by the stimulation of growth of cultures in stationary phase at concentrations below 5 nM following the addition of 10 nM Se. The kinetic

terms μ_{max} (maximum growth rate) and K_S (half saturation constant) were derived using Michaelis-Menton curve fitting functions within Kaleidagraph (Synergy Software) version 4.1.1.

Identification of genes coding for selenoproteins in A. anophagefferens

Identification of selenoprotein genes in the *Aureococcus* genome was performed using SECISearch (Lobanov et al 2007), which analyzes primary sequences and secondary structures and then calculates the free energy for various parts of potential SECIS elements. The search was carried out using both default and loose patterns of SECISearch to accommodate identification of unusual SECIS structures. The searches also were then extended to identify organism-specific structures by a modified SECISearch (Lobanov et al 2007). For identified candidate SECIS elements, ORFs were predicted in the regions upstream of the SECIS elements. The presence of at least one homologous protein in the NCBI non-redundant database was used as an additional requirement. On the final step, a manual sequence and homology analysis of predicted selenoprotein ORFs located upstream of candidate SECIS elements was carried out.

In addition to the SECIS-based search procedure, the *A. anophagefferens* genome was analyzed with TBLAST against all known selenoprotein sequences to identify homologs of previously described selenoprotein genes. A third implemented procedure included a search for Sec/Cys pairs in homologous sequences (Lobanov et al 2007). We extracted all ORFs containing in-frame UGA codons, and homologs containing Cys in place of Sec were found by identifying the Sec/Cys pairs flanked by homologous sequences. For this procedure, we used TBLASTX to examine all potential ORFs with in-frame UGA codons against NCBI non-redundant protein database. All candidates were then tested with SECISearch for the presence of SECIS elements. Finally, PSI-BLAST was used to identify additional more distant homologs. All datasets

obtained with the three independent methods were combined and the proteins were classified as homologs of previously known selenoproteins, novel selenoproteins and candidate selenoprotein genes.

Metabolic labeling of the A. anophagefferens selenoproteome

To experimentally verify the occurrence of selenoproteins, *A. anophagefferens* cells were subjected to metabolic labeling with ^{75}Se. *Chlamydomonas reinhardtii* was labeled in parallel and used as a control. The metabolic labeling was performed as described previously (Novoselov et al 2002). Briefly, 100 ml of each culture were metabolically labeled with 100 μCi of ^{75}Se ([^{75}Se]selenious acid, specific activity 1,000 Ci/mmol, from, the Research Reactor Facility, University of Missouri, Columbia, MO) for 48 h. Cells were collected, resuspended in PBS and sonicated. Next, 25-50 μg of total soluble protein from each organism were resolved by SDS-PAGE and transferred onto a PVDF membrane (Invitrogen). Selenoprotein patterns were visualized with a PhosphorImager.

Field collections

During the spring and summer of 1998, seawater samples were collected twice weekly to bi-weekly from West Neck Bay, NY, a small (1 km^2), shallow (2 - 4 m), enclosed embayment on Shelter Island, within the Peconic Estuary of eastern Long Island. The absence of tributaries and point source anthropogenic inputs make this bay a relatively simple system, where chemical inputs can be largely attributed to groundwater seepage or internal recycling (Gobler and Sañudo-Wilhelmy 2001). All sampling material used in this study was prepared using trace metal clean techniques (Gobler et al 2002). Water samples were collected with a peristaltic pump connected to acid-washed Teflon tubing on a bamboo pole extended 4 m upwind and lowered to a depth of 1 m. Dissolved samples were obtained by filtration through trace metal clean, polypropylene capsule filters (0.2 μm). All analyses and sample manipulation performed

after collection were conducted within vertical flow clean benches supplied with HEPA-filtered air. Triplicate chlorophyll *a* samples were collected on precombusted GF/F glass fiber filters (nominal pore size = 0.7 μm) and analyzed by standard fluorometeric methods (Parsons et a 1984). *A. anophagefferens* densities were determined on field samples fixed with glutaraldehyde (1% final concentration). Samples were enumerated by direct count methods employing a Zeiss model D-7082 epi-fluorescent microscope and an immunofluorescent label as described by Anderson et al 1989. The amount of chlorophyll *a* in West Neck Bay due to the presence of *A. anophagefferens* was estimated by assuming a constant chl *a* per cell value (0.035 ± 0.003 pg $cell^{-1}$ for nutrient replete cultures (Gobler and Sañudo-Wilhelmy 2001) and multiplying this value by the *Aureococcus* density. Although such calculations can be biased by variability of cellular chlorophyll content due to changes in light and nutrient regimes, such approximations have been used successfully in the past to compare *Aureococcus* biomass to that of the total algal community (Gobler and Sañudo-Wilhelmy 2001; Caron et al 2004).

Groundwater entering West Neck Bay through the intertidal zone was sampled from 1 m, Teflon-lined, PVC piezometers with 2.5 cm horizontal screened slits along the lower 25 cm. High groundwater seepage rates at West Neck Bay allowed piezometers to fill with fresh groundwater (salinity < 0.1 ppt) when sampled during low tide. Groundwater was sampled using a peristaltic pump equipped with acid-washed Teflon tubing. To ensure representative groundwater was sampled, piezometers were purged at < 100 mL min^{-1} and samples were not obtained until the conductivity, dissolved oxygen, temperature, and pH of the pumped groundwater stabilized (Puls and Paul 1995). Filtered groundwater or baywater samples were analyzed for total dissolved Se as described by Cutter (1982, 1983). Briefly, total dissolved Se was determined by boiling a 4 M HCl acidified sample, with the addition of potassium

persulfate. Next, within a glass stripping vessel selenite was quantitatively converted to hydrogen selenide by adding sodium borohydride to a sample containing sulfanilamide (to eliminate interference due to nitrite). The evolved hydrogen selenide was stripped from solution using helium and trapped in a borosilicate U-tube packed with silanized glass wool and immersed in liquid nitrogen. After the trap was removed from the LN2, an atomic absorption spectrometer fitted with an open quartz tube furnace burning an air–hydrogen flame was used to detect the hydride. Instrument response (as peak area) was recorded on a chromatographic integrator. To ensure accuracy, all determinations utilized the standard additions method of calibration, and all samples were analyzed in triplicate to quantify precision (found to be 4%). Detection limits for dissolved Se was 0.02 nM.

Field experiments

Field experiments were conducted on June 4th, June 22nd, July 3rd, and July 21st of 1998 using water from West Neck Bay. Within two hours of collection, 50 mL of seawater were transferred to trace metal clean, 60 mL polycarbonate flasks in a HEPA laminar flow hood to prevent trace metal and bacterial contamination. Triplicate flasks were amended with sodium selenite (10 nM) or were left unamended as a control treatment. The selenite solution was cleaned of trace metals with Chelex-100 ion exchange resin (Bruland, 1980), filter sterilized (0.2 µm) and frozen before use. Amended and unamended flasks were incubated at the same temperature found in West Neck Bay during water collection under 125 µEin m^{-2} sec^{-1} of light on a light:dark cycle which mimicked summer conditions in West Neck Bay (14hr : 10hr). The average incoming solar radiation during day light hours (6 am to 8 pm) to Long Island during June and July 1998 was 2140 µEin m^{-2} sec^{-1} (V. Cassella, Brookhaven National Lab, pers. comm), and the average extinction coefficient in the West Neck Bay water column during

experiments was 2.26 (Gobler and Sañudo-Wilhelmy 2001). Therefore, the 125 µEin m^{-2} sec^{-1} used in our experiments was equivalent to the light levels found at 1.4 m in the water column of West Neck Bay during this period, or ~ 4% of incident radiation. After 48 hr, samples from each flask were preserved to a final concentration of 1% glutaraldehyde in sterile polycarbonate test tubes for enumeration of *A. anophagefferen*s as described above. Net specific growth rates of *A. anophagefferens* were determined as described for culture experiments.

Results

Requirement of Se for growth of A. anophagefferens

Axenic *A. anophagefferens* (CCMP9184) cultures displayed a sigmoidal increase in growth rate in response to increasing selenite concentrations (Fig 1). Cultures did not grow below 10 pM selenite, but displayed a gradual increase in growth above 10 pM and near maximal growth rates (~0.6 d^{-1}) at concentrations of selenite at or above 5 nM (Fig 1). A half-saturation constant (K_s) for selenite was 0.27 ± 0.08 nM and a µmax 0.60 ± 0.03 d^{-1}, with a Pearson product-moment correlation coefficient (R) of 0.97.

The Aureococcus anophagefferens selenoproteome

We detected 59 selenoprotein genes in *A. anophagefferens*, which is an extremely large number of such proteins for any living organism. For example, this is twice the number of selenoproteins in *Ostreococcus lucimarinus*, which previously had the largest known eukaryotic selenoproteome (Palenik et al 2007; Lobanov et al., 2007) and four-fold more than in the genomes of various diatoms and green algae (Fig 2). Currently, the *Aureococcus* selenoproteome is the largest selenoproteome of all living organisms. It is characterized by the occurrence of

almost all known eukaryotic selenoproteins and also has selenoproteins that were previously described only in bacteria. Nearly half of the selenoproteome are comprised of methionine sulfoxide reductases, thioredoxin reductases, glutathione peroxidases, glutaredoxins, and peroxiredoxins, as several selenoproteins occurred in multiple copies (Table 1, Supplementary Table 1). For example, selenoprotein methionine sulfoxide reductases were represented by four MsrA isozymes and three MsrB isozymes (Table 1). Many novel selenoproteins with unknown functions were also detected (Table 1). Further analysis of the selenoproteome revealed dominance of the thioredoxin fold, which accounted for two-thirds of detected selenoproteins. In all these proteins, Sec was located in place of the catalytic Cys found in thioredoxin. This observation and an additional presence of many other oxidoreductases (MsrAs, MsrBs, GILT, TRs, etc.) point to the redox function of Sec in selenoproteins, which in turns suggests the role of Se in regulating redox homeostasis in *Aureococcus*.

As in other organisms, Sec is inserted into *A. anophagefferens* selenoproteins with the help of SECIS elements (Fig. 3, 4, 5), which are RNA structures present in the 3'-UTRs of selenoprotein genes. However, *A. anophagefferens* SECIS elements differ from the canonical SECIS element in that most such structures in this organism have a very small apical loop and lack a mini-stem and a conserved unpaired AA sequence (Fig. 5). Therefore, to identify selenoprotein genes, SECISearch was adjusted to recognize the *A. anophagefferens* consensus structure developed based on the analyses of SECIS elements in known selenoprotein genes. The use of this modified program allowed us to define the selenoproteome as having 59 selenoproteins. It should be noted, however, that we detected several additional candidate selenoproteins, but they could not be unambiguously confirmed due to lack of homologs in other organisms, poor SECIS elements, and possibly incorrect gene models. We think, however, that

the actual number of selenoproteins in *Aureococcus* will exceed 59. We also detected components of Sec insertion machinery, including Sec-specific elongation factor EFsec, a SECIS-binding protein, and Sec tRNA (Fig 3). The latter was not detected by tRNAscan, but could be identified in a maximum sensitivity mode of this program.

A characteristic feature of *Aureococcus* SECIS elements is their unusually high diversity (Fig 4). While several SECIS elements found in the *A. anophagefferens* genome correspond to the canonical SECIS model (such as the structure in DUF1000, Fig 4A), a surprisingly large number of SECIS elements does not. The most common, characteristic differences with canonical model are shown with Selenoprotein W and PDI SECIS elements as examples. One of these features is an unusually long stem. While a typical length of the stem (distance from the SECIS core to the apical loop) is 10-13 nucleoides, in *Aureococcus* SECIS elements it can be nearly twice as long (Fig 4B). Another characteristic feature is the extremely small, virtually absent apical loop. The size of the apical loop in the canonical model varies from 7 to 27 nucleotides, while many *Aureococcus* SECIS elements have apical loops consisting of only 3 nucleotides (Fig 4C). It remains to be seen if there are any extra constrains that disrupt nucleotide pairing near the end of the stem to increase the size of the apical loop.

Metabolic labeling of A. anophagefferens with ^{75}Se

We next subjected *A. anophagefferens* cells to metabolic labeling with ^{75}Se. For comparison, *Chlamydomonas reinhardtii*, which has 12 selenoprotein genes (Merchant et al., 2007, Novoselov et al 2002) was labeled in parallel (Fig 6). This experiment revealed both a large number of selenoproteins in *Aureococcus* and a significant increase in abundance of these proteins in comparison with *Chlamydomonas*. In addition, abundant selenoprotein bands in the 8-

28 kDa region corresponded the predicted masses of the majority of *Aureococcus* selenoproteins. These observations further highlight the significant use of Sec and selenoproteins in *Aureococcus* in comparison with other organisms.

