

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Vadim Gladyshev Publications

Biochemistry, Department of

---

January 2004

## Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution

Sergi Castellano

*Grup de Recerca en Informàtica Biomèdica, Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra*

Sergey V. Novoselov

*University of Nebraska-Lincoln*

Gregory V. Kryukov

*University of Nebraska-Lincoln*

Alain Lescure

*UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 Rue René Descartes, 67084 Strasbourg Cedex, France*

Enrique Blanco

*Grup de Recerca en Informàtica Biomèdica, Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, Dr. Aiguader 80, 08003 Barcelona, Catalonia, Spain*

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/biochemgladyshev>



Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

---

Castellano, Sergi; Novoselov, Sergey V.; Kryukov, Gregory V.; Lescure, Alain; Blanco, Enrique; Krol, Alain; Gladyshev, Vadim N.; and Guigo, Roderic, "Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution" (2004). *Vadim Gladyshev Publications*. 66.

<https://digitalcommons.unl.edu/biochemgladyshev/66>

This Article is brought to you for free and open access by the Biochemistry, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Vadim Gladyshev Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Sergi Castellano, Sergey V. Novoselov, Gregory V. Kryukov, Alain Lescure, Enrique Blanco, Alain Krol, Vadim N. Gladyshev, and Roderic Guigo

Submitted November 14, 2003; revised February 4, 2004; accepted February 17, 2004; published online April 23, 2004.

Scientific Report

# The prokaryotic selenoproteome

Gregory V Kryukov and Vadim N Gladyshev\*

Department of Biochemistry, University of Nebraska–Lincoln, Lincoln, Nebraska 68588-0664

\*Corresponding author. E-mail: [vgladyshev1@unl.edu](mailto:vgladyshev1@unl.edu)

**Abstract:** In the genetic code, the UGA codon has a dual function as it encodes selenocysteine (Sec) and serves as a stop signal. However, only the translation terminator function is used in gene annotation programs, resulting in misannotation of selenoprotein genes. Here, we applied two independent bioinformatics approaches to characterize a selenoprotein set in prokaryotic genomes. One method searched for selenoprotein genes by identifying RNA stem–loop structures, selenocysteine insertion sequence elements; the second approach identified Sec/Cys pairs in homologous sequences. These analyses identified all or almost all selenoproteins in completely sequenced bacterial and archaeal genomes and provided a view on the distribution and composition of prokaryotic selenoproteomes. In addition, lineage-specific and core selenoproteins were detected, which provided insights into the mechanisms of selenoprotein evolution. Characterization of selenoproteomes allows interpretation of other UGA codons in completed genomes of prokaryotes as terminators, addressing the UGA dual-function problem.

**Keywords:** selenocysteine, selenoprotein, TGA codon, genome, annotation

## Introduction

In the genetic code, the codon UGA differs from other codons in that it has a dual function: although it most often terminates protein synthesis, it also encodes the 21st amino acid in protein, selenocysteine (Sec) (reviewed in Low & Berry, 1996; Bock, 2000; Hatfield & Gladyshev, 2002). Available gene annotation programs interpret UGA as a stop codon, resulting in misannotation or completely missing selenoprotein genes. This problem is particularly serious for prokaryotic genomes as they currently account for >90% of completely sequenced genomes. Yet, neither which prokaryotes contain selenoproteins nor the number of selenoprotein genes in any of the prokaryotic genomes is known. Recent identification of the 22nd amino acid, pyrrolysine, which is encoded by UAG codon in several methanogenic organisms (Srinivasan *et al*, 2002), suggests that misannotation of genes due to dual functions of termination signals is not limited to Sec.

To insert Sec at UGA codons, selenoprotein genes evolved an RNA stem–loop structure, designated Sec insertion sequence (SECIS) element. SECIS elements are present immediately downstream of Sec UGA codons in bacteria (Zinoni

*et al*, 1990; Huttenhofer *et al*, 1996; Liu *et al*, 1998) and in 3' untranslated regions (UTRs) in archaea (Wilting *et al*, 1997) and eukaryotes (Berry *et al*, 1991; Walczak *et al*, 1996). The bacterial, archaeal and eukaryotic SECIS elements have no similarities to each other with regard to sequence and structure. The eukaryotic SECIS element consensus has been well characterized, which allowed identification of these structures in sequence databases (Kryukov *et al*, 1999; Lescure *et al*, 1999; Castellano *et al*, 2001; Martin-Romero *et al*, 2001), including mammalian genomes (Kryukov *et al*, 2003). However, conservation of bacterial SECIS elements has been insufficient for their computational description, which precluded identification of bacterial selenoprotein genes by searching for the stem–loop structure.

Conversely, SECIS elements in known selenoprotein genes in archaea exhibited significant conservation (Wilting *et al*, 1997), suggesting that searches for these stem–loop structures might be useful in identifying archaeal selenoprotein genes. Previous homology screens and manual analyses of selected genomic regions identified seven selenoprotein genes in *Methanococcus jannaschii* (Wilting *et al*, 1997) and, subsequently, Sec-specific translation elongation factors were identified in *Methanococcus* and *Methanopyrus* species (Rother *et al*, 2000, 2001a, 2001b).

In the present work, we applied independent bioinformatics approaches to identify entire selenoprotein sets in completed bacterial and archaeal genomes.

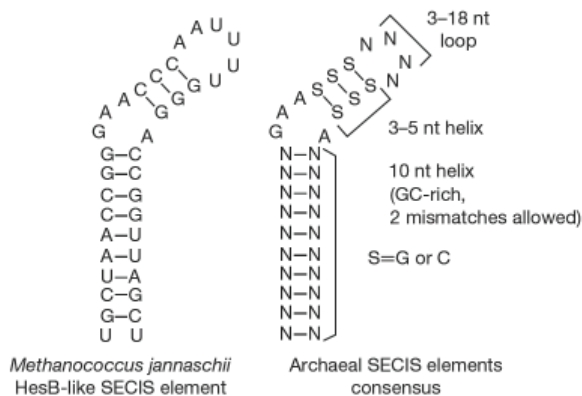
## Results And Discussion

### SECISearch-based identification of selenoproteins

Selenoprotein homologues of known archaeal selenoprotein genes (Wilting *et al*, 1997) were compiled and their SECIS elements were extracted and structurally aligned (Fig 1). This procedure revealed conserved SECIS regions and resulted in an archaeal SECIS element consensus (Fig 2, right structure). The consensus was very similar to that previously reported (Wilting *et al*, 1997), although the primary sequence conservation was limited to the unpaired GAA\_A region. We identified conserved structural features of these structures, as shown in Fig 2, and developed a computer program, archaeal SECISearch, to recognize an archaeal consensus SECIS

	Helix I	GAA	Helix II		Helix II	A	Helix I
<i>M. jannaschii</i> coenzyme F420-reducing hydrogenase, $\alpha$ -subunit:	GCCTCGAGGG	GAA	CCC	GAAA	GGG	A	CCCGAGAGGC
<i>M. jannaschii</i> coenzyme F420-reducing hydrogenase, $\delta$ -subunit:	GTTTCTCTCGG	GAA	CCC	GTCAA	GGG	A	CCGAGAGAAC
<i>M. jannaschii</i> formylmethanofuran dehydrogenase, subunit B:	TGTTGGAGGG	GAA	CCC	GTAA	GGG	A	CCCTCCAACA
<i>M. jannaschii</i> selenophosphate synthetase:	ACGATGTGCC	GAA	CCC	TTTAA	GGG	A	GGCACATCGA
<i>M. jannaschii</i> heterodisulphide reductase, subunit A:	GGCACCCTC	GAA	GGC	AAT	GCC	A	AAGTGGTCT
<i>M. jannaschii</i> methylviologen-reducing hydrogenase, $\alpha$ -subunit:	GCTCACAACC	GAA	CCC	ATTT	GGG	A	GGTTGTGAGC
<i>M. jannaschii</i> formate dehydrogenase, $\alpha$ -subunit:	GCCACCCTGC	GAA	CCC	AATATAAATAATACAA	GGG	A	GCAGGTGGCG
<i>M. jannaschii</i> HesB-like:	TGCTAACCGG	GAA	CCC	AATTTT	GGG	A	CCGGTTAGCT
<i>M. kandleri</i> coenzyme F420-reducing hydrogenase, $\alpha$ -subunit:	CCGCCGCGGG	GAA	CCCC	AAA	CGGGG	A	CCCGCGGCGC
<i>M. kandleri</i> coenzyme F420-reducing hydrogenase, $\delta$ -subunit:	CGCCCCGGGG	GAA	CCCC	GCAAGGA	GGGG	A	CCCCCGGGTC
<i>M. kandleri</i> formylmethanofuran dehydrogenase, subunit B:	CCCACGGGGC	GAA	CCCC	TCCC	CGGGG	A	GCACCCTGGG
<i>M. kandleri</i> selenophosphate synthetase:	CCCCGCGGG	GAA	CCCC	GTAGGTGT	GGGG	A	CCCCGCGGGG
<i>M. kandleri</i> formate dehydrogenase, $\alpha$ -subunit:	AGGTCGCGCG	GAA	CCCC	GAAGGA	GGGG	A	CCGTCGACCC
<i>M. kandleri</i> heterodisulphide reductase, subunit A:	CCCCGCCCCC	GAA	GGGC	GAAA	GCCC	A	GGGGGTGGGG
<i>M. kandleri</i> methylviologen-reducing hydrogenase, $\alpha$ -subunit:	TGGCCCGGG	GAA	CCC	TAAC	GGG	A	CCCCGGCCG