Field dynamics and experiments

During a field study in the spring and summer of 1998, dissolved Se concentrations in West Neck Bay averaged 0.71 ± 0.37 nM (Fig 7). *A. anophagefferens* densities steadily increased from 1.0 x 10^4 cells mL^{-1} in late May to > 1.0 x 10^5 cells mL^{-1} in late June (Fig 7), representing 10% to 25% of the algal population. *A. anophagefferens* reached peak densities in early July, as cell densities grew to peak of 5.6 x 10^5 cells mL^{-1} (Fig 7) and represented over 95% of algal biomass. During this bloom peak dissolved selenium levels decreased to their lowest levels of the year (0.05 nM; Fig 7). Following the peak of the bloom, dissolved selenium concentrations rose sharply to > 1 nM and remained elevated for the sampling year as *A. anophagefferens* densities steadily declined to nearly zero (Fig 7). Groundwater entering West Neck Bay contained, on average, 2.71 ± 0.54 nM of dissolved Se (n=6). During an incubation experiments conducted in June and late July, the addition of selenite did not significantly alter the growth of *A. anophagefferens*. During the incubation experiment conducted on July 8th during the peak of the *A. anophagefferens* bloom, the addition of 10 nM selenite yielded a significant increase in *A. anophagefferens* growth rates compared to an unammended control (Fig 8; $p<0.0001$; T-test).

Discussion

A. anophagefferens is the first example of a pelagophyte that requires Se and the fact that it is currently the organism with both the largest and the most diverse selenoproteome on the

planet suggests that other pelagophytes may also extensively utilize Se. Culture experiments demonstrated that nM concentrations of Se are required for maximal growth rates of this species and field evidence shows that blooms reduced concentrations of dissolved Se below this level and caused the growth rate of this alga to be limited by the Se supply. Collectively, our findings demonstrate the essential role of the micronutrient Se as a component of selenoproteins in this alga and perhaps other pelagophytes and/or harmful algae.

Currently, the genome of *A. anophagefferens* contains the largest number of selenoproteins in any living organism with a fully sequenced genome. Moreover, the set of selenoproteins reported here represents a conservative estimate: the true number of selenoproteins is likely even larger. Several candidate selenoproteins were not included in the selenoproteome because they were lacking homologs in other organisms or identifiable SECIS elements in the 3'-UTR (mostly due to insufficient length, or unusual form of the potential SECIS elements). In addition, the proteins that were highly similar to each other were not always reported as separate entities, since additional verification was required to avoid overreporting caused by misassembly, splicing variants, and sequencing errors. Thus, it is very likely that the number of selenoproteins in *Aureococcus* exceeds 60.

Organism-specific changes in SECIS element structure such as those found here for *A. anophagefferens* have been reported previously. For example, the majority of *Ostreococcus* SECIS elements possess an extra long mini-stem (Lobanov et al 2007). It was also found that *Dictyostelium discoideum* SECIS elements for evolutionarily unrelated proteins are highly conserved, with five nucleotides preceding and two nucleotides following the SECIS core being identical in all identified SECIS elements. This example points to a case of convergent evolution in SECIS structure and possibly to constraints associated with SECIS binding proteins. Finally,

SECIS elements in *Neospora* and *Toxoplasma* were found with a noncanonical GGGA sequence in the SECIS core (Novoselov et al 2007). Thus, deviations from the eukaryotic SECIS model, while rare, may be expected, and *Aureococcus* joins such group of organisms with lineage-specific SECIS functions.

Another feature of the *Aureococcus* selenoprotein set is that it represents both the majority of eukaryotic selenoproteins and several selenoproteins that were previously only detected in bacteria. It is an attractive possibility that these proteins were acquired by *Aureococcus* through lateral transfer from bacteria. Finally, many completely new selenoproteins were discovered, mostly with unknown functions. Location of Sec in these proteins points to both oxidoreductase function of these proteins and identifies the catalytic Cys residues in homologs of these proteins in other organisms (Fomenko et al., 2007).

The set of predicted selenoprotein functions provides insight into the ecology of this organism. About two-thirds of *Aureococcus* selenoproteins possess a thioredoxin fold, and several additional proteins are homologs of known thiol oxidoreductases. The *Aureococcus* selenoproteins represent all known redox regulatory systems (thioredoxin, glutaredoxin, methionine sulfoxide reductase, etc) as well as systems involved in disulfide bond formation and isomerization (PDI, GILT, etc). The functions of several selenoproteins are not known, but it is clear that at least some (and possibly all) of them are oxidoreductases. These findings indicate that *Aureococcus* uses Se to fine-tune its redox regulation. We suggest that this feature provides this organism with competitive advantage over its competitors during algal blooms. The shallow nature and rapid vertical mixing times (~1 h; Milligan and Cosper 1997) of estuaries where *A. anophagefferens* chronically exposes phytoplankton cells to extremes in light and temperature that are likely to create high levels of intracellular oxidative stress. The large number and

multiple redundancies of selenoproteins in *A. anophagefferens* help protect these cells against oxidative through the removal of hydroperoxides and the repair of oxidatively damaged proteins (Stadtman 1996) and thus likely infer a competitive advantage to this alga, particularly during summer when it blooms and vertical gradients in temperature light attenuation in estuaries can be extreme (Cosper et al 1987).

A highly increased use of Sec in *A. anophagefferens* compared to other organisms that occupy the same environmental niche should make this organism particularly competitive during blooms when other, non-Se resources such as nitrogen and light are scarce (Gobler et al 2011). Sec residues are often superior catalysts compared to Cys (Stadtman 1996; Hatfield and Gladyshev 2002; Kim and Gladyshev 2005), and therefore, selenoproteins help *A. anophagefferens* to more efficiently organize its metabolism. Increased reliance on Se to support biosynthesis of so many proteins also suggests that *A. anophagefferens* can efficiently sequester Se from environment, which is also confirmed by direct metabolic labeling of *Aureococcus* with ^{75}Se as well as by the decline of dissolved Se in estuaries with brown tides. Moreover, reliance on Se would also make *Aureococcus* sensitive to levels of this element in the environment and vulnerable to electrophilic compounds and alkylating agents that target Sec residues in proteins (Hatfield and Gladyshev 2002).

While maximal growth rates for *A. anophagefferens* occurred at Se at or above 5 nM, the half saturation constant for growth was 0.27 nM. In the environment, dissolved Se concentrations ranged from 0.05 to 1.35 nM, above and below the half saturation and below the levels that elicited for maximal growth of cultures. Importantly, multiple species of Se are present within the dissolved pool in marine environments (Cutter and Bruland 1984) and some of the species are less biovailable or unavailable to phytoplankton (Cutter and Bruland, 1984;

Cutter and Cutter, 1998), indicating the bioavailable pool of Se to *A. anophagefferens* was a fraction of the levels measured within the total dissolved pool. Half-saturation constants for nutrients are often utilized as proxies for nutrient limitation in marine ecosystems with concentrations below half-saturation constants are often considered limiting (Caperon and Meyer, 1972; Fisher et al., 1992). The lowest concentrations of Se (0.05 - 0.22 nM) were observed during the peak of the *A. anophagefferens* bloom were below its half saturation constant and likely due to the large Se demand created by this bloom, a pattern that has been observed for other essential nutrients during blooms such as nitrogen (Gobler and Sunda 2012). These low Se concentrations present at the peak of the bloom seem to restrict the growth of *A. anophagefferens* as the addition of 10 nM Se during this period significantly enhanced *A. anophagefferens* growth rates. Again, since a substantial portion of the total dissolved Se pool may be comprised of species that are not assessable by phytoplankton, the actual bioavailable pool of Se during the peak of the brown tide may be lower than the 0.05 nM measured. As such, a lack of Se may have contributed to the demise of the 1998 *A. anophagefferens* bloom in West Neck Bay and thus is likely a factor that influences the occurrence of brown tides.

A. anophagefferens blooms in shallow, enclosed estuaries such as West Neck Bay where concentration of Se averaged ~1 nM but never blooms in deep estuaries or continental shelf regions that are characterized by lower selenium concentrations (Measures et al 1984; Cutter and Cutter 2001; Cutter and Cutter 1995; Cutter and Bruland 1984). The formation of blooms exclusively in shallow estuaries ensures that *A. anophagefferens* has access to a rich supply of the Se required to synthesize these ecologically important and catalytically superior enzymes (Stadtman 1996; Hatfield and Gladyshev 2002; Kim and Gladyshev 2005). Since *A. anophagefferens* relies on selenoproteins for growth and since a scarcity of Se may prohibit

bloom formation in off-shore waters, Se is likely to play a key role in shaping the niche space and bloom occurrences of this species. Moreover, since some phytoplankton do not require Se (Harrison et al 1988), Se availability is likely to shape the succession and composition of phytoplankton communities, in general.

Se can be anthropogenically enriched in aquatic environments by numerous activities (Bodek et al 1988, Cutter 1989). Consistent with this hypothesis, the watershed of West Neck Bay is entirely surrounded by residential dwellings and levels of Se were measured in groundwater entering West Neck Bay were four-fold greater than the bay and other surface waters (Cutter and Cutter 1995, 2001), perhaps due to the strong influence of wastewater or fertilizers from the watershed (Bodek et al 1988, Cappon 1991, McBride and Spiers 2001). Since groundwater is nearly the exclusive source of freshwater to this (Schubert 1998) and other estuaries which host brown tides (LaRoche et al 1997), the anthropogenic loading of Se into this and perhaps other coastal ecosystems may lead to the intensification of blooms of *A. anophagefferens*. Since Se is required by many other phytoplankton that form HABs including *Chattonella verruculosa* (Imai et al 1996), *Scrippsiella trochoidea* (Harrison et al 1988), *Katodinium rotundatum* (Harrison et al 1988), *Gymnodinium catenatum* (Doblin et al 2000, Band-Schmidt et al 2004), *Alexandrium minutum* (Doblin et al 1999), *Protoceratium reticulatum* (Mitrovic et al 2004), *Karenia selliformis* (Mountfort et al 2006), *Cochlodinium polykrikoides* (Lee 2008), *Prorocentrum minimum* (Baines and Fisher 2001), and *Peridinium gatunense* (Lindstrom 1991), anthropogenic loading of this element may contribute to the occurrence of these events globally.

Acknowledgements: This study was supported by awards from National Oceanic and Atmospheric Administration Center for Sponsored Coastal Ocean Research award

#NA09NOS4780206 to CJG and NIH grant GM061603 to VNG as well as an award from Suffolk County, NY, USA, to GTT and SSW.

References

Anderson, D.M., Kulis, D.M., and Cosper, E.M. (1989). Immunofluorescent detection of the "brown tide" organism, *Aureococcus anophagefferens*. In: Cosper, E.M., Bricelj, V.M., and Carpenter, E.J. (ed.), Novel phytoplankton blooms: Causes and impacts of recurrent Brown Tides and other unusual blooms, lecture notes on coastal and estuarine studies. Springer-Verlag. p. 213-228.

Baines, S.B., Fisher, N.S. 2001. Interspecific differences in the bioconcentration of selenite by phytoplankton and their ecological implications. Mar Ecol Prog Ser. 213: 1-12.

Band-Schmidt, C.J., Morquecho, L. , Lechuga-Deveze, C.H., Anderson, D.M. 2004. Effects of growth medium, temperature, salinity and seawater source on the growth of Gymnodinium catenatum (Dinophyceae) from Bahia Concepcion, Gulf of California, Mexico. J Plankton Res 26: 1459-1470

Berry, M.J., Banu, L., Chen, Y.Y., Mandel, S.J., Kieffer, J.D., Harney, J.W. and Larsen, P.R. (1991) Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. Nature 353: 273–276.

Bodek, I., Lyman, W.J., Reehl, W.F., and Rosenblatt, D.H. 1988 Environmental inorganic chemistry: properties, processes, and estimation methods. Pergamon Press, NY

Bruland, K.W. (1980). Oceanographic distribution of cadmium, nickel, and copper in the North Pacific. Earth Planet. Sci. Lett. 47: 176-198

Caperon, J. and J. Meyer. 1972. Nitrogen-limited growth of marine phytoplankton. II Uptake kinetic and their role in nutrient growth of phytoplankton. Deep-Sea Research 19:619-632.

Cappon CJ. 1991. Sewage sludge as a source of environmental selenium. Sci Total Environ. 100:177-205.

Caron, D.A., Gobler, C.J., Buck, N.J., Lonsdale, D.J., Cerrato, R.M., Schaffner, R.A, Rose, J., Taylor, G. Boissonneault, K.R. .Mehran, R.. 2004. Microbial herbivory on the brown tide algal, *Aureococcus anophagefferens:* results from natural ecosystems, mesocosms and laboratory experiments. *Harmful Algae* 3: 439–457

Cosper, E.M., Dennison, W.C., Carpenter, E.J., Bricelj, V.M., Mitchell J. G., Kuenstner S.H, Colflesh D.C., Dewey M., 1987. Recurrent and persistent "brown tide" blooms perturb coastal marine ecosystem. Estuaries 10, 284-290.

Cutter, G. A. (1983). Elimination of nitrite interference in the determination of selenium by hydride generation. Anal. Chim. Acta, 149:391-394.

Cutter, G. A. (1989). Selenium in freshwater systems. In *Occurrence and distribution of Selenium*. Ed. M. Ihnat, CRC Press, Florida, p. 243-262.

Cutter, G. A and Cutter, L. S. (1995). Behavior of dissolved antimony, arsenic, and selenium in the Atlantic Ocean. Mar. Chem., 49: 295-306.

Cutter, G. A. and Cutter, L. S. (2001). The biogeochemical cycling of selenium in the western Atlantic Ocean. Deep-Sea Res. II, *in press*.

Cutter, G.A. 1982. Selenium in Reducing Water. Science 217:829-831

Cutter, G.A. 1989. The Estuarine Behavior of Selenium in San Francisco Bay. Estuarine, Coastal, and Shelf Science 28:13-24

Cutter, G.A. and Bruland, K.W. 1984. The marine biogeochemistry of selenium: a re-evaluation. Limnology and Oceanography 29:1179-1192

Cutter, G.A., Cutter, L.S. (2004) Selenium biogeochemistry in the San Francisco Bay estuary: changes in water column behavior. Estuarine Coastal Shelf Sci 61:463–476.

Doblin, M.A., Blackburn, S.I., Hallegraeff, G.M. (1999) Comparative study of selenium requirements of three phytoplankton species: Gymnodinium catenatum, Alexandrium minutum (Dinophyta) and Chaetoceros cf. tenuissimus (Bacillariophyta). J Plankton Res 21: 1153-1169.

Doblin, M.A., Blackburn, S.I., Hallegraeff, G.M. 2000. Intraspecific variation in the selenium requirement of different geographic strains of the toxic dinoflagellate Gymnodinium catenatum. J. Plankton Res 22: 421-432

Fisher, T.R., Peel, E.R., Ammerman, J.W., Harding L.W. 1992. Nutrient limitation of phytoplankton in Chesapeake Bay. Marine Ecology Progress Series. 82:51-63.

Fomenko, D.E., Xing, W., Adair, B.M., Thomas, D.J., Gladyshev, V.N. (2007) High-Throughput Identification of Catalytic Redox-Active Cysteine Residues. Science **315**, 387-389.

Fries, L. (1982). Selenium stimulates growth of marine macroalgae in axenic culture. J. Phycol. 18: 328-331

Gobler, C.J, Sunda, W.G. 2012. Ecosystem disruptive algal blooms of the brown tide species, *Aureococcus anophagefferens* and *Aureoumbra lagunensis*. In press to *Harmful Algae.*

Gobler, C.J. and 32 co-authors. 2011. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc Natl Acad Sci USA* 108:4352–4357.

Gobler, C.J., Donat, J.R., Consolovo, J.A. and S.A. Sañudo-Wilhelmy. 2002. Physicochemical speciation of iron during coastal algal blooms. *Marine Chemistry* 77: 71-89

Gobler, C.J., and S.A. Sañudo-Wilhelmy. 2001. Temporal variability of groundwater seepage and Brown Tide bloom in a Long Island embayment. *Marine Ecology Progress Series* 217: 299-309

Gobler, C.J., and S.A. Sañudo-Wilhelmy. 2001. Effects of organic carbon, organic nitrogen, inorganic nutrients, and iron additions on the growth of phytoplankton and bacteria during a Long Island Brown Tide bloom. *Marine Ecology Progress Series* 209: 19-34

Harrison, P. J., Yu, P. W., Thompson, P. A., Price, N. M., Phillips D. J. 1988. Survey of selenium requirements in marine Phytoplankton. Mar. Ecol. Prog. Ser. 47**: 89-96.**

Hamilton, S.J., Buhl, K.J., Faerber, N.L., Wiedermeyer, R.H., and Bullard, F.A. 1990. Toxicity of organic selenium in the diet of chinook salmon. Environmental Toxicology and Chemistry 9:347-358

Hoffman, D.J., Ohlendorf, H.M., and Aldrich, T.W. 1988. Selenium teratogenesis in natural populations of aqautic birds in central California. Archives of Environmental Contamination and Toxicology 17: 519-525

Hatfield, D.L., Gladyshev, V.N. (2002) How Selenium Has Altered Our Understanding of the Genetic Code. Mol Cell Biol 22, 3565-3576.

Imai, I., Itakura, S., Matsuyama, Y., et al. 1996. Selenium requirement for growth of a novel red tide flagellate Chattonella verruculosa (Raphidophyceae) in culture. Fisheries Sci 62: 834-835

Keller, M., Guillard, R. R. L., Provasoli, L., Pintner, I. J. (1984). Nutrition of some marine ultraplankton clones from the Sargasso Sea. Eos 65: 898

Kim, H.Y., Gladyshev, V.N. (2005) Different Catalytic Mechanisms in Mammalian Selenocysteine- and Cysteine-Containing Methionine-R-Sulfoxide Reductases. *PLoS Biol.* **3**, e375

Kim, H.Y., Fomenko, D.E., Yoon, Y.E., Gladyshev, V.N. 2006. Catalytic Advantages Provided by Selenocysteine in Methionine-S-Sulfoxide Reductases. Biochemistry 45:13697-13704.

Kryukov, G.V., Castellano, S., Novoselov, S.V., Lobanov, A.V., Zehtab, O., Guigo, R., Gladyshev, V.N. (2003) Characterization of mammalian selenoproteomes. Science 300, 1439-1443

LaRoche, J., Nuzzi, R., Waters, R., Wyman, K., Falkowski, P.G., Wallace, D.W.R. (1997) Brown tide blooms in Long Island's coastal waters linked to variability in groundwater flow. Global Change Biol 3: 397-410.

Lee, Y.S. 2008. Utilization of various nitrogen, phosphorus, and selenium compounds by Cochlodinium polykrikoides. J Environ Biol 29: 799-804

Lemly, A.D. 1985. Toxicology of selenium in a freshwater reservoir: Implications for envirnmental hazard evaluation and safety. *Ecotoxicology and Envirnomental Safety* 10:214-338

Lindstrom, K. 1991. Nutrient requirements of the dinoflagellate Peridinium gatunense. J Phycol 27, 207-219

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 and small with terrestrial life. *Genome Biol.* 8, R198

Lobanov, A.V., Hatfield, D.L., Gladyshev, V.N. (2009) Eukaryotic Selenoproteins and selenoproteomes. Biochim. Biophys. Acta. 1790:1424-1428.

McBride, M.B., Spiers, G. (2001) Trace element content of selected fertilizers and dairy manures as determined by ICP-MS. Comm Soil Sci Plant Anal 32:139–156.

Measures, C. I., Grant, B., Khadem, M. and Edmond, J. M. (1984). Distribution of Be, Al, Se and Bi in the surface waters of the western North Atlantic and Caribbean. Earth Planet. Sci. Lett., 71: 1-12.

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, Marshall WF, Qu LH, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen CL, Cognat V, Croft MT, Dent R, Dutcher S, Fernandez E, Fukuzawa H, Gonzalez-Ballester D, Gonzalez-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV., Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV., Paulsen IT, Pazour G, Purton S, Ral JP, Riano-Pachon DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen CJ, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN., Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martinez D, Ngau WC, Otillar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318: 245-250.

Milligan, A.J. and Cosper, E.M. (1997). Growth and photosynthesis of the "brown tide" microalga *Aureococcus anophagefferens* in subsaturating constant and fluctuating irradiance. Mar. Ecol. Prog. Ser. 153:67-75.

Mitrovic, S.M., Amandi, M.F., McKenzie, L., Furey, A., James, K.J. 2004. Effects of selenium, iron and cobalt addition to growth and yessotoxin production of the toxic marine dinoflagellate Protoceratium reticulatum in culture. J Expt Mar Biol Ecol 313: 337-351

Mountfort, D., Beuzenberg, V., MacKenzie, L. , Rhodes, L. 2006. Enhancement of growth and gymnodimine production by the marine dinoflagellate, Karenia selliformis. Harmful Algae 5: 658-664

Novoselov, S.V., Rao, M., Onoshko, N.V., Zhi, H., Kryukov, G.V., Xiang, Y., Weeks, D.P., Hatfield, D.L., Gladyshev, V.N. 2002 Selenoproteins and selenocysteine insertion system in the model plant cell system. Chlamydomonas reinhardtii. EMBO J. 21: 3681-3693.

Novoselov, S.V., Lobanov, A.V., Hua, D., Kasaikina, M.V., Hatfield, D.L., Gladyshev ,V.N. 2007. A highly efficient form of the selenocysteine insertion sequence element in protozoan parasites and its use in mammalian cells. Proc. Natl. Acad. Sci. USA 104:7857-7862.

Palenik et al. 2007. The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation Proc Natl Acad Sci USA 104:7705-7710

Parsons, T.R., Maita, Y., and Lalli, C.M. 1984. A manual of chemical and biological methods for seawater analysis. Pergamon.

Puls, R.W. and Paul, C.J. 1995. Low-flow purging and sampling of ground water monitoring wells with dedicated systems. Ground Water Monitoring Rev. 15: 116-123

Schubert, C.E. 1998. Areas contributing ground water to the Peconic Estuary, and ground-water budgets for the North and South Forks and Shelter Island, Eastern Suffolk County, New York. U.S. Geological Survey Water Resources Investigations Report 97-4136.

Stadtman, T.C. 1996. Selenocysteine. Annu Rev Biochem 65, 83-100.

Wheeler, A.E., Zingaro, R.A., Irgolic, K., and Bottino, N.R. 1982. The effect of selenate, selenite and sulfate on the growth of six unicellular marine algae. *Journal of Experimental Marine Biology and Ecology.* 57:181-194

Figure 1. Growth rates of *A. anophagefferens* in media with differing concentrations of selenite.

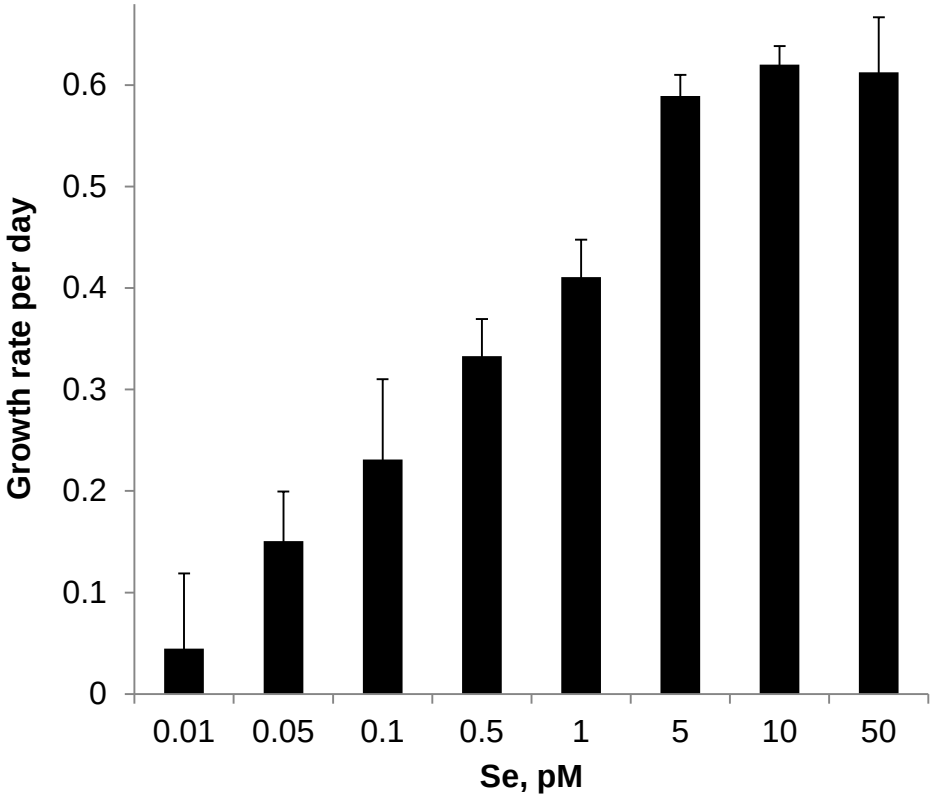


Figure 2. Size of selenoproteomes of six eukaryotic phytoplankton for which complete genomes are available.