**Figure 1.** Alignment of SECIS elements in archaeal selenoprotein genes. SECIS elements from eight selenoprotein genes in the *M. jannaschii* genome and seven selenoprotein genes in the *M. kandleri* genome were manually aligned on the basis of their primary sequences and secondary structure features. Strictly conserved nucleotides are highlighted.



**Figure 2.** Archaeal SECIS element structures. Archaeal SECIS element consensus sequence (right structure) and the SECIS element in *M. jannaschii* HesB-like gene (left structure) are shown. Conserved structural features in the consensus structure are also indicated.

element in sequence databases. This threestep program searched for primary sequence, secondary structure and free energy criteria of the predicted stem-loop structures. Additional 'fine structural criteria' were also applied that required the presence of a conserved GAA\_A bulge in the predicted structure and removed candidate sequences with predicted Y-shaped structures. A total of 14 completely sequenced archaeal genomes (all that were available at the time of the searches) were analysed with this program (Table 1). Subsequently, sequences flanking the predicted SECIS elements were analysed for the occurrence of open reading frames (ORFs).

Interestingly, only *M. jannaschii* and *Methanopyrus kandleri* had selenoprotein genes (Table 1). Consistent with this finding, known Sec insertion machinery genes, selenophosphate synthetase (SPS) and a Sec-specific elongation factor SelB, were detected in *M. jannaschii* and *M. kandleri*, but not in other archaeal genomes. The SECIS-based analysis revealed eight selenoprotein genes in *M. jannaschii* and seven in *M. kandleri* genomes, with only one false positive and no false negatives in 14 analysed genomes. Of these 15 selenoprotein genes, 13 contained SECIS elements in the 3'-UTRs, and one SECIS element in each organism was found in the 5'-UTR, the

location not observed in eukaryotic and bacterial selenoprotein genes. Although incorrectly annotated in genome databases, 14 archaeal selenoproteins were homologues of previously identified selenoproteins (Wilting *et al.*, 1997). The 15th archaeal selenoprotein was a new selenoprotein with distant homology to HesB protein (Figs 2, 3). This ORF was entirely missing in the *M. jannaschii* genome annotation.

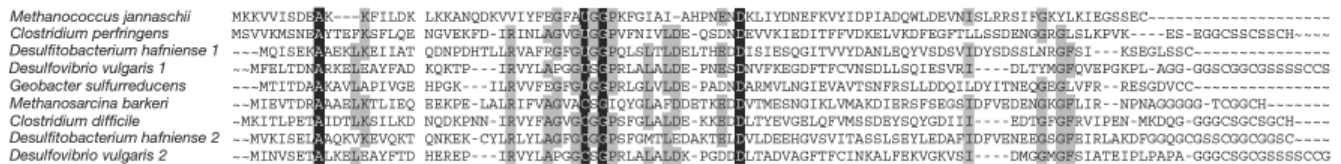
### SECIS-independent identification of selenoproteins

We also applied a SECIS-independent method for identification of selenoprotein genes that searched for Sec/Cys pairs in homologous sequences. It took advantage of the fact that all known prokaryotic selenoproteins except one (selenoprotein A) had homologues in NCBI non-redundant and/or microbial databases that inserted Cys in the place of Sec. This method was applied to both bacterial and archaeal genomes that could contain selenoprotein genes as follows (Fig 4). A database of all available prokaryotic genomes that coded for selenoprotein genes was developed, which was composed of 12 completely and 33 incompletely sequenced bacterial genomes and two completely sequenced archaeal genomes (*M. jannaschii* and *M. kandleri*) that possessed components of the Sec insertion machinery (archaeal and bacterial searches were combined in this experiment because sets of known selenoproteins in these organisms showed significant overlap). We also constructed a prokaryotic protein database, which included 227,930 protein sequences from all available annotated prokaryotic genomes. This protein database was searched against the selenoprotein genome database with tblastn to identify local alignments, in which TGA in a genomic sequence corresponded to cysteine in a protein query and the corresponding Sec/Cys flanking sequences showed significant homology. The identified TGA-containing sequences were subjected to computational filters (tests for the presence of upstream in-frame stop signals preceding translation initiation signals and superior blastx and rpsblast hits in different frames than the TGA codon; see Methods for details) and clustered, pseudogenes were removed, and the resulting candidate selenoprotein genes were screened for homologues against NCBI microbial and NR databases. To decrease the possibility that some of the hits were due to se-

**Table 1.** Selenoprotein genes and SECIS elements identified by SECISearch in completely sequenced archaeal genomes

	Archaeal SECISearch				
	Primary sequence/ secondary structure	Free energy	Fine structure	Selenoprotein genes	False positives
<i>Aeropyrum pernix</i>	0	0	0	0	0
<i>Archaeoglobus fulgidus</i>	1	0	0	0	0
<i>Halobacterium</i> sp. NRC-1	7	3	1	0	1
<i>Methanobacterium thermoautotrophicum</i>	0	0	0	0	0
<i>Methanococcus jannaschii</i>	8	8	8	8	0
<i>Methanopyrus kandleri</i>	9	7	7	7	0
<i>Pyrobaculum aerophilum</i>	2	0	0	0	0
<i>Pyrococcus abyssi</i>	2	0	0	0	0
<i>Pyrococcus furiosus</i>	1	0	0	0	0
<i>Pyrococcus horikoshii</i>	0	0	0	0	0
<i>Sulfolobus solfataricus</i>	0	0	0	0	0
<i>Sulfolobus tokodaii</i>	2	0	0	0	0
<i>Thermoplasma acidophilum</i>	1	0	0	0	0
<i>Thermoplasma volcanium</i>	0	0	0	0	0

See text for details of the searches.



**Figure 3.** Amino-acid sequence alignment of archaeal and bacterial HesB-like selenoproteins and their Cys-containing homologues. U is Sec.

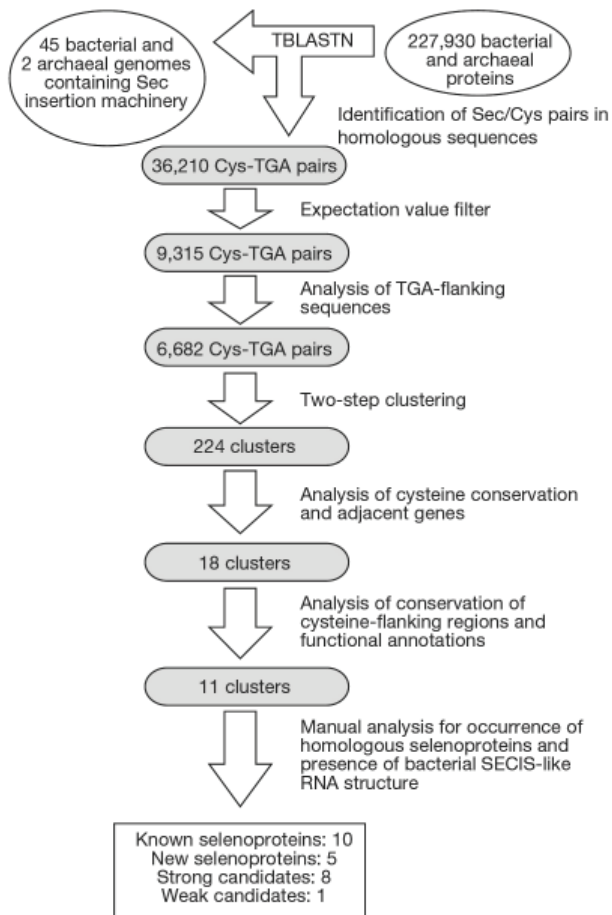
quence errors that replaced TGC or TGT codons with TGA, an ORF containing an in-frame TGA was considered a selenoprotein gene if at least two different sequences in two distant genomes were found that conserved this TGA and in addition at least two corresponding Cys-containing homologues were detected. Protein families in which all homologues conserve Sec would be missed by this approach, but such families appear to be extremely rare (currently, only one family (the selenoprotein A family) conserves Sec in 100% of sequences).

Analysis of the archaeal subset of identified sequences revealed the same set of seven and eight archaeal selenoprotein genes in the *M. kandleri* and *M. jannaschii* genomes (Table 2), with no other archaeal selenoprotein candidates or false-positive sequences. Thus, both SECIS-based and Sec/Cys homology approaches, being independent methods, were efficient in identifying selenoprotein genes in archaeal genomes. Analysis of the bacterial hits revealed ten previously known and five new selenoprotein families (Fig 4, Table 2 and supplementary Figs S1–S5 online). In addition, candidate selenoproteins, each of which was found only in one bacterial genome, were identified (Fig 4, Table 2 and supplementary Figs S6–S13 online). Potential bacterial SECIS structures were identified downstream of TGA in known, new and candidate selenoprotein genes (supplementary Fig S14 online). Only one previously known selenoprotein, sele-

noprotein A of the glycine reductase complex, was not detected, because no Cys-containing homologues were found for this protein. Thus, our ability to complete the analyses of genomic sequences with an excellent true-positive rate suggested that we identified all or almost all selenoprotein genes in completely sequenced prokaryotic genomes.

**Prokaryotic selenoproteomes**

Our data allowed a view on entire selenoproteomes in completed bacterial and archaeal genomes. Although selenoproteins were present in all three major domains of life, only ~20% of completed bacterial and ~14% of completed archaeal genomes had them. In addition, the number of selenoproteins in a genome varied from one to more than ten, with *Carboxydotherrnus hydrogenoformans*, *Eubacterium acidaminophilum*, *Geobacter metallireducens*, *Geobacter sulfurreducens* and *M. jannaschii* having the largest numbers of selenoproteins among partially and completely sequenced genomes. Analysis of the composition of selenoproteomes revealed that most selenoproteins were redox proteins, which used Sec either to coordinate a redox-active metal (molybdenum, nickel or tungsten) or in Sec: thiol redox catalysis. Interestingly, in contrast to mammals, where new selenoproteins identified through SECIS elements represent a significant fraction of selenoproteomes (Kryukov *et al*, 2003), the data suggested that in many prokaryotes, entire selenoprotein sets



**Figure 4.** Identification of prokaryotic selenoproteins by searching for Sec/Cys pairs in homologous sequences (SECIS-independent method). For details of the searches, see Methods.

were already known before our study. For example, seven out of eight selenoproteins in *M. jannaschii* (Wilting *et al.*, 1997) and two out of two selenoproteins in *Haemophilus influenzae* (Wilting *et al.*, 1998) were previously identified.

The presence of selenoproteins that were found in a small number of genomes (mostly homologues of thiol-dependent oxidoreductases) (Table 2) contrasted with the broad occurrence of their Cys homologues. This observation suggested that these selenoproteins evolved recently, probably from Cys-containing proteins. Conversely, formate dehydrogenases, which were present in most genomes with functional Sec insertion systems (supplementary Fig S1 online), appeared to be of ancient origin. In addition, bacterial formate dehydrogenase genes often clustered with Sec insertion machinery genes (data not shown), suggesting correlated expression. Thus, both a lineage-specific expansion (recent origin) and the presence of core selenoproteins (ancient origin) contribute to the composition of selenoproteomes. Our data suggest that as new genomic sequences become available, additional examples of lineage-specific selenoproteins are likely to emerge.

In conclusion, we identified genes encoding selenocysteine-containing proteins in completely sequenced genomes of archaea and bacteria. This essentially solved the UGA dual-function gene prediction problem in prokaryotes, as other in-

frame UGA codons may now be assigned a terminator function. Further systematic characterization of selenoprotein functions should reveal a full set of biological processes that are dependent on the trace element selenium.

## Methods

**Databases and resources.** Completely and incompletely sequenced genomes of archaea and bacteria were obtained from the NCBI data repository (<ftp://ftp.ncbi.nlm.nih.gov/genomes>) (Wheeler *et al.*, 2003) and TIGR, and blast programs were obtained from NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>) (Altschul *et al.*, 1990). The searches were performed on a Prairiefire 256-processor Beowulf cluster supercomputer at the Research Computing Facility of the University of Nebraska-Lincoln.

**SECIS-based identification of selenoprotein genes in archaeal genomes.** Archaeal SECISearch contains three modules. The first module is based on the PatScan program (<http://www-unix.mcs.anl.gov/compbio/PatScan/HTML/patscan.html>) (Dsouza *et al.*, 1997) and searches for RNA structures that match the archaeal SECIS element primary sequence and the secondary structure consensus. The second module, based on the RNAfold program from Vienna RNA package (<http://www.tbi.univie.ac.at/~ivo/RNA>) (Hofacker *et al.*, 1994), predicts secondary structure and calculates free energy for the entire putative SECIS element. The predicted bulge-forming GAA\_A nucleotides are constrained to be unpaired. Predicted RNA structures, the calculated free energies of which are above the  $-16$  kcal/mol threshold determined from the analysis of known archaeal SECIS elements, are excluded from further analysis. The third module of archaeal SECISearch filters out structures that are either Yshaped or do not have a pronounced GAA\_A bulge (nucleotide pairs exactly 'above' and at most one nucleotide 'below' the GAA\_A bulge are required to be present in the SECIS element stem).

**Identification of selenoprotein genes by searching for Sec/Cys pairs in homologous sequences (SECIS-independent method) in prokaryotic genomes.** *Identification of Sec/Cys pairs in homologous sequences:* The protein database was constructed by automated extraction of bacterial and archaeal protein sequences from the NCBI genome data repository. This database had 227,930 unique ORFs, which at the time of analysis represented nearly all predicted protein sequences in completely sequenced prokaryotic genomes. To identify prokaryotic genomes that encode Sec-containing proteins, the NCBI microbial genome database was searched for homologues of known protein components of Sec insertion machinery (Sec synthase, SPS and Sec-specific elongation factor). We identified 12 completely and 33 partly sequenced bacterial and two completely sequenced archaeal genomes, in which at least one Sec insertion machinery gene was detected. These 47 genomes were extracted and combined to generate a selenoprotein genome database.