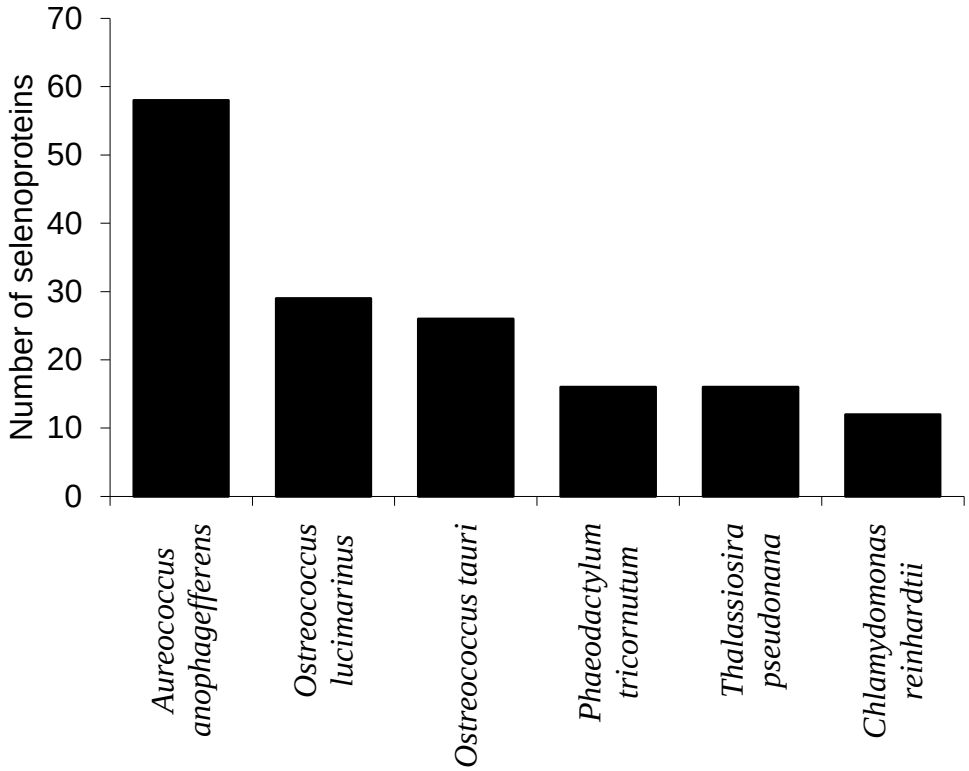
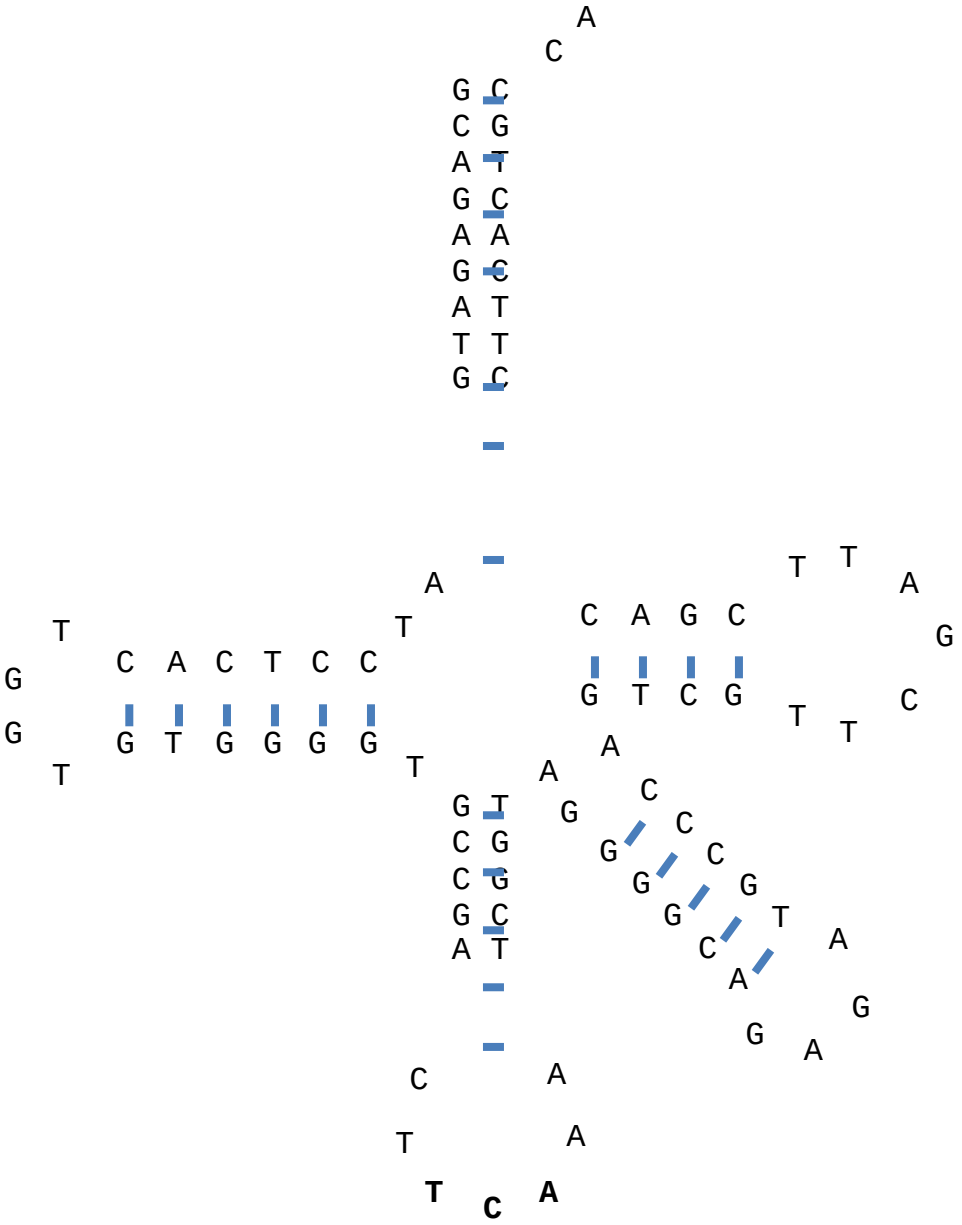


Figure 3. *A. anophagefferens* Sec tRNA sequence and its secondary structure.



>scaffold_12[1212565,1212654]
gcagagatgatcctcactggtgtggggtgccgact**tca**aatcggtaggggcagagatgcccagt
cgttcgattcgaccttcactgctcca

Figure 4. Examples of *A. anophagefferens* SECIS elements. (A) A typical eukaryotic SECIS element. Such structure occurred only in several *Aureococcus* selenoprotein genes. (B, C) Typical Aureococcus SECIS structures featuring a long stem and a small apical loop that lacked a conserved AA sequence and also lacked a mini-stem. An SBP2-binding site at the bottom of the main step is shown in bold.

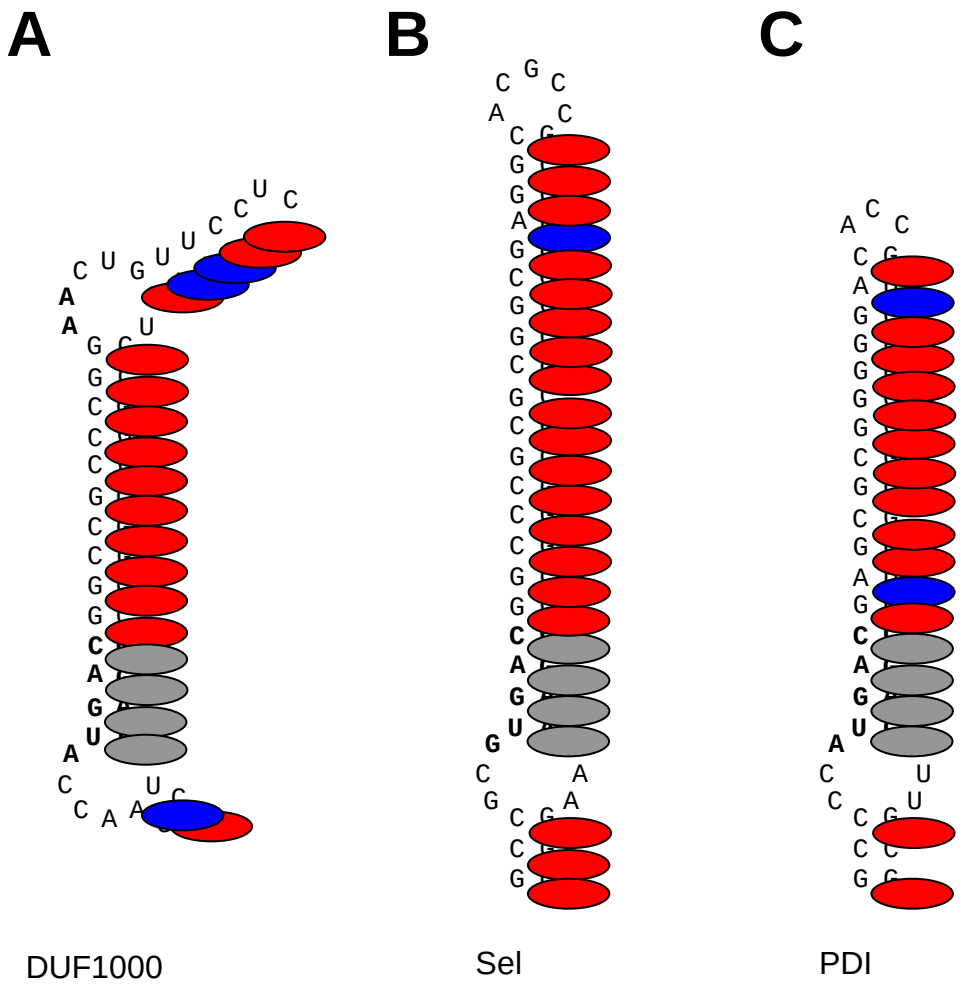


Figure 5. Nucleotide alignment of *Aureococcus* SECIS elements. Conserved nucleotides are highlighted. The SBP2-binding site is shown in red above the sequences.

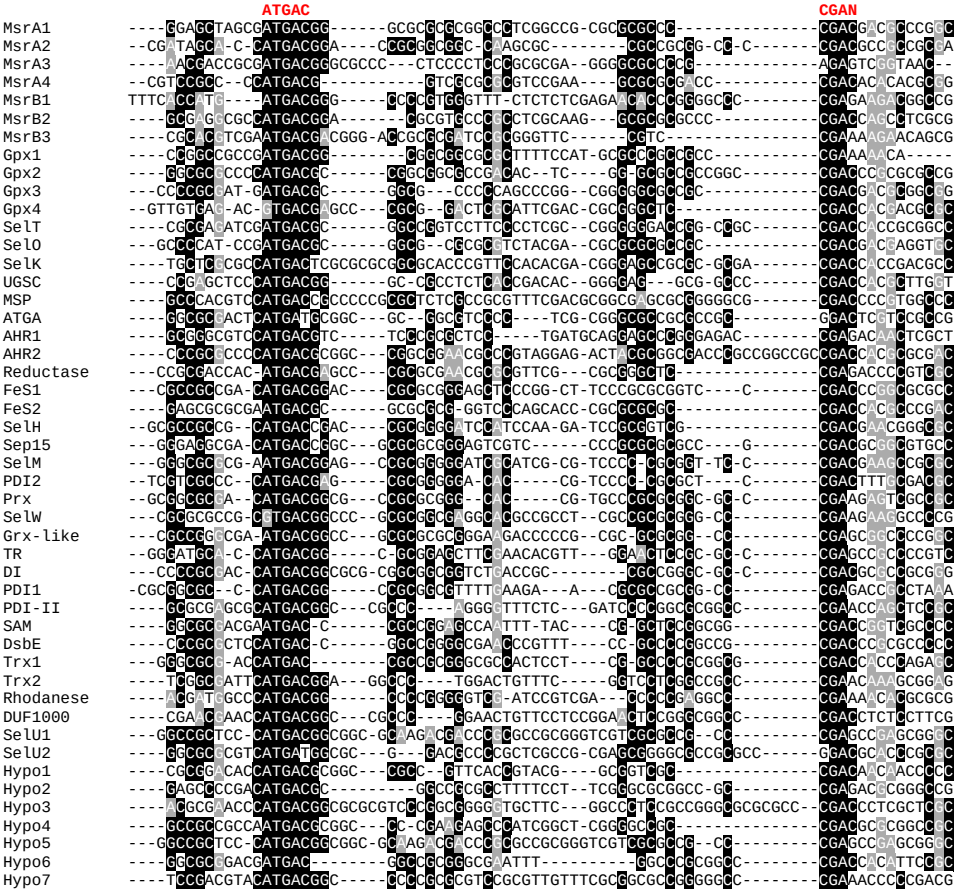


Figure 6. Metabolic labeling of *Aureococcus anophagefferens* (A.a.) and *Chlamydomonas reinhardtii* (C.r.) with ^{75}Se. (A) 20 ug of A.a and 50 ug of C.r proteins were separated by SDS-PAGE and selenoprotein patterns were visualized with a PhosphorImager. (B) Coomassie blue staining to show protein load. Migration of size markers (in kDa) is shown on the right.

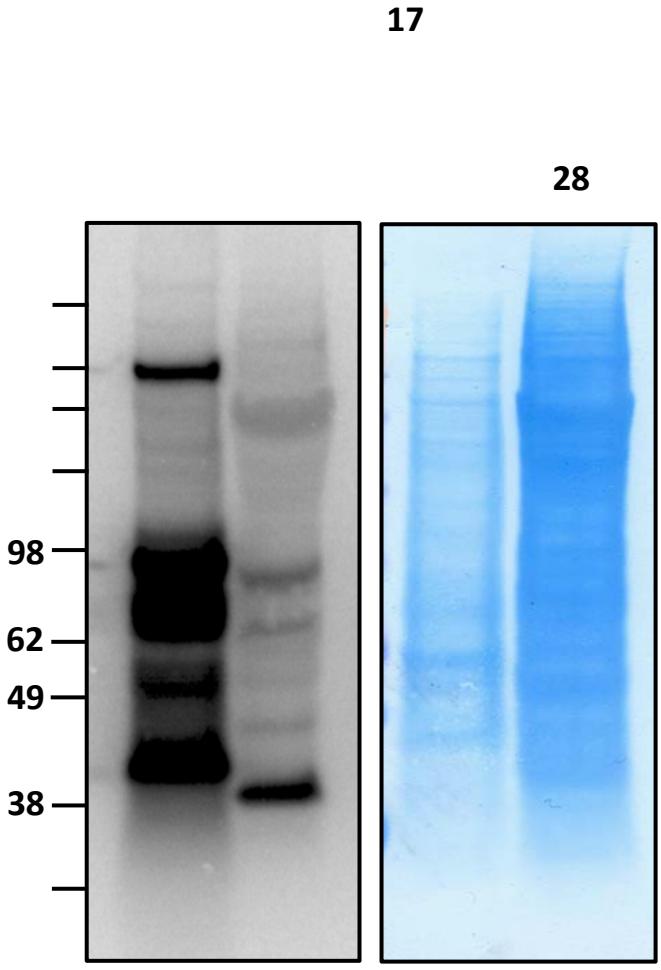


Figure 7. Temporal dynamics of dissolved Se and *A. anophagefferens* abundances in West Neck Bay, NY, USA, during the spring and summer of 1998.

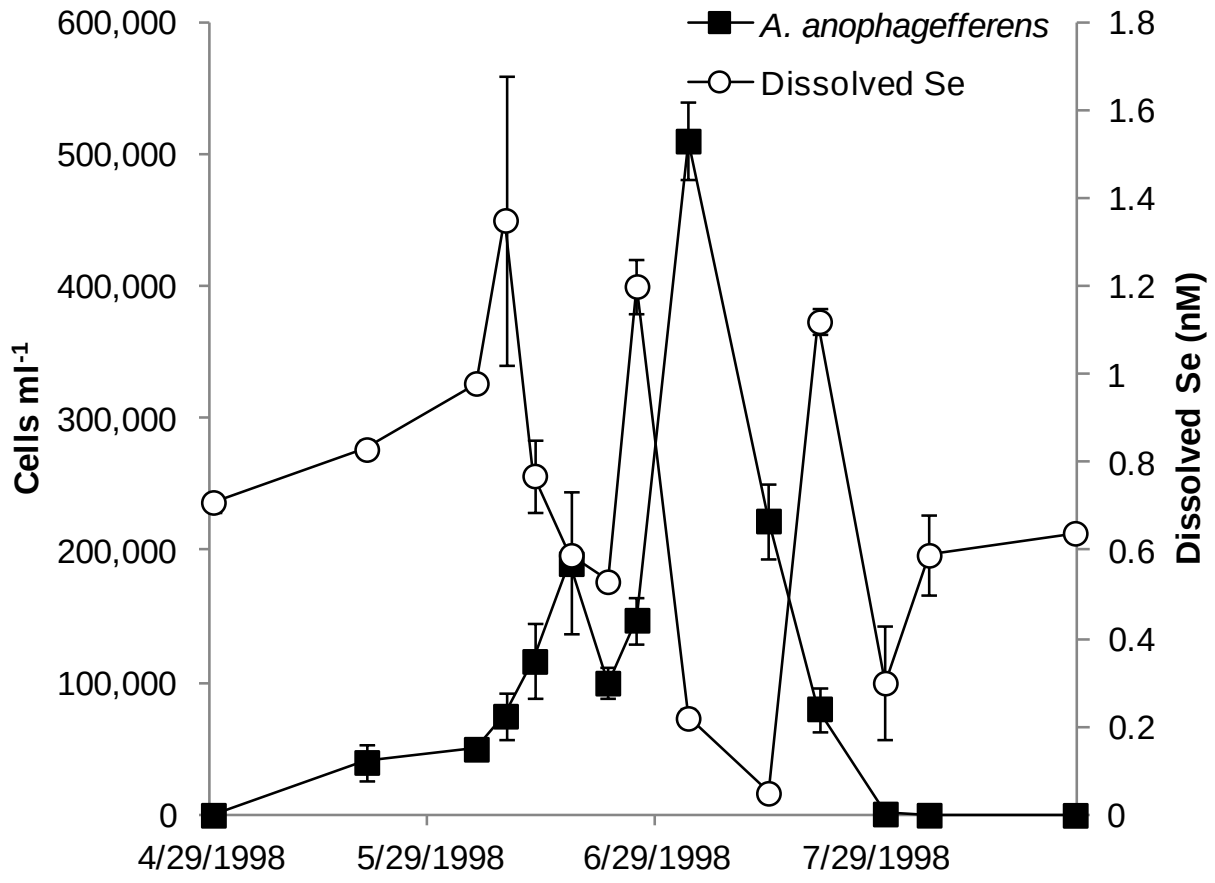


Figure 8. Net growth rates *A. anophagefferens* during incubations experiments conducted using water from in West Neck Bay, NY, USA, during the spring and summer of 1998.

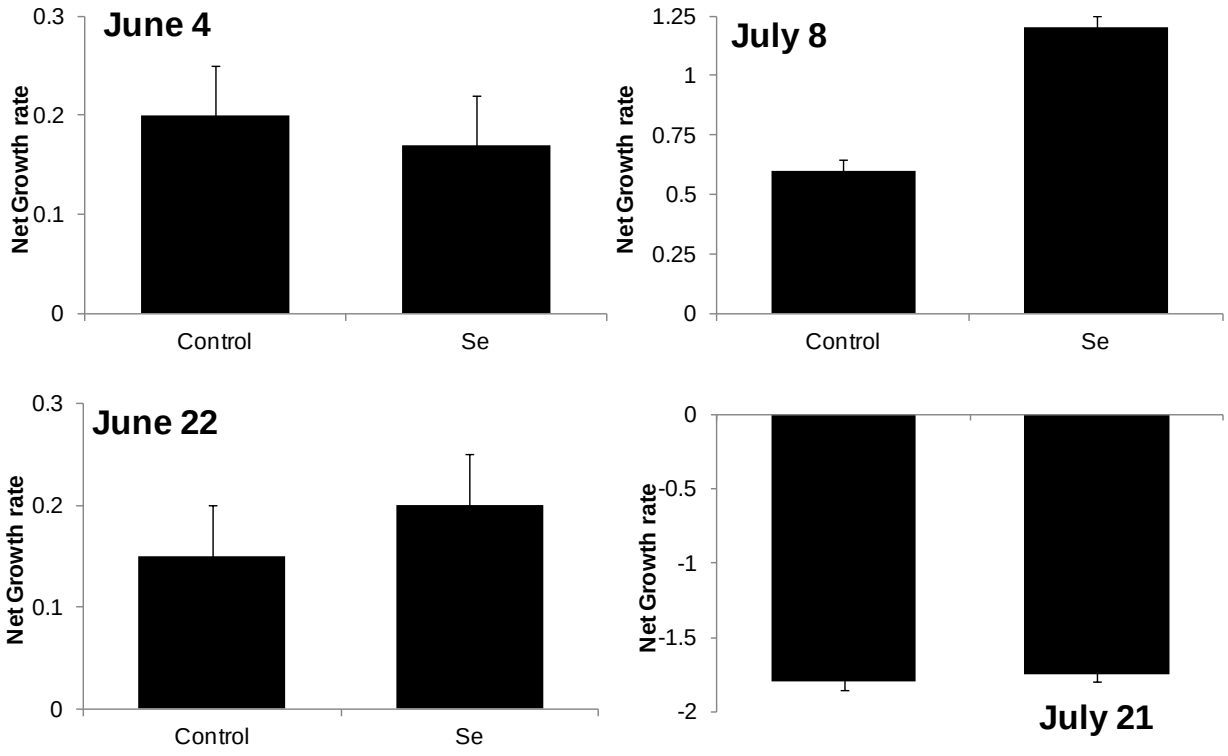


Table 1. Selenocysteine-containing proteins in the *A. anophagefferens* genome.

Selenocysteine-containing proteins	PID
Thioredoxin domain containing protein	77962
Thioredoxin domain containing protein	78111
Thioredoxin domain containing protein	78110
Thioredoxin domain containing protein	78109
Thioredoxin domain containing protein	78108
Glutathione peroxidase	77994
Glutathione peroxidase	77995
Glutathione peroxidase	77996
Glutathione peroxidase	77997
Glutathione peroxidase	78003
Methionine sulfoxide reductase A	77998
Methionine sulfoxide reductase A	77999
Methionine sulfoxide reductase A	78000
Methionine sulfoxide reductase A	78001
Methionine sulfoxide reductase B	78002
Methionine sulfoxide reductase B	77940
Methionine sulfoxide reductase B	78004
Methionine sulfoxide reductase B	77942
Glutaredoxin	77970
Glutaredoxin	77969
Glutaredoxin	77966
Peroxiredoxin	78116
Peroxiredoxin	78117
Peroxiredoxin	77974
Selenoprotein T	77992
Selenoprotein W	77920
Selenoprotein W	77991
Selenoprotein W	78120
Selenoprotein W	78121
Selenoprotein W	78119
Selenoprotein O	77993
Selenoprotein U	77989
Selenoprotein U	77988
Selenoprotein M	77987
Selenoprotein M	78118
Fe-S oxidoreductase	77968
Selenoprotein Sep15	77986
Selenoprotein H	77985
Iodothyronine deiodinase	77984
Thioredoxin reductase	77983
Selenoprotein K	77979
Methyltransferase	78115
Methyltransferase	77963
Thiol:disulfide interchange protein	77965
Fe-S reductase	77972
UGSC-containing protein	77971
Membrane SelenoProtein	77964
GILT superfamily protein	77961
Rhodanase	78106
Protein disulfide isomerase	78005
Protein disulfide isomerase	77982
Protein disulfide isomerase	77977
Hypothetical protein	78107
Hypothetical protein	78112
Hypothetical protein	78113
Hypothetical protein	77980
Hypothetical protein	77978
Hypothetical protein	77975
Hypothetical protein	77967

Supplementary Table 1. Sequences of Sec-containing proteins encoded in *the A. anophagefferens* genome.

>SelT
...APVAAAKRAAPAAGGARANGDAEVLVKLCTS**U**GTQRNYLELRKFLEDAYPGLRSVAAEQYPPPAVGVFAAQAAG
MAQVACVALLLGGEKVFQLFGAATPGWYHSVAENKMMAFGAVWMANNVAAQMVATGAFEIHVNGELAFSKLETGRLP
SAADVVRGMKRAGFATDLEARYVGDAEF
>SelW1
MRAPLLFVLATAQLCDALSATTPIKLHVEFCQQ**U**KYSLKVDRLETMLKDRFGDDIVITGRNDPDQLSRLSPSGEWRV
GAFEVTRLGDDELLYSKLETKRHLVGTRIKVDGEVVHDDKPFNEWVEKWGF
>SelW2
MGSLFSRAAPVKIEIEYCHT**U**SFKPKAAWFGETVLARVATLKSEGFAVEATAEQIEQKGKMFVRVDGAQVYSKGGFR
AENPQPGPEEVDELVTSLTAKAGKALPLSDDEKAAFLELHEAQIKAGKSKH
>SelO
MLRLILLAAAARGARSSDDAGEDLEELAARVTDDFAAALDVDPDAAANAPNTAARDVRGGHFVRPLALSPLGAPRLL
AFSNATLALLGLDAAEAARPEFLAVFAGGDPRAPALAPLHAAPWATPYALSIYGSEQVTDGGVGDGYGDGRAASLFV
AGAWEAQLKGSGPTPFRRRGDGFAVVRSSTREFLASEAMAALGVPTTRALALCASASDVAERPWYSDRTAVDDDFQP
VRHGGDVLRRERRAITTRVARSFVRVGSLELFARRLRRTGEPRARRELEALFRYAYDREGLGPAGAALEDAAVAAAT
ETRGRFASLACHWLRVGYAQSNYNSDNSLVGGATVDYGPFGFVERYDKTWAMWVGSGDHFAFGNALEAARRNWATLV
DALRPLAGGRDAALDAAVRGWPAEARRAEGAMRAAKLGLVGAAALAAATVDGAVRWDRVDAPRGAAALGLWARADAL
LSRGPPYADHTLFWRLLGRAARAADGAAALGALGGAWYDAPDADLAAAWRAWLDDWRALGPDAAAMDAANPAFVPRE
WMLVEAYEAAERGDGAVVDRLAALFRRPYAEDQAQADSEAYFRRAPDDAHLTGGVGWMS**U**SS
>MsrA-1
MGSSVEGGASAPAVRRRYTVVGKRGVVVRAGVELDSPVVARLPRGAVVEGDASSSVPARPGGARLRVAGGWCAARLV
APGPLVGGCAPLAASLGLPPRPEGYLCDLEAAPPPHERLEVATVALG**U**FWEPPALLRATHAGVAATTVGYCGGRAPW
PTYGDMGDHTEAVQVLYDPGVTTFDDILRFVLPSGGRNRSQTRQYMTAVWPHTDAQAAAVAAATRDLPEDFLFVEPF
SCFYRAEDPPPGPELAMTGARGPRPRAPRRRPAP
>MsrA-2
MARRAACACLVAQALALAPSRRVATFALG**U**FWTPQQKFDLRDGVIWSTVGYTSGTTPAPNYDDLGDHTEALRLEFDE
DVVSYEALVEQFWDLHAPNKGSRQYRSAIFYHDEAQRECAERVQAQLFPSKRYVRDTAIEPAAEFFPAEAYHQKYRE
KLAEDPLPFPFSLFSR
>MsrA-3
MAAPAAAAPRKSDPRTELGLGPMPPGHPCPPQPIGALSDRRNVATMAMG**U**FWGPESVFASTEGILEHRVGYCGGKEA
YPTYEAIKDHTEAVQVEFDPEKIKYEDVIRRVLDEAPREPAYSTQYRSAIFYHDDEQKRILERKVADMGAAAQYVAV
EPYHVMYRAEEYHQQYLSKYGGGYTGGYGLM
>MsrA-4
MHRSKEYQRNSPQVALLLAAGASALKTPCPLYAPTSPRVCTLAMG**U**FWHPQGSFDALAGVEKTLVGYTGGDDFEPTY
QRIGDHTEAMLVEFDESLAFEDLLVKFWEEHTPMVGSRQYRSAVFFHDDAQRDAAESMKAALIADGKKWVKNTAVEP
AKDFWQAEEYHQKYLAKMLTPRYYSGEDRYA
>MsrB-1
MELLSDLSTWGERFLDFWDRGVYACARCERPLYASDDKWSGPCLWPSWRRDLPGAARHAPVEGYNGYVVDVHELYCG
GCDLFLGHAFADARDKGDAHPDARWRH**U**LLSLSLEFVPEGR
>MsrB-2
MALPPSRSDGTPFPCQLSEEEYKAKLTGSQYRCLRQGGTEAYRRGEFCNFFPEDGYMACAACDIPLYSAKSKFADPG
WDAYASCYWTGSTCHVGVRPDGSALENFCNNCGSHLGHVFFRDPHSPAPTKERH**U**VNSCCTKYVKGPPPPGLTEKVV
FVPSNMQSFTK
>MsrB-3
MGNPTVTAKVVAKKGCKIRSERSLKSTCVALLAPGTLLSIEGCLDDARVATISPVAGFASIKTLRIVDDASPEDLIR
SAWRHFHPLRSDDELAKTLDAEAVRCLRAHGTEPAGTGEYNAFEPPGGRGRFDCRACRFPLYHASRKFADQGWIAFD
QCFFTGDVAHVGLAAGNMDSIEVHCSNCRSHLGHVFTDGVSGERH**U**LNSCCVAYVDAPPSHDYGQPLRQGSKVAAIS
DLLKTMSS
>SelU-1
MRLFYSALAVGAAGLAAKRMSTAKRPVVTAADLGSTMLLPVADGAVTGAAVPASSLWSKAGALIFVVRRPG**U**PLCRE
EARGLSELVKIRGSRQNFVGIIKEVAPTSQAANDTVLGVGEFERQYFGGGAVYLDEGKQFYGFLGDRKLISTRNVFK
ALVRPLKTWRSLRAVGDRMKSKGIEGNMVGEGLLLGGVLVVDRDGRVTYSYPETTGEPAPVDAVDEALDALDD
>SelU-2