The protein database was searched against a selenoprotein genome database (47 genomes, ~200 Mb) with the NCBI-tblastn program. In this search, the threshold for extending hits was set to 11 and the expectation value was cut off to 10. The resulting data set was analysed for the presence of local alignments, in which cysteine in a protein query from

**Table 2.** Prokaryotic selenoproteins

Protein name	Bacterial genomes	Archaeal genomes	Example	
			Genome	Sec location (protein length)
<i>Known selenoproteins</i>				
Formate dehydrogenase $\alpha$ -chain	37	3	<i>Methanococcus jannaschii</i>	144 (686)
Formylmethanofuran dehydrogenase, subunit B	0	2	<i>Methanococcus jannaschii</i>	121 (435)
Coenzyme F420-reducing hydrogenase, $\alpha$ -subunit	0	3	<i>Methanococcus jannaschii</i>	389 (413)
Methylviologen-reducing hydrogenase, $\alpha$ -subunit	3	2	<i>Methanococcus jannaschii</i>	441 (464)
Coenzyme F420-reducing hydrogenase, $\delta$ -subunit <sup>a</sup>	2	3	<i>Methanococcus jannaschii</i>	13,64 (134)
Heterodisulphide reductase, subunit A	1	2	<i>Methanococcus jannaschii</i>	207 (668)
Selenophosphate synthetase	17	2	<i>Methanococcus jannaschii</i>	19 (349)
Peroxiredoxin (Prx)	2	0	<i>Eubacterium acidaminophilum</i>	47 (203)
Glycine reductase complex, selenoprotein A <sup>b</sup>	9	0	<i>Clostridium difficile</i>	45 (157)
Glycine reductase complex, selenoprotein B	8	0	<i>Clostridium difficile</i>	350 (436)
Proline reductase	3	0	<i>Clostridium difficile</i>	151 (241)
<i>New selenoproteins</i>				
HesB-like protein	5	1	<i>Methanococcus jannaschii</i>	35 (95)
Thioredoxin (Trx)	4	0	<i>Treponema denticola</i>	32 (107)
Prx-like thiol:disulphide oxidoreductase	3	0	<i>Geobacter sulfurreducens</i>	66 (209)
SelW-like protein	3	0	<i>Campylobacter jejuni</i>	12 (81)
Glutathione peroxidase	1	0	<i>Treponema denticola</i>	35 (155)
<i>Candidate selenoproteins</i>				
Glutaredoxin (Grx)	1	0	<i>Geobacter sulfurreducens</i>	49 (122)
Protein similar to the N-terminal domain of Prx reductase AhpF	1	0	<i>Carboxydotherrnus hydrogenoformans</i>	147 (218)
Thiol:disulphide interchange protein	1	0	<i>Geobacter metallireducens</i>	33 (190)
DsbG-like protein	1	0	<i>Chloroflexus aurantiacus</i>	104 (253)
Fe-S oxidoreductase	1	0	<i>Desulfovibrio vulgaris</i>	244 (432)
DsrE-like protein	1	0	<i>Desulfovibrio vulgaris</i>	70 (107)
NADH oxidase	1	0	<i>Geobacter metallireducens</i>	45 (451)
Distant homologue of peroxidase/peroxynitrite reductase system component AhpD	1	0	<i>Geobacter sulfurreducens</i>	50 (98)

A total of 11 previously known prokaryotic selenoproteins are shown, along with five new selenoproteins (supported by occurrences in multiple genomes and the presence of potential SECIS elements) and eight candidate selenoproteins (supported by occurrences in single genomes and the presence of potential SECIS elements). The number of bacterial and archaeal genomes in which indicated selenoproteins were found is shown, followed by names of representative genomes (full list is in supplementary Fig S1 online) and Sec locations and lengths of representative selenoproteins.

<sup>a</sup>Contains two Sec residues.

<sup>b</sup>Protein was not detected in genomic searches.

a protein database corresponded to TGA in a translated nucleotide sequence from the selenoprotein genome database. If at least one such alignment with an expectation value below  $10^{-3}$  or at least two alignments with expectation values below 1 were identified for a particular TGA codon, the corresponding TGA-containing sequence was chosen for further analysis. We identified 9,315 such local alignments.

**Analysis of TGA-flanking sequences:** In each of the TGA-containing sequences from 9,315 alignments, a region upstream of the TGA was analysed for the presence of in-frame start and stop codons. If a stop codon (TGA, TAA or TAG) oc-

curred closer to the TGA codon than an appropriate start codon (ATG or GTG), such sequences were discarded. For each remaining sequence, a 1 kb TGA-flanking region (500 nt–TGA–500 nt) was searched against the protein database with NCBI-blastx program. If the best hit that covered the TGA codon with at least a seven-nucleotide overlap was in a different frame from the TGA, the corresponding sequence was filtered out. The 1 kb TGA-flanking regions (500 nt–TGA–500 nt) were then translated in all three possible ORFs and searched using rpsblast against an NCBI collection of known conserved domains ([ftp://ftp.ncbi.nih.gov/pub/mmd\\_b/cdd](ftp://ftp.ncbi.nih.gov/pub/mmd_b/cdd)).

If the best hit that covered the TGA codon with at least six amino-acid residue overlap was not in the same frame as the TGA codon, the sequence was removed from further analysis. In addition, if additional stop codons were found within the predicted conserved domain that covered TGA in the correct frame, such a sequence was considered a pseudogene and was also discarded. A total of 6,682 local alignments remained after application of these filters.

**Clustering:** The set of 2,071 unique proteins that corresponded to the remaining 6,682 TGA-containing sequences was extracted. These protein sequences were compared pairwise by the NCBI-bl2seq program. If two proteins produced a local alignment that had an expectation value below  $10^{-5}$  and was at least 20 amino-acid residues long, they were assigned to the same cluster. Clusters containing proteins that were initially aligned with the same TGA-containing region were joined into larger clusters. This analysis resulted in 244 clusters.

**Analyses of cysteine conservation and adjacent genes:** Cysteines that corresponded to the TGA codons in the local alignments were analysed for conservation in other homologous proteins. If a cysteine was not conserved, the hit was discarded. In addition, ORFs that contained TGA-flanking regions were analysed for domain conservation in homologues. The 1 kb TGA-flanking regions were searched with the NCBI-blastx program against the NCBI microbial and NR protein databases. If a TGA codon was located on the edge of the homology region (on either side of the sequence), the candidate hit was removed from further analyses. The remaining sequences were manually analysed for the occurrence of homologous selenoproteins and corresponding Cys-containing proteins in the NCBI microbial and NR databases and for the presence of potential SECIS elements immediately downstream of the TGA codon using mfold.

**Supplementary information** is attached. It is also available at *EMBO reports* online (<http://www.nature.com/embor/journal/v5/n5/7400126s1.pdf>).

### Acknowledgments

We thank the Research Computing Facility of the University of Nebraska-Lincoln for the use of Prairiefire 256-processor Beowulf cluster supercomputer. This work was supported by NIH grant GM61603 (to V.N.G.).

### References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410 .
- Berry MJ, Banu L, Chen YY, Mandel SJ, Kieffer JD, Harney JW, Larsen PR (1991) Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. *Nature* 353: 273–276 .
- Böck A (2000) Biosynthesis of selenoproteins—an overview. *Biofactors* 11: 77–78 .
- Castellano S, Morozova N, Morey M, Berry MJ, Serras F, Corominas M, Guigo R (2001) *In silico* identification of novel selenoproteins in the *Drosophila melanogaster* genome. *EMBO Rep* 2: 697–702 .