MVKAPLAPPGGAMLVDLTSESRAPFAASKLWEQDPALIVVVRRPGUQICRAEVKRVYAQKQAFEALGVKMSAVVKED
LPGEDGKPGEIAAFREGFWPDLPVYMDERLAFYEAIAGGAVTKTSLATFLMKILNPWSRLKANTKRVPKTVEGNLTG
EGFVHGGCYVVRGKTGDVVLAHHEAEIGDHPAKGELLAACRKAAGMD
>SelM
MLRRLVLACAAALAAAATGKIETCSAUKLNKLPEVKRFVKESGHADTYEGLEIDYVRGKPPTLIMMDGDAEVERVDL
APYSTDELHALMQEKGFARKEAEPAGTREQVLHRAPRSAQADAVADAETTL
>Sep15
MARALVFAAALGSAASLDCLEAGFTASLRCSSCAKLESLVPDPELAADCGACCSEDLSASASYASAILEVCEUKLGH
YPAIGDFIERRSDTYGSLDVRYKQGQKPRLLMVDDDGEVAETVPIGGWNEDTVAEYLDDNLRGAVAEA
>DI
MEAAERSAASDWIEVARAAQRRVARESLAPGSSEAAVDARVDALRGAALRHPEICHWVRHNRARRGDLRVGDAAPDV
ALWRLDGSKTSLLAGRDAAKPLVVVAGSVSUPPFRRVGPELDALARECPGADFTFVYVAEAHAADEWPLRSARFAAA
PVVVDQPRTLADRVALARRYAADYGIASPVVVDDPTDEAFERAYAPWPLRLYVVRGAALAWIAEPDGASYERSVAEL
RAMLRA@RPRAPATMTARGGGLTAAGRPTRRGARPGASFLSRPLSCVPRPKSFVPP@NLGRVGHVEVVLPLRVDGLL
PRDEEALGVGRLLAAAAP
>SelH
MAPKRKASAKQDSEPKKAAKTKADAPPSEGGVVVVEACKSUGAFKTRANKVLKALEGKAEVKINEEKPRKGAFVVSV
GDTKVVELLDMKRPFPALKALDMDDVAADVEKALAA
>Gpx-1
...GQVVLVENVASIUGTTVRDFTQMNDLTEKFAGKFTCLGFPTNQFGKQTNEKDWELLPMLAHVRPGGGFVPNFPI
FTKTEANGEGASDLFKFLRSSLGAPSDDFKGQGSDYVISTKNIIWTPVTRRDQRTDLAWNFEKFLINQEGKPVKRYS
PGFLTADVAADVEALLEHGPDALG
>Gpx-2
MGVFASKEPPCAHVGGFHDLMSEDIDGAEVRFSAFRGQVVCVTNLASRUGKTNSHYKGLAKLGAQFSGRGLAIVGFP
CNQFMGQEPGAESEIKAFARSRGFLDPALGDASAVGAFNLMRKCDVSGAAVDPVFSYLKREAPCKISWNFGAYFLVS
RDGAVEAHENAHPRDLAPRVEALLGAVDAADVDVAAD
>Gpx-3
MLTSCKDACAAHHVASSFYDLKAKTAAGATLTFDAFKGKVVLLTNVASAUGKTAAHEKSKTAAHYSQMVELHDELGH
EGLEILAFPCNQFGNQEPATCEVVAEFAKSAYGADFTMMEKVAVHGPRAHPVYNFLRATTGATPSWNFGVYFVVSRS
GEVKAYPDVAPAQLAGVLRGEL
>Gpx-4
MARRAATTLALATLATAFAPMPATARSAPVRAAFYDLADVANDGSEVAFSQFEGKVCYAVNVASAUGATRPGYKLMK
QLSDEFGDDLAVIAFPTGEFGGQELPTDEAIAKFAYDRVKFPAAPRGVLLQKADLGRASWQAMQAEAGADAPAWNFK
GKFLVSKDGAVTDASGVDDIAAAVKALL
>Gpx-5
MSFFDLKGRTNAGEAVDFATHKGKCVLAVNVARLUDVTSPRTRIRPRAPTTTRRAPVRPKSKHEDTLERTRTRASES
RARSCPDPAVTS
>TR
MVDDGFTKEHGFDYDVVVVGGGSGGLACSKECAKLGAKVAVLDFVKPSPAGSTWGLGGTCVNVGCIPKKLMHTASII
GETLKDDAAAFGWKDAAAGAHDWESMKSTVQDHIKSLNFKYRVALREAGITYLNRLGEFEDKNTISLTDKKGNKTSV
TAARFIVAVGGRPNKLGCPGGEHAIDSDDLFSLPKAPGRVLCVGGGYIALECGGFTAGMGYPTTCMVRSILLRGFDR
ECVGKIEAHIKHHGVQLQVGVTPAKIEKDAATGVLTVTDSGGGVHEYDTVLVATGRYADTAALGIDNVPGASANVNA
KTKKLECVDEQLPGAPHVYAIGDVVEGRPELTPVAIEAGLRLARRLFGGKTEPMDYECVATTIFTPLEYGTIGLSED
DAKAQLGECNVESYISEFAPLEYALSETRSERGDGCFAKLVVDKSTGKVVGFHYLGPNAGEITQGFSIAMRKGATYA
DFISTVGIHPTVAEEFTSMTVTKSSGESAAKGGCUG
>PDI-1
...VTIELISDTMUPNGFIGKRHLQNAMAEVHANPLGAEPLGFTVLRVPFFLEPDYPEGEAFEETNRVRLVRKWGGQ
AGWDAQKKRHDLKGRGRAVGIEKFNLDRVASNTLTSHRLVQWVTKHYGVNAAEKMYADLNHRHFELGKKLNDRAMLL
EVAVAAGAELSKAMDFLDDPDAGREEITAAQAKLRELGVSGIPTLLLGGEWQLPSGALHADDIVPALRMVEERGGAT
GSFFAETLGISDAVMEETLHFAP
>PDI-2
MCRGEVTKHAKEVLNLLPGSWKGSPRPRTLRVEADWVTCTQUPHGFIGKKNVQLGLNMFHERHPDVGLLLNVTRHPY
SFNGDSRRGSLAGHGTKEGNEPTWHDSLLGYCGGDAARRAAAEDGMRMLGRRAGIELDYGVQTNWQPIDSQRSMLWA
RRFGLAEEFMDHLGHRHFEQRKSASHRATLLDAAAASGLDASALGEFLDSDELVADVWTSYGSTIHEKGINQRPVAL
TIFEDFSGKRRKIVTDMLGRHQRHPLLRLRARRHAVAVPERSPRGDDRQRLGQPGRVPPGLRGPAGPRAAGGHRRPV
RQAAPERLQGPRHQHGRLQREGRPRRAPRGVPLSGARPFVAP
>SelK

MVYLRADGTVVEKRSMWRLSIVPDFLWGVVDFFWLFVSTLINPQSEHNSLRPKRSSGGATGRGGGGGRPGGSGGGAR
KVKGLKDCSSSSMSAAGAGGUG

>similar to *O. tauri* hypothetical protein 3

MMRAAALLLLAATADGFLHGKTANKPLLRATESFENPVPLLQEAIDRAPQFDEGRSVVFGVLSTEVADAPAPAEQAR
RRELAAASLTNIDDAERDRRGAVAVVGGAATLAYAASLVANHASFAGRLSVFPLLALSLGYYESKKEGLUNIAQSGL
WDVTGNGMEKIADAKLARALLDKVNDFNVKTGLKGVAFTLAFAATAFFDV

>SAM dependent methyltransferase

MLGYDGVELEAGADLGLGCGNPLIAANLKPGEVVVDLGSGAGVDCFAAAKLVGPSGRVIGVDMTPEMLSRARAAAAP
YGNVVSFRLGEIEHLPVGDGAVDCVISNCVINLSPDKPAVYREMNRVLAPGGRVSISDVLRTSDIPEALKTAEAFAC
UVAGASAEDEIEAMLVAAGFVDVKIAVKENGRDIVKGWIPGSGAEKYVTSAYVTATKPRSTHGFRDAVLYGLGAPPA
PPASAGGCCPPPAAAACCPPPAAAAAAAPPAAACCPPPSGGGSAKPAA

>PDI-II

MSSLDAAYAKWADLDLDTASEDDEPQRPRAAPSKKRSKVVVDIVSDPNUYWAWPSKRRVEAVAAEFPDVAFEIRFQP
FQLYPDLPRDSRGVDKLEYFTALSERRRPNASMAEKKARIAGLVGAWKADGLDLTSPFGVDGGRWGSSFDAQRLIWL
ARQQGREDPMIEAIYKLNHVDNEPLSDHAKLLEAARAAGVQRAPELLARADGLGAREVLERYQHYVAMGINAVPVIV
LDDKHVISNGAPEKDFLRRTFKHLIDTGTVPPHLGA

>Peroxiredoxin 1 (alkyl hydroperoxide reductase)

MLRNSLRPLARLARPGGRAFAAVGDAIPDVSLDVEFPPKEVSLKQRLAGRKVILVGLPGAFTPTUSETQVPGYLGAA
DALKAKGIDEVIIFCVNDSAVMKFWAVDQGVEDNDKCYSMRYVNRSEQKVSFMADTSGALTRALDLELTHENPVRDL
GPGRCKRFAAYYDDGVLKALEVSEAPGDPAGDDDPHASCVDNMLAHV

>thioredoxin fold protein 1

MRVSRFALCSLGVATALRPQTSSFARRAARAASTAEDCGCEVTYGGDVPASAKTMDHLAAVRDLEAFDAATGGATTL
GAKLGGDRVAVVSFLRSFGUPFCQELLVQLERRRPALEAAGVGLVAVGIGTPEKGRLVADHVGYDASRLLADPENAL
YDALALNAGVGRTFFNPATPYAILDRLQSGTGGDLGDVLGKWLPGGGPSGEGAFIIPPKQAQAFNQGGMFVFADGAS
RFVHYDASTGAHADVDAVVDLALSL

>thioredoxin fold protein 2

MRRFALLIATATALAPTAYDRLAGAQLKNIATKQPVALTEQWNSDQKVVVEFLRHFGUVFCWERVMQLQRDALPALN
AAGVKLLVVGIGSVESGETFAKQTSFSPELLFVDDSELSDAYAAAGTRNTKRDDSGKAVFEGVGSMWSAATNDAIKE
RGRDDLNAVTGNLFNPGPYKPLMPKASTMRRSMEQTMVQGGSFVFDGDSVVAEHYDESSGAHLSIEALLAAAGAA

>Prx 2

MWRLAPLVAAAGALSVGDVAPAFSGTSFDGRTIDLSYAKQTEKGLIIWFYPRAGTGGUTKEGENFERLREQFKEAGY
AIAGCSTDPPEFNRKFAEEHDFGYPLISNGDDVIKAFDSCKQAACSSAKRTTVVISPDGLVVWREDPFDAADGPDEL
LQRLTGQGGASPSKKRKKHLRAHGAHESAHHSGEHRNARA

>Trx2

..EPLRRNRRPTCRRRSRPSPITRPPSNPARTRALRRLERARARVSRSPHRAASARVPRAASKPRARVPRAASTPRA
TPQVVILFTASWUPPCKTFGPKYDARAATAAALLCKVDVENPETQSLCQKHGVSCMPTVVVIKDGKEAGKMEGVDQA
KFEVWASGG

>Trx3

MKLVLALLASAGAMELNADNFKTEVFESGKNAFVKFLAPWUGHCKSMKPAWDSLMKEYDGHPSVLIGDVDCTVHNDL
CSEAGVSGYPTIKYWTDGADMAGASPYQGGRDLAALQKHVEDNMLPKCDAKDPENSGCDDQEIAYVAKMTEKGADAI
AKEATRLEGMKGSAMKPDKKAWLLKRINVLKGLSA

>Trx4

MRSAARTPRTQRLTLPPPPRFKTILALALASAAATELTPETWDDAVDGKTIFVKFLAPWUGHCKKMKPAWDALMADY
KDHPTKLVADVDCTAAGKPLCDSNGVKGFPTIKYGDPADLQSYEGGRDEKALKAFAETKLVAMCSVKNQDLCSEEKK
AEIAKFMAMEVGDLNTKIEEQEAELKNIEADFKKSVEGLQKLYEGYTKEKEEKMAALKDSGLGLMKSVKAAKAKAAK
DEL

>Trx5

MAVLKFLLLATASATLELTGDTFDEAVLSSGKSAFVKFLAPWUGHCKAMKPAWDALSQDYEGSKTVLVADVDCTAAG
KSLCARFKVSGFPTIKYFNPPNHVGEDYDGGRSLDQLRAFAESELKIACSPTSKASCSKEELAQLEKDLAIPAAART
AELVKIYEAMDADAEKHDAYLEELQARYDEAEEALEARNAAVEPRVKQLKAAGTTLPEGFKRRDEEDEDEEDEDEED
DDDDDEGDADGDDDDDDEGSDEL

>Rhodanase

MLEGFSWAGDVCVLDGGLRAWTEAGFAVDAGSAPAPPPAASPPPPRACRPGAFVDASRVLAAIGDVALVDTLKPAAF
AGAKASRYGRRGHIPTATNLPYTAVVDSSSGRFFGADAIRAAAAAAGLRPGAPVLAYUGGGISATAVLFALVEILGE
DPALLSLYDASLSEWAADDAYPMVCPADDGP

>hypothetical protein, distant homolog of Trx

MQGGYEAVDSPRAAFRSRSRVVGAATFLFAAGALAHLGRTAQRPGAAELLDEVAASSGKPTVHAYADALUSDCRRHA
LTLHELMLSKSDVLDAIDLKIDYIGAAASVEEHGITLIPVNEYMLCAEELLDLPQSEWWPFTHCAFEIQKCMNYVSR
EEAGMTSCDEADSGSDDAMALAGTERDLSTCTCTLEGVVEYCAETHTSTTLAKLATCKDSDEASKLFKTSNAVADAI
NSGHPLWVKINGVEYAGPSDPDETTATMDAWAEQVLNVTCQSLGSAVSSCAPYV
>UGSC
MICAGFGAGKPGMAAHRMSVPVSRPVAALEPGTGNLPDGAVLDPRGATPPPLALAPRSGRLEGAVGLLDISKGGGKV
FLDGIEAFLRTAFEGVEIRRYAKPTFSRNCPAGLLDDVAAACAHVVVGLADUGSCATCEIHDVVALEARGLASVAVL
SDAFPSNARAQAAQLGLPAAPAVFVAHPISDQTDAQLLARAEAIGPAVADALTTAAPAIPAGAEDAACGA
>Grx-like 1
MLKTTCTEHLRAMMLATRSRSRYTMLALLLLAAGSAALQPATTTTLWNSGVUPFAQRAWIALLERDIPFTHELVDLS
AKPPRLVELYEQARPGSKGAAKVPLLEHGGDVVVESVDVCRFVDGIEACAGPRLRPDGAAELVDAFLDAWPRVEAAY
YAVLTAKDQPGADAAVAGFGATLGELEALLEERGGPLLLGADVSLAEAVAAPWVQRWAVTLPRFRKVDLEADVLAPR
GLARVAAWARAVAARPAVVASAAPREEMIAAAERYYVTFQTR
>Grx-like_2
MLMRRIASLSFATSLGRARGLGRNAHKSCVRRFAAGKQGGAFPPGPEQWDSWTEIGFGVWAPTALCATFWNSGTUPY
AQRAWIAMEETRVPFVTETVDLQHKSEAFLAKYEEANPGGRAKVPIVEIDGLVLTESAVVVEYLAEKDGPGPFYLED
PARRAAARLMAEVHPFGDYFKFLKLRDDPEALAAAVADLTGKLEVFEAFLVKHGDAAGPFFNGEDLCFAEANLAPFL
QRMVPTLKHYVDVDVRRLCEPFPRVDRLVSAVLARWTVRKTGVPEDKLIEGMDKMLARIAAQAPPP
>Tryparedoxin-like protein
MASQALAACAPPAACSIYDADAASACVMWSPSAGAASYAVEMRAAGDEAWTTLSTTLKACKCRKKGLDASRAYEFRV
AAVDAAGAVGEFSGIAKVASRADVAQQPAPVVSAVDGESVTVQWSGGEGPYALQFAAEEDLRADATTPWRLAAASVN
GTVAKKKNLDGGKRYAFRVKPADASGYEWSRASETALIPRAAPFMARCFGTDLVNRSGGTKTTGPSLAGKIVAVYAS
ANWUPPCRNFSPNLVRDYNVLKAAGRPFEVVWLSCERDPEAFSSTFAQFPFLAVPFDNDERERALGNFNVSGIPRLV
ILGPDGRELVNNAVGMSLSIATVDGWMRQAKGHK
>SelM-like
MLRSILLAAAAVGARGGLFSSSERTCEVESCSTUWTQRLPVADQFVKNEDGAESYGCTVNYIPHHPPELVIKANGRE
QERISLERFRNPNDIRRLMESKGFAQTDEYFKKQKDKRDVEFVAGQEADVYWLDGDGGKHHSGIAEVGAPLMVSTYV
GHNFEVQDGFGDPIIMYTVSEKARQRVSIPAETPEL
>Grx-like
MLRRLVFALALLAGAAEAFVAPRASSSARALHAEDPTKVWYAEIANTVQNLLTNSPLNEGKKAVVKMLAGPYDEVAV
RAKLDGLIAGEPVLMLSFVRUPFCVKAKQLLDAKGAKYTVVELDEVSDGPALRAELGGLVGRTSVPAIWIQSDFVGG
CNDGPGIITLDKQGELDARLKAAGAL
>thioredoxin-like 1
MEPLDERLAKVLPAHFKAEKALVVRKNIHMTSNVVTEVAAGAVVKLDRVRKLPGDPKKLRGRVVVAPAAGWVTVTSK
NLTRVDASAAPADDREDWEYDPRDSYDHYLTRAERCAWPAAKQRAKVAKLLEKEFPTVKLPAAARTSWSVSRMRAFL
ESKGAKGRPHAVVGVAAGGDPWAAAQAEYRLAMEGVEREALRAVERDNREWCKAWDEDRFRGAFLRPELASRVDDCR
RALRFASSPGLERWLREGGKPNWDGPRDWLPPNGGWGDPLPSKTENLLGRLKSGTRRVERLDELTGTWHGFRRSYLP
GAKPVALRPYDFRTYRHLASEWRRDNLLRRSGALPVEHVKAPYDGLLPRLPGDKDAPEPWEKKLDERWTTLRDFADA
HLPGPEAATPYPEVDDAAAEALAGRCRHRPCVCQCLSADVDVDDPVFSVFASLPAVPPHFEAPTQHLPYQPTNRMAA
HLAKCKPPILLAAHLQSDDDKTEYESSTDDGEWQPEYDTRLETPDALYPGRFELHVLPALGGERLSQRGGAAWHVLV
HGRVRWCLLPPNGGRYVKRRYHELQRAPPKPAASCRHQDFVSGLDWFRSEFETLADPAARVPLSEIVQEAGEALWVP
EGWSAARLALKPSVGFALDVGSLNAPGPRGNPPGIMDDDRPARPSIFTKHTDPEYLESDYSWIPDDAFLKQLEGLNT
NLKKAMDEAAEKAAAKWAEQKFESGEWIRPENIAAAAAEAVRQVDELLPPPPPDSIAAAEPSEMLKRWAASPFHPDQ
NPRLVCQSEDDEQIIINIVFREAVCIHSINIVAPPGEEAPKEVKLFCNQPSLGFPDCEDGPCAQALSLTAEDLAADR
VNELKMAKFNYVNILTVFVGETHGDSVCSISSIKLNGKTRESTNMANFKKVGUG
>Alkyl hydroperoxide reductase (peroxiredoxin)
MSVKAGDTIPSVVLCDGMQFPPKLVNVADEIKGKKVLIMGLPGAFTPCUSGHQVPGYLAAEEQLKAKGIEKMYVFCV
ND--
VMKAWAKDQSITGDGLITFWADNLSVLTKALGMAISHPGPSGDLGPARSKRFVLLVEDGVVKFVQLSEAPDDPAGDN
DAASPVAAETMVEKILTLL
>Gpx-like_3
MARRAATTLALATLATAFAPMPATARSAPVRAAFYDLADVANDGSEVAFSQFEGKVCYAVNVASAUGATRPGYKLMK
QLSDEFGDDLAVIAFPTGEFGGQELPTDEAIAKFAYDRVKFPAAPRGVLLQKADLGRASWQAMQAEAGADAPAWNFK
GKFLVSKDGAVTDASGVDDIAAAVKALL
>Fe-S oxidoreductase
MRPATLRAPLARRRPPVPLQVARGLAAHSPPGARRLSTRSLPDYNTPVDVANIRLSKPKRPPPTKRRKAADGEWQGL
KDLRARLKAEGSMVDQTLIDMESDADFQRTLKAYEEKGAAKLTLEEKKERRRQLDALGVPDFFDFLGRELEEPSKAR