- Dsouza M, Larsen N, Overbeek R (1997) Searching for patterns in genomic data. *Trends Genet* 13: 497–498 .
- Hatfield DL, Gladyshev VN (2002) How selenium has altered our understanding of the genetic code. *Mol Cell Biol* 22: 3565–3576 .
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer S, Tacker M, Schuster P (1994) Fast folding and comparison of RNA secondary structures. *Monatsh Chem* 125: 167–188.
- Huttenhofer A, Westhof E, Böck A (1996) Solution structure of mRNA hairpins promoting selenocysteine incorporation in *Escherichia coli* and their basespecific interaction with special elongation factor SELB. *RNA* 2: 354–366 .
- Kryukov GV, Kryukov VM, Gladyshev VN (1999) New mammalian selenocysteine-containing proteins identified with an algorithm that searches for selenocysteine insertion sequence elements. *J Biol Chem* 274: 33888–33897 .
- Kryukov GV, Castellano S, Novoselov SV, Lobanov A, Zehtab O, Guigo R, Gladyshev VN (2003) Characterization of mammalian selenoproteomes. *Science* 300: 1439–1443 .
- Lescure A, Gautheret D, Carbon P, Krol A (1999) Novel selenoproteins identified *in silico* and *in vivo* by using a conserved RNA structural motif. *J Biol Chem* 274: 38147–38154 .
- Liu Z, Reches M, Groisman I, Engelberg-Kulka H (1998) The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in *Escherichia coli*. *Nucleic Acids Res* 26: 896–902 .
- Low SC, Berry MJ (1996) Knowing when not to stop: selenocysteine incorporation in eukaryotes. *Trends Biochem Sci* 21: 203–208 .
- Martin-Romero FJ, Kryukov GV, Lobanov AV, Carlson BA, Lee BJ, Gladyshev VN, Hatfield DL (2001) Selenium metabolism in *Drosophila*: selenoproteins, selenoprotein mRNA expression, fertility, and mortality. *J Biol Chem* 276: 29798–29804 .
- Rother M, Wilting R, Commans S, Bock A (2000) Identification and characterisation of the selenocysteinespecific translation factor SelB from the archaeon *Methanococcus jannaschii*. *J Mol Biol* 299: 351–358 .
- Rother M, Resch A, Wilting R, Bock A (2001a) Selenoprotein synthesis in archaea. *Biofactors* 14: 75–83 .
- Rother M, Resch A, Gardner WL, Whitman WB, Bock A (2001b) Heterologous expression of archaeal selenoprotein genes directed by the SECIS element located in the 3' non-translated region. *Mol Microbiol* 40: 900–908 .
- Srinivasan G, James CM, Krzycki JA (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* 296: 1459–1462 .
- Walczak R, Westhof E, Carbon P, Krol A (1996) A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA* 2: 367–379 .
- Wheeler DL et al (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31: 28–33 .
- Wilting R, Schorling S, Persson BC, Böck A (1997) Selenoprotein synthesis in archaea: identification of an mRNA element of *Methanococcus jannaschii* probably directing selenocysteine insertion. *J Mol Biol* 266: 637–641 .
- Wilting R, Vamvakidou K, Bock A (1998) Functional expression in *Escherichia coli* of the *Haemophilus influenzae* gene coding for selenocysteine-containing selenophosphate synthetase. *Arch Microbiol* 169: 71–75 .
- Zinoni F, Heider J, Bock A (1990) Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc Natl Acad Sci USA* 87: 4660–4664 .



## Supplementary Figures

### (Kryukov and Gladyshev, The Prokaryotic Selenoproteome)

**Figure 1. Prokaryotic selenoproteins.** Archaeal and bacterial genomes that encode selenoproteins are indicated. New selenoproteins were represented by at least two archaeal, bacterial or eukaryotic genomes and possessed predicted SECIS elements. Candidate selenoprotein genes were represented only in single genomes, but also had predicted SECIS elements. In addition, Sec corresponded to catalytic Cys in these proteins.

**Figure 2. Thioredoxin alignment.** GenBank accession numbers for *Treponema pallidum*, *Chlamydomonas reinhardtii*, *Oryza sativa*, *Desulfotobacterium hafniense* and *Clostridium perfringens* thioredoxins are NP\_219354.1, CAA44209.1, Q9ZP20, ZP\_00097586.1 and NP\_563454.1, respectively. In addition, the Trx-domain (152 N-terminal residues) of a *Carboxydotherrmus hydrogenoformans* homolog is shown. Bacterial genomes, in which new bacterial selenoproteins were found, are shown in bold. U is Sec. Conserved Cys/Sec residues are highlighted in red. Amino acid sequence alignments were generated with ClustalW program and shaded by BoxShade program v3.21.

**Figure 3. Alignment of Prx-like thiol:disulfide oxidoreductases.** GenBank accession numbers for *Synechocystis* sp. PCC 6803, *Nostoc* sp PCC 7120, *Thermosynechococcus elongates*, *Streptomyces coelicolor*, *Azotobacter vinelandii* and *Chloroflexus aurantiacus* peroxiredoxin-like thiol:disulfide oxidoreductases are NP\_441148.1, NP\_488682.1, NP\_682079.1, NP\_624490.1, ZP\_00092626.1 and ZP\_00020860.1, respectively.

**Figure 4. Alignment of SelW-like proteins.** GenBank accession numbers for *Sinorhizobium meliloti*, *Agrobacterium tumefaciens*, *Vibrio cholerae*, *Chlamydomonas reinhardtii*, *Homo sapiens* and *Ciona intestinalis* SelW/SelW-like proteins are NP\_384371.1, NP\_353260.1, NP\_230628.1, AAN32901.1, NP\_003000.1 and AK116508, respectively.

**Figure 5. Alignment of glutathione peroxidases.** GenBank accession numbers for *Saccharomyces cerevisiae*, *Clostridium perfringens*, *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Homo sapiens*, *Schistosoma mansoni* and *Drosophila melanogaster* glutathione peroxidases are S48499, NP\_561827.1, NP\_564813.1, AAL14348.1, NP\_002076.1, AAC14468.2 and AAF47761.1, respectively.

**Figure 6. Alignment of glutaredoxins.** GenBank accession numbers for *Escherichia coli*, *Nostoc* sp PCC 7120, *Brucella suis*, *Magnetospirillum magnetotacticum* and *Arabidopsis thaliana* glutaredoxins are NP\_290193.1, NP\_488913.1, NP\_698856.1, ZP\_00055463.1 and AAM61279.1, respectively.

**Figure 7. Alignment of a protein homologous to the N-terminal domain of peroxiredoxin reductases.** GenBank accession numbers for *Thermoplasma acidophilum*, *Ferroplasma acidarmanus*, *Thermotoga maritime*, *Aquifex aeolicus* and *Chloroflexus aurantiacus* proteins homologous to N-terminal domain of peroxiredoxin reductases are

NP\_393603.1, ZP\_00000333.1, NP\_228677.1, NP\_213313.1 and ZP\_00018901.1, respectively.

**Figure 8. Alignment of thiol:disulfide interchange proteins.** GenBank accession numbers for *Aquifex aeolicus*, *Microbulbifer degradans* and *Magnetococcus* sp. MC-1 thiol:disulfide interchange proteins are NP\_214242.1, ZP\_00065619.1 and ZP\_00042769.1, respectively.

**Figure 9. Alignment of DsbG-like proteins.** GenBank accession numbers for *Chloroflexus aurantiacus*, *Thermobifida fusca*, *Streptomyces coelicolor*, *Archaeoglobus fulgidus* and *Sinorhizobium meliloti* DsbG-like proteins are ZP\_00019418.1, ZP\_00057296.1, NP\_630109.1, NP\_070183.1 and NP\_385036.1, respectively.

**Figure 10. Alignment of Fe-S oxidoreductases.** GenBank accession numbers for *Archaeoglobus fulgidus*, *Desulfovibrio desulfuricans*, *Aquifex aeolicus*, *Magnetococcus* sp. MC 1 and *Pyrobaculum aerophilum* Fe-S oxidoreductases are NP\_069383.1, ZP\_00128545.1, NP\_213656.1, ZP\_00042574.1 and NP\_559230.1, respectively.

**Figure 11. Alignment of DsrE-like proteins.** GenBank accession numbers for *Methanococcus jannaschii*, *Methanopyrus kandleri*, *Magnetococcus* sp. MC 1 and *Methanosarcina mazei* DsrE-like proteins are ZP\_00019418.1, ZP\_00057296.1, NP\_630109.1, NP\_070183.1 and NP\_385036.1, respectively.

**Figure 12. Alignment of NADH oxidases.** GenBank accession numbers for *Deinococcus radiodurans*, *Bacillus halodurans*, *Archaeoglobus fulgidus*, *Methanococcus jannaschii*, *Oceanobacillus iheyensis* and *Giardia intestinalis* NADH oxidases are NP\_294716.1, NP\_244643.1, NP\_069231.1, NP\_247633.1, NP\_691780.1 and AAL59603.1, respectively.

**Figure 13. Alignment of a distant homolog of peroxidase/oxynitrite reductase system component AhpD.** GenBank accession numbers for *Neisseria meningitidis*, *Haemophilus influenzae*, *Pseudomonas aeruginosa*, *Methanothermobacter thermautotrophicus* and *Methanosarcina mazei* distant homologs of peroxidase/oxynitrite reductase system component AhpD are NP\_274596.1, NP\_439212.1, NP\_249256.1, NP\_276072.1 and NP\_632869.1, respectively.