KRCEVFQLNVGLYCNQACAHCHVESSPKRFEAMSEAVAERCVEILAASPSVHTLDITGGAPELNPQFRAIVEGARRA
VGDRVDIIDRCNLTVLTEPGQEDLADFLAHHGVRVVASLPCYSSKNVNQQRGRGVFGRSIEGLRVLNEAGYGAEGSD
LRMDLVYNPLGAFLPPPQEALEAKYREELKREFGVDFDELFTMTNMPIKRFADFLSRRGELKDYLDLLVRNFNPDTV
DTLMCRNTLSVSWDGSVYDCDFNQQLGNPLALQDAPLASVFDFRSTDEFLGSDIRFDNHCFGCTAGMGSSUQGTTA
>Hypothetical selenoprotein
MSKGSKDRVQGALWGLLAGDALASPTHWYYGGLRQVQGDYGRQGIRDYTKPVAEMQGSIMPRSNTDGAGRGSYNEDR
TTVIGDVINHGKKPYWNPRTSFHYHGTLAAGENTLEAQLSRLLVRTIAARGGVDDAAFRDAYVAWMTTPGSHNDSYA
STCHRMFFANYAHGKRDPADCPDNDRHNVDTIDGLVLPSVAAVAAAYGGGPGARAEAVDAAVRVARVTRDSKPLERA
AAAVAGVLFDSVHGVGTMADRVDAATDALGMSRVHAGAPDEMISUYLQQSLPSALAMAKKYGGDDVFDALLVNANVG
GENVHRGAVLGALLGANKGRSRLPPRLVDGLAATSEIAAEIDALVAALAPK
>thiol_disulfide (TlpA-like)
MSGWAGKQVPSFDVTMVSSNGTIGAKMPLSAMLGKPLVIHLYNGGUGGCRPCAQQMDHWQRAYGAGALFLCVCVESA
SVAKQFSQMFFKAATNCFVDEWRGLFPAQLGCSGFIVIDAEGKFVTTKSMSFLDYRERAFRAVEQMIAPMLAAPEDD
AFPVGAEVVLRGLAKAPELEGAHGIVAQTPVDLKRKGRVAVTVGGRAGRTVALRVENVALVGDAPELFDEAGDDEDP
KRRKTAGGVRPLPKVGHAEMDDEHAELTALLETLARDRSLEALAAVRTEFAEHSAHEEALMAEVGFGWTGPLSALAS
HAKDHARIVALADKVAAKGIAKAVHLVAPEDIAALGDAIHDHALHFDQLYVDAVKRGPKVACAGGS
>MSP
...IVGSUCWPILVGAIFGVAAGQVPGGLVAMFHGIQVAIMLALMTNIIQFAYWKQRGKSQKPIYLLCLATLLVMVQ
PTSMLVIGSYHDAALATGGLCISQISPVTYPVTNTSVWRDCVSEPPFGNFFFDGDDDPNALTPNTSDRLVHPDLRHL
RRLPAHVRRRLRGHAARQEAQEALGRHPRSRGLGPIPAISAAPPPVVRRRARLVGGAGPRP
>Methyltransferase
MAAPSGGPDGVQEAYTTYAAQGGDGADNWATRAKTWGYAPAVLESLGLEGAERRLFAGACGGGCPLRVDGGPRPGET
VVDLGCGFGHDLVLASRMTRGGRIVGVDVTAAMLDAAARTVAAFSLPNVELRRAPLDDPSLGSGVADVVVSNGVFNL
TRDKRAAFAAAFRLLKPGGRLLLADVCKTRPAGSCAAGPRDEWSGUVSGCESVEAYLDHLRAVGFVDGAHLSWTAFY
TSPTTRGAHFFAVKPDPRRRLAAAAVLAAAAAVVAVIFARRR
>GILT
MIYWASCRECRHTEGCFSSAMLSMAKLLPLLAASASSSSSVEVAFFGEALUPGCQAFVQGPMKDLMMAEGIDDVLDF
SLVPWGNAYVASATCPTTSTEEGYALYNVTARECWNDRCGSEGFDCFAGEFIYQHSPSEGLADLVEACAVDAYPERA
NWWPFVYCFEGLKLNYCPCGEACDLDDCDLHYTGAPQIYLKDYAPSGTEDVPYNKSAADAYIADAGRRCAGTVGLDW
APVASCALADAADPTSWPGARGLALEKAAAEKTADLQPEHGGVPWVVVDGTPLSNTTALLATVCAAAEAKGLPAPAG
CAN
>Hypothetical protein (Thioredoxin-like)
MLEDVRGLSGPRRVCEPSSACSVHLVGFYYADULGCRTRAAKLVRLASTVAGACVVVLNNRMPMACLMNASAPCDTS
YGGWQAPMVAAVANGSATPVLQDDGAAWSAIGAGHDDVVLYDDACRAFAAAKKNVADVEDDAGFRAVAVLAGAAASA
AGGAGVCRETSCFASSAAAAAPAGAAGRRLLLFAIFLSSFLTGGLVVRFYVSRASASPASSDPVFSPLGFGAQMVRR
RPRSRSGSDDDDDDDFDDIAL
>PRX5-like
MAATQLKAGDTVPAGVTLDLGFPPSKIDLAEHVKGKKVILMGLPGAFTPTUSSVQIPGYLAAQDALKAQGIDEIVVY
CVNDGAVMTAWRKDQGCGEEAVGGSLLKFYGDPTGAFTRACGLVLDHPGPTYKGLIMRTRRHAAYVVDGVVRALEVS
EKPDDPAGDDFPESSCVESMFKHIAPGVALDVKLPEADC
>glutaredoxin
MGLFQSTTADHDPKQLVESKIALRDTVVVFSKSYUPYAQNTLKLLRALPGLRLEVMELDQMGEPRNGPVQQLLKARS
GRTSVPQVFVNDAFLGGNDEIVGLGAGLRSAVSAALPA
>GST family dehydroascorbate reductase
MLMRRIASLSFATSLGRARGLGRNAHKSCVRRFAAGKQGGAFPPGPEQWDSWTEIGFGVWAPTALCATFWNSGT&PY
AQRAWIAMEETRVPFVTETVDLQHKSEAFLAKYEEANPGGRAKVPIVEIDGLVLTESAVVVEYLAEKDGPGPFYLED
PARRAAARLMAEVHPFGDYFKFLKLRDDPEALAAAVADLTGKLEVFEAFLVKHGDAAGPFFNGEDLCFAEANLAPFL
QRMVPTLKHYVDVDVRRLCEPFPRVDRLVSAVLARWTVRKTGVPEDKLIEGMDKMLARIAAQAPPP
>Peroxiredoxin
MKLALALALLPAASAFAPSRSGARATTVTNIAVGDSVPADTVLVKEFPFDTIDVPARLAGKKTIVLGLPGAFTPTUT
GARSQVPGYMSKEDELKAKGIDEILVYCVNDAAVMEAWSDKMNVKPKSLVTMLADPGCKFTEAMGLAMPADEVPPQL
GYVRSKRYAAVFNDGTLEQLYVSYAKGDPAGDDDPSASLVENVLPKL
>PRX-like 2
MGMRQAVLALALLGAAAAFAPPSAARALAPLRAADSCGECDDWNPFGDGCKPCNDGKSVFVNEALVTSKVLRDVDVV
GANGQRAEMSKLMGDSGSVVVFLRHLGUPYCWDYANAWCQPSYLGSLRAAKVAGPIFISVAPADSAVEKMEKFLDLN
PLVPRKSLFVDDSPTFDAYASAGFKKIGDDTAGGMAAASKLQAPKLNAGAWWRYLTNVAALSPVPKDLKFGEIPEGV
LRLGGTFVVDDDKVAFAHADKFPGDHPAIADVLRAARINILDATGEVQNA

Supplementary Table 2. Sequences of SECIS elements in the *Aureococcus* selenoprotein genes.

>MsrA1
GGAGCTAGCGATGACGGGCGCGCGCGGCCCTCGGCCGCGCGCGCCCCGACGACGCCCGGC
>MsrA2
CGATAGCACCATGACGGACCGCGGCGGCCAAGCGCCGCCGCGGCCCCGACGCCGCCGCGA
>MsrA3
AACGACCGCGATGACGGGCGCCCCTCCCCTCCCGCGCGAGGGGCGCCCCGAGAGTCGGTAAC
>MsrA4
CGTCCGCCCCATGACGGTCGCGCGCGTCCGAAGCGCGCGACCCGACACACACGCGG
>MsrB1
TTTCACCATGATGACGGGCCCCGTGGGTTTCTCTCTCGAGAACACCCGGGGCCCCGAGAAGACGGCCG
>MsrB2
GCGAGGCGCCATGACGGACGCGTGCCCGCCTCGCAAGGCGCGCGCCCCGACCAGCCTCGCG
>MsrB3
CGCACGTCGAATGACGACGGGACCGCGCGATCCGCGGGTTCCGTCCGAAAAGAACAGCG
>Gpx1
CCGGCCGCCGATGACGGCGGCGGCGCGCTTTTCCATGCGCCCGCCGCCCGAAAAACA
>Gpx2
GGCGCGCCCCATGACGCCGGCGGCGCCGACACTCGGGCGCCGCCGGCCGACCCGCGCGCCG
>Gpx3
CCCCGCGATGATGACGCGGCGCCCCCAGCCCGGCGGGGGCGCCGCCGACGACGCGGCGG
>Gpx4
GTTGTGAGACGTGACGAGCCCGCGGACTCGCATTCGACCGCGGGCTCCGACCACGACGCGC
>SelT
CGCGAGATCGATGACGCGGCCGGTCCTTCCCCTCGCCGGGGGGACCGGCCGCCGACCACCGCGGCC
>SelO
GCCCCATCCGATGACGCGGCGCGCGCGTCTACGACGCGCGCGCCGCCGACGACGAGGTGC
>SelK
TGCTCGCGCCATGACTCGCGCGCGGCGCACCCGTTCCACACGACGGGAGCCGCGCGCGACGACCACCGACGCC
>UGSC
CCGAGCTCCCATGACGGGCCGCCTCTCACCGACACGGGGAGGCGGCCCCGACCACGCTTGGT
>MSP
GCCCACGTCCATGACCGCCCCCGCGCTCTCGCCGCGTTTCGACGCGGCGAGCGCGGGGGCGCGACCCCGTGGCCC
>ATGA
GGCGCGACTCATGATGCGGCGCGGCGTCCCCTCGCGGGCGCCGCGCCGCGGACTCGTCCGCCG
>AHR1
GCGGGCGTCCATGACGTCTCCCGCGCTCCTGATGCAGGAGCCCGGGAGACCGAGACAACTCGCT
>AHR2
CCCGCGCCCCATGACGCGGCCGGCGGAACGCCCGTAGGAGACTACGCGGCGACCCGCCGGCCGCCGACCACGCGCGAC
>Reductase
CCGCGACCACATGACGAGCCCGCGCGAACGCGCGTTCGCGCGGGCTCCGAGACCCCGTCGC
>FeS1
CGCCGCCGACATGACGGACCGCGCGGGAGCTCCCGGCTTCCCGCGCGGTCCCGACCCGGCGCGCC
>FeS2
GAGCGCGCGAATGACGCGCGCGCGGGTCCCAGCACCCGCGCGCGCCGACCACGCCCGAC
>SelH
GCGCCGCCGCATGACCGACCGCGGGGATCCATCCAAGATCCGCGGTCGCGACGAACGGGCGC
>Sep15
GGGAGGCGACATGACCGGCGCGCGCGGGAGTCGTCCCCGCGCGCGCCGCGACGCGGCGTGCC
>SelM
GGGCGCGCGAATGACGGAGCCGCGGGGGATCGCATCGCGTCCCCCGCGGTTCCCGACGAAGCCGCGC
>PDI2
TCGTCGCCCCATGACGAGCGCGGGGGACACCGTCCCCCGCGCTCCGACTTTGCGACGC
>Prx
GCGGCGCGACATGACGGCGCCGCGCGGGCACCGTGCCCGCGCGGCGCCCGAAGAGTCGCCGC
>SelW
CGCGCGCCGCGTGACGGCCCGCGCGGCGAGGCACGCCGCCTCGCCGCGCGGGCCCGAAGAAGGCCCCG
>Grxlike
CGCCGGGCGAATGACGGCCGCGCGCGCGGGAAGACCCCCGCGCGCGCGGCCCGAGCGGCCCCGGC
>TR
GGGATGCACCATGACGGCGCGGAGCTTCGAACACGTTGGAACTCCGCGCCCGAGCCGCCCCGTC
>DI
CCCCGCGACCATGACGGCGCGCGGCGGCGGTCTGACCGCCGCCGGGCGCCCGACGCGCCGCGGG
>PDI1

CGCGGCGCCCATGACGGCCGCGGCGTTTTGAAGAACGCGCCGCGGCCCGAGACCGCCTAAA
>PDIII
GCGCGAGCGCATGACGGCCGCCCAGGGGTTTCTCGATCCCCGGCGCGGCCCGAACCAGCTCCGC
>SAM
GGCGCGACGAATGACCCGCCGGAGCCAATTTTACCGGCTCCGGCGGCGACCGGTCGCCCC
>DsbE
CCCGCGCTCCATGACCGGCCGGGGCGAACCCGTTTCCGCCCCGGCCGCGACCCGCGCCCCC
>Trx1
GGGCGCGACCATGACCGCCGCGGGCGCCACTCCTCGGCCCCGCGGCGCGACCACCCAGAGC
>Trx2
TCGGCGATTCATGACGGAGGCCCTGGACTGTTTCGGTCCTCGGCCGCCCGAACAAAGCGGAG
>Rhodanese
ACGATGGCCCATGACGGCCCCGGGGGTCGATCCGTCGACCCCCGAGGCCCGAAAACACGCGCG
>DUF1000
CGAACGAACCATGACGGCCGCCCGGAACTGTTCCTCCGGAACTCCGGGCGGCCCGACCTCTCCTTCG
>SelU1
GGCCGCTCCCATGACGGCGGCGCAAGACGACCCGCGCCGCGGGTCGTCGCGCCGCCCGAGCCGAGCGGGC
>SelU2
GGCGCGCGTCATGATGGCGCGGACGCCCCGCTCGCCGCGAGCGGGGCGCCGCGCCGGACGCACCCGCGC
>Hypo1
CGCGGACACCATGACGCGGCCGCCGTTCACCGTACGGCGGTCGCCGACAACAACCCCC
>Hypo2
GAGCCCCGACATGACGCGGCCGCGCCTTTTCCTTCGGGCGCGGCCGCCGAGACGCGGGCCG
>Hypo3
ACGCGAACCCATGACGGCGCGCGTCCCGGCGGGGGTGCTTCGGCCCTCCGCCGGGCGCGCGCCCGACCCTCGCTCGC
>Hypo4
GCCGCCGCCAATGACGCGGCCCCGAAGAGCCCATCGGCTCGGGGCCGCCGACGCGCGGCCGC
>Hypo5
GGCCGCTCCCATGACGGCGGCGCAAGACGACCCGCGCCGCGGGTCGTCGCGCCGCCCGAGCCGAGCGGGC
>Hypo6
GGCGCGGACGATGACGGCCGCGGGCGAATTTGGCCCGCGGCCCGACCACATTCCGC
>Hypo7
TCCGACGTACATGACGGCCCCCGCGCGTCCGCGTTGTTTCGCGGCGCCGGGGGCCCGAAACCCCCGACG