**Figure 14. Predicted SECIS elements in bacterial selenoprotein genes.** Stem-loop structures were predicted with mfold (Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415). Numbers correspond to distances from selenocysteine UGA codons.

## Supplementary Figure 1

	<b>Bacterial genomes</b>	<b>Archaeal genomes</b>
Formate dehydrogenase, alpha chain	<i>Actinobacillus actinomycetemcomitans</i> <i>Aquifex aeolicus</i> <i>Actinobacillus pleuropneumoniae</i> <i>Burkholderia mallei</i> <i>Burkholderia fungorum</i> <i>Campylobacter jejuni</i> <i>Carboxydotherrnus hydrogenoformans</i> <i>Clostridium difficile</i> <i>Desulfitobacterium hafniense</i> <i>Desulfovibrio desulfuricans</i> <i>Desulfovibrio gigas</i> <i>Desulfovibrio vulgaris</i> <i>Escherichia coli</i> <i>Enterobacter aerogenes</i> <i>Eubacterium acidaminophilum</i> <i>Gemmata obscuriglobus</i> <i>Geobacter metallireducens</i> <i>Geobacter sulfurreducens</i> <i>Haemophilus influenzae</i> <i>Klebsiella pneumoniae</i> <i>Moorella thermoacetica</i> <i>Mycobacterium avium</i> <i>Mycobacterium smegmatis</i> <i>Pasteurella multocida</i> <i>Pseudomonas aeruginosa</i> <i>Pseudomonas fluorescens</i> <i>Pseudomonas putida</i> <i>Salmonella dublin</i> <i>Salmonella enteritidis</i> <i>Salmonella enterica</i> <i>Salmonella paratyphi A</i> <i>Salmonella typhimurium</i> <i>Shewanella oneidensis</i> <i>Shewanella putrefasciens</i> <i>Shigella flexneri</i> <i>Yersinia enterocolitica</i> <i>Yersinia pestis</i>	<i>Methanococcus jannaschii</i> <i>Methanococcus voltae</i> <i>Methanopyrus kandleri</i>
Formylmethanofuran dehydrogenase, subunit B	-	<i>Methanococcus jannaschii</i> <i>Methanopyrus kandleri</i>
Coenzyme F420-reducing hydrogenase, alpha subunit	-	<i>Methanococcus jannaschii</i> <i>Methanococcus voltae</i> <i>Methanopyrus kandleri</i>
Methylviologen-reducing hydrogenase, alpha subunit	<i>Desulfovibrio baculatus</i> <i>Desulfovibrio desulfuricans</i> <i>Desulfovibrio vulgaris</i>	<i>Methanococcus jannaschii</i> <i>Methanopyrus kandleri</i>
Coenzyme F420-reducing hydrogenase, delta subunit	<i>Carboxydotherrnus hydrogenoformans</i> <i>Geobacter metallireducens</i>	<i>Methanococcus jannaschii</i> <i>Methanopyrus kandleri</i> <i>Methanococcus maripaludis</i>
Heterodisulfide reductase, subunit A	<i>Carboxydotherrnus hydrogenoformans</i>	<i>Methanococcus jannaschii</i> <i>Methanopyrus kandleri</i>
Selenophosphate synthetase	<i>Actinobacillus pleuropneumoniae</i>	<i>Methanococcus jannaschii</i>

	<i>Aquifex aeolicus</i> <i>Campylobacter jejuni</i> <i>Carboxydotherrnus hydrogenoformans</i> <i>Chloroflexus aurantiacus</i> <i>Clostridium botulinum</i> <i>Clostridium difficile</i> <i>Clostridium perfringens</i> <i>Desulfovibrio desulfuricans</i> <i>Desulfovibrio vulgaris</i> <i>Eubacterium acidaminophilum</i> <i>Geobacter metallireducens</i> <i>Geobacter sulfurreducens</i> <i>Haemophilus ducreyi</i> <i>Haemophilus influenzae</i> <i>Thermoanaerobacter tengcongensis</i> <i>Treponema denticola</i>	<i>Methanopyrus kandleri</i>
Peroxiredoxin (Prx)	<i>Eubacterium acidaminophilum</i> <i>Geobacter metallireducens</i>	-
Glycine reductase complex, selenoprotein A	<i>Carboxydotherrnus hydrogenoformans</i> <i>Clostridium botulinum A</i> <i>Clostridium difficile</i> <i>Clostridium litorale</i> <i>Clostridium sticklandii</i> <i>Clostridium purinolyticum</i> <i>Eubacterium acidaminophilum</i> <i>Thermoanaerobacter tengcongensis</i> <i>Treponema denticola</i>	-
Glycine reductase complex selenoprotein B	<i>Carboxydotherrnus hydrogenoformans</i> <i>Clostridium botulinum A</i> <i>Clostridium difficile</i> <i>Clostridium litorale</i> <i>Clostridium sticklandii</i> <i>Eubacterium acidaminophilum</i> <i>Thermoanaerobacter tengcongensis</i> <i>Treponema denticola</i>	-
Proline reductase	<i>Clostridium botulinum A</i> <i>Clostridium difficile</i> <i>Clostridium sticklandii</i>	-
<b>New selenoproteins</b>		
HesB-like protein	<i>Clostridium botulinum</i> <i>Clostridium perfringens</i> <i>Desulfitobacterium hafniense</i> <i>Desulfovibrio vulgaris</i> <i>Geobacter sulfurreducens</i>	<i>Methanococcus jannaschii</i>
Thioredoxin (Trx)	<i>Carboxydotherrnus hydrogenoformans</i> <i>Geobacter metallireducens</i> <i>Geobacter sulfurreducens</i> <i>Treponema denticola</i>	-
Prx-like thiol:disulfide oxidoreductase	<i>Chloroflexus aurantiacus</i> <i>Geobacter metallireducens</i> <i>Geobacter sulfurreducens</i>	-
SelW-like protein	<i>Campylobacter jejuni</i> <i>Gemmata obscuriglobus</i> <i>Geobacter sulfurreducens</i>	-
Glutathione peroxidase	<i>Treponema denticola</i>	-

<b>Candidate selenoproteins</b>		
Glutaredoxin (Grx)	<i>Geobacter sulfurreducens</i>	-
Protein similar to the N-terminal domain of Prx reductase AhpF	<i>Carboxydotherrmus hydrogenoformans</i>	-
Thiol:disulfide interchange protein	<i>Geobacter metallireducens</i>	-
DsbG like protein	<i>Chloroflexus aurantiacus</i>	-
Fe-S oxidoreductase	<i>Desulfovibrio vulgaris</i>	-
DsrE-like protein	<i>Desulfovibrio vulgaris</i>	-
NADH oxidase	<i>Geobacter metallireducens</i>	-
Distant homolog of peroxidase/peroxynitrite reductase system component AhpD	<i>Geobacter sulfurreducens</i>	-

## Supplementary Figure 2 - Thioredoxin

```

Geobacter metallireducens      1 -----MAASGESFLVTCSACGTANRVPAEKEGVAGRCGNCRGTL
Geobacter sulfurreducens     1 -----MAESGQSFLVACPACGTSNRVPAAREGVAGRCGSCRGVL
Treponema denticola         1 -----
Treponema pallidum          1 -----
Chlamydomonas reinhardtii Ch2 1 -----
Oryza sativa TrxM           1 MALETCFRAWATLHAPQPPSSGGSRDRLLLSGAGSSQSKPRISVASPSPLRPASRFACQC
Desulfitobacterium hafniense 1 -----
Carboxydotherrnus hydrogenoformans 1 -----MEKKSTIAIIVVFVIALLAFFVYYATN
Clostridium perfringens     1 -----MEVKNKRLIILGLLIFALIGIYIGKN
  
```

```

Geobacter metallireducens     40 PPLYLHPVPLTDRSEDPFVAGYPG--PVLTEFWAPWUPHCORFAPVREIARELAG--RG
Geobacter sulfurreducens     40 PPLYFQPVPLTDRSEDPFVAGYHG--PVLVEFWAPWUPHCDFAPVREIARELAG--TA
Treponema denticola         1 --MIMAVLDTNANED-ETVKTAK--PVLIDFWAPWUPGCVQLSPEIQAAEALGD--KA
Treponema pallidum          1 ---MALDISSGNVR-KTLETNP--LVVDFWAPWCGSCRMIGPVLEEVESVGS--GV
Chlamydomonas reinhardtii Ch2 1 ----EAGAVNDDTKNVVLESSV--PVLVDFWAPWCGPCRIIAPVDEIAGEYKD--KL
Oryza sativa TrxM           61 SNVVDDEVVADEKNWDSMVLGSEA--PVLVEFWAPWCGPCRMIAPIVDELAKEYVG--KI
Desulfitobacterium hafniense 1 -MAGENVKFTTTANVNEEVVLSGDK--AVLVDFWAAWCGPCRMVAPIEELADEMAG--RV
Carboxydotherrnus hydrogenoformans 28 RELGSPANSGNVIDQYNEALKNHE--PFLLEFSQATUPTCAQMAPVQELKKKEYEGRMRV
Clostridium perfringens     28 YIEGIKQINFSQINNVVENLEAKEGIPTIIMFKTDTCPYCVEMQKELSYVSKEREGKFNII
  
```

```

Geobacter metallireducens     96 VVAQVNTQGNPOVASRFGIRGIPALVLLQR-----GKILGSLNGAQSKEVLSW
Geobacter sulfurreducens     96 AVVQVNTQENPOLAARFGIRGIPALVLLRR-----GQVLATWSGALPREAVLSR
Treponema denticola         54 VIAQSNVDNARELAVKFKFMSIPTLIVLKD-----GKEVDRHTGYMDKSLVNF
Treponema pallidum          52 VIGKLNVDLDDQLAVEFNVASIPTLIVLKD-----GKEVDRSTGFVDRSKILTL
Chlamydomonas reinhardtii Ch2 52 KCVKLNTEDESPNVASEYGIRSIPTIMVEKGD-----GKKCETIIGAVPKATIVQT
Oryza sativa TrxM           117 KCKKVNTEDESPNIATNYGIRSIPTVLMFKN-----GEKKEVSVIGAVPKTTLATI
Desulfitobacterium hafniense 56 IIGKLNVDDEEPAIAGQYQVMSIPTLAVEKKN-----QQVVDKSVGFRKADLVKM
Carboxydotherrnus hydrogenoformans 86 IVAEVSRQDSQQLAAKFGIQYVPTFIVWDQNGNIVPWTDAQGNQMPMFVGGITKEELKAO
Clostridium perfringens     88 YYARLEEEKNIDLAVKYDANVVPTTFVLDK-----EGNKFYVHQGLMRKNNIETI
  
```

```

Geobacter metallireducens     145 VRSTLR-
Geobacter sulfurreducens     145 VRDALR-
Treponema denticola         103 VSKHI--
Treponema pallidum          101 IQKNA--
Chlamydomonas reinhardtii Ch2 101 VEKYLN-
Oryza sativa TrxM           166 IDKYVSS
Desulfitobacterium hafniense 105 IEKHA--
Carboxydotherrnus hydrogenoformans 146 MDKVALK
Clostridium perfringens     138 IINSLGVK
  
```

### Supplementary Figure 3 - Prx-like thiol:disulfide oxidoreductase

*Geobacter sulfurreducens* 1 -----MSPEHAATLQRTATE--LVLSGIVG-----  
*Synechocystis* sp. PCC 6803 1 -----MNLQIELYKQQESLKRSSPERAAIFSDFIQGLSEEFNR  
*Nostoc* sp. PCC 7120 1 MNADRHRYKISVNSGVHLWFYHCIIYQLVDNFERFMLTSTDFSGILLNERFFRNFLPIPAS  
*Thermosynechococcus elongatus* 1 -----MAVQLP-----FTLSTNFSGLFNERFWQNAWPLPPQ  
*Streptomyces coelicolor* 1 -----MTTTPADIADQAAVLAEGMADQTPSEALEAFGAEQAELEDAAGVPS  
*Azotobacter vinelandii* 1 -----MSESLNRLLAELHAERERTWDPAALRVNVEQRRRLVVEEARAE  
*Chloroflexus aurantiacus* (Sec) 1 -----MSRPFIPMTAKWRQEFKPRGPNVPAVVGSEAPD  
*Chloroflexus aurantiacus* (Cys) 1 -----VMNAQSHNWLSTILIIICVAVIISRPPDMNPTV

*Geobacter sulfurreducens* 24 HAATIGDRAQDFITLPA-NVGRQIRLSEVTAQSTAVVTFYR---GAWUPYCSLQLRAYC-  
*Synechocystis* sp. PCC 6803 41 RLLRIGDFAPDFITLKN-KGETIILSEQLKTPILLKFFR---GYWCPYCGLELRAYC-  
*Nostoc* sp. PCC 7120 61 NELRLDVGTPDFIQLEPDTNNGTLVKSNIYRCKQPIILAEITRIFTEKQYCPFCFPHIKALN-  
*Thermosynechococcus elongatus* 32 NELRRCALVPDVALPGVGLSDRVLSNEWKKQPLLVFTRIFTAHQYCPICYPYLTFLN-  
*Streptomyces coelicolor* 44 GIATAGSLMPDASLLDV-HGNATLQOVREGRPAVVFYR---GAWCPYCNIALRTYER  
*Azotobacter vinelandii* 43 RFVYACERVEFFELLRV-EGGRLTLDLVLVANGPAVVFYR---FAGCPACNIATPYQR  
*Chloroflexus aurantiacus* (Sec) 35 FTLPYAQFLSGPPEDRVEYGRITITLSALRCR-PVVLNLTRIVSDRFFUPHCAQQLDALR-  
*Chloroflexus aurantiacus* (Cys) 35 PVPQGGFVAFAIVLPQL-DGQTLSSLSELCQVVIVNFWAS-----WCGPCRAEMPMLL-

*Geobacter sulfurreducens* 77 AVLPRLRELGGELLALSPQTPDK---SQATLLKNFLQYEVLSVGNLVARSEGGLVYPLGE  
*Synechocystis* sp. PCC 6803 95 KVVNKIRALGTTILALSPQTLVA---SQKTIDRHDLYDLSDSGFQTAQDYGLVFTVFD  
*Nostoc* sp. PCC 7120 120 ENYEQFTNRGTEVLLVSTDEKQ---SQIVVKDLGLKMPLLSDPSCRAFRTYQVGGQALG-  
*Thermosynechococcus elongatus* 91 ENHETFQGGKAVLVVSTDAQQ---SEKVKADMALKMPLLYDFSCQVFRKYRTGQALG-  
*Streptomyces coelicolor* 99 ELAPRLAERGVSMIAVSPQRPDG---SLTMAQTNDLSYDVLSDPKNHIGRALGIVTRPTD  
*Azotobacter vinelandii* 98 NLQPIILHAWGMPVAVSPQIPER---LGEIKSRHGLLLEVASDRDNALGRREGILYEFDE  
*Chloroflexus aurantiacus* (Sec) 93 EHYDLFVQRNAHLVVSSTDLEM---TSYVAEVLRAVYPIILSDPEWGVFYRYGMSAMG-  
*Chloroflexus aurantiacus* (Cys) 87 QLYQAERRRGLTVLAVNSTVQDNPADVSNMQRDFGLSFPVILVYDGSVGNRYGVR-----

*Geobacter sulfurreducens* 134 EMRIYLGFGVNLADYNGDESWEPLPFGTFVIDGTMTRYSFVDADYTRRLEP--ATLLD  
*Synechocystis* sp. PCC 6803 152 AVKQIYLQSGCVIPEHNGTEEWLFPVPATFVIDRRGHIALAYANVDFRVRMEP--EDATA  
*Nostoc* sp. PCC 7120 176 -----APLPAQFVLDKDGRLRYKHLFSFFEDHNASV--EKLLG  
*Thermosynechococcus elongatus* 147 -----APLPAQFLLDQEGKLHYKHLFSFLEPNAPL--ERLFQ  
*Streptomyces coelicolor* 156 RVQHAQASLGLDLTEVNDGTPDILMPTVAIVDAEGVLRWIDVHPNYVTRSEP--ARTLE  
*Azotobacter vinelandii* 155 PSRRASLAKGPGIGALTGTGTWELQPAATVIGRDRRVHFAEVSFDWLVRTFA--LPLLE  
*Chloroflexus aurantiacus* (Sec) 149 -----VPLPGVFVIDADGTRRWSWAAPLSVVFVTPRPAELAA  
*Chloroflexus aurantiacus* (Cys) 142 -----MLPTTFIDRKGVVRVLFGGPLSEANLR---NVIE

*Geobacter sulfurreducens* 192 VLERIREERGRDDNQAS  
*Synechocystis* sp. PCC 6803 210 ILLSLFVGN-----  
*Nostoc* sp. PCC 7120 211 KFD-----  
*Thermosynechococcus elongatus* 182 EIDALAQGATVTAA--  
*Streptomyces coelicolor* 214 ALARAVR-----  
*Azotobacter vinelandii* 213 AVRALLAEPALRATP-  
*Chloroflexus aurantiacus* (Sec) 186 VLDALASEG-----  
*Chloroflexus aurantiacus* (Cys) 175 PLLAETE-----

## Supplementary Figure 4 - SelW-like

<i>Sinorhizobium meliloti</i>	1	MSE-KPRV	V	T	I	L	L	Y	C	T	C	N	W	L	L	R	A	G	W	M	A	Q	E	L	L	S	T	F	A	D	T	L	G	E	V	A	L	--	I	F	G	T	G	N	F	E	I	R	V	D	G	A	L
<i>Agrobacterium tumefaciens</i>	1	MTETKPR	A	T	R	Y	C	T	C	N	W	L	L	R	A	G	W	M	A	Q	E	L	L	S	T	F	A	S	D	I	G	E	V	S	L	--	I	P	S	T	G	G	L	F	E	I	T	V	D	G	T	I	
<i>Vibrio cholerae</i>	1	MNK--AQ	E	T	Y	Y	C	R	C	N	W	M	L	R	S	A	W	L	S	Q	E	L	L	H	T	F	S	E	E	I	E	V	A	L	--	H	P	D	T	G	G	R	F	E	I	F	C	N	G	V	Q		
<b><i>Campylobacter jejuni</i></b>	1	MMK---	V	K	I	A	Y	C	N	L	Y	R	P	Q	A	R	V	A	E	L	Q	S	D	F	K	--	V	E	V	E	F	--	E	I	G	G	R	G	D	F	I	V	E	V	D	G	K	V					
<i>Chlamydomonas reinhardtii</i>	1	-MAP-VQ	H	V	L	Y	C	G	G	Y	G	S	R	Y	R	S	L	E	N	A	T	R	M	K	F	P	N	A	D	I	K	F	S	F	E	A	T	P	Q	A	T	C	F	F	E	V	E	V	N	G	E	L	
<i>Homo sapiens</i>	1	-MA--LAV	R	V	V	Y	C	G	A	U	G	Y	K	S	K	Y	L	Q	L	K	K	L	E	D	E	F	-	G	R	L	D	I	C	G	E	G	T	P	Q	A	T	C	F	F	E	V	M	V	A	G	K	L	
<i>Ciona intestinalis</i>	1	-MPNKV	K	H	V	V	Y	C	G	G	Y	R	P	R	Y	E	R	L	K	D	L	S	K	D	Y	D	Q	N	E	V	E	F	S	S	E	G	T	P	E	V	T	C	Y	L	E	V	L	V	N	G	T	L	
<b><i>Gemmata obscuriglobus</i></b>	1	-----	M	N	V	E	L	K	Y	C	S	L	G	Y	E	P	K	A	V	S	L	A	A	T	L	T	S	L	K	Q	K	V	K	G	L	T	--	L	V	P	A	G	G	C	F	E	V	T	V	N	G	E	L
<b><i>Geobacter sulfurreducens</i></b>	1	-----	M	N	V	E	L	F	C	P	T	S	Q	Y	P	I	A	A	G	L	A	R	L	T	E	Q	T	E	N	V	S	V	E	L	D	---	K	Q	A	P	R	S	E	F	A	V	I	D	G	E	I		

<i>Sinorhizobium meliloti</i>	58	I	W	E	R	K	R	D	G	G	F	P	G	P	K	E	-	L	Q	R	I	R	D	V	I	E	P	E	R	D	L	G	H	T	D	R	S	-----						
<i>Agrobacterium tumefaciens</i>	59	I	W	E	R	K	R	D	G	G	F	P	G	P	K	E	-	L	Q	R	I	R	D	L	I	D	P	E	R	D	L	G	H	V	D	R	T	K	H	E	G	L	D	T
<i>Vibrio cholerae</i>	57	I	W	E	R	K	Q	E	G	G	F	E	A	K	V	-	L	Q	R	V	R	D	L	I	D	P	E	R	D	L	G	H	V	D	R	P	S	T	Q	S	--			
<b><i>Campylobacter jejuni</i></b>	53	I	F	S	K	T	Q	L	I	N	C	E	S	E	R	F	P	Y	Q	N	E	I	N	Q	L	I	K	N	R	V	-----													
<i>Chlamydomonas reinhardtii</i>	59	V	H	S	K	K	N	G	G	H	V	D	N	Q	E	-	K	V	E	R	I	F	A	K	I	G	E	A	L	A	K	-----												
<i>Homo sapiens</i>	57	I	H	S	K	K	K	G	D	C	Y	V	D	T	E	S	-	K	F	L	K	L	V	A	A	I	K	A	A	L	A	Q	G	-----										
<i>Ciona intestinalis</i>	60	V	H	S	K	K	N	G	D	C	Y	I	D	S	E	A	-	K	L	K	I	C	N	A	I	D	K	C	L	Q	-----													
<b><i>Gemmata obscuriglobus</i></b>	54	I	N	S	K	L	Q	T	G	T	F	P	D	E	Q	S	-	V	L	E	S	V	R	E	R	L	K	R	-----															
<b><i>Geobacter sulfurreducens</i></b>	53	I	F	S	R	L	E	R	C	R	M	P	E	P	L	D	-	I	I	P	A	T	R	A	R	R	H	G	T	S	G	-----												



## Supplementary Figure 5 - Glutathione peroxidase

*Saccharomyces cerevisiae* HYR1 1 -----MSEFYKLAIPVDKKGQPFPPDQ  
*Clostridium perfringens* 1 -----MEIYDLSVKDINGENVSLER  
***Treponema denticola*** 1 -----MGIYNYIVKDSLGNDFSFND  
*Arabidopsis thaliana* 1 -----MATKEPESVYELSTEDAKCENNLALSQ  
*Chlamydomonas reinhardtii* 1 MLLTRKNVAVRPARAARRDVRAMSLGNLFGGSKPTSSTSNFHQLSALDIDKKNVDFKS  
*Homo sapiens* Gpx4 1 --MSLGRLCRLLKPALLCGALAAPGLAGTMCASRDDWRCARSMHEEFSAKDIDGHMVNLDK  
*Schistosoma mansoni* 1 -----MSSSHKSWN---SIYEFIVKIDINGVDVSLER  
*Drosophila melanogaster* 1 -----MSAN-GDYKNAASIIYEFIVKIDINGVDVSLER

*Saccharomyces cerevisiae* HYR1 22 LKGVVLLIVNVASKCGFT-POYKELEALYKRYKDEGFTIIGFPCNQFGHQEPGSDDEELAQ  
*Clostridium perfringens* 21 YRGKVVLLIVNVASKCGFT-KOFDGLLEELYEKYKDEGFEVLGFPCNQFKEQDPGNSNSEIMN  
***Treponema denticola*** 21 YKDYLIVNVNTACEUGLT-PHFQGLEALYKEYRDKKFLVAAFPCNQFGQDPGTNEEIEIN  
*Arabidopsis thaliana* 27 YKDKVLLIVNVASKCGMTNSNYTELNELYNRYKDKGLEILAFPCNQFGDEEPTINDQITD  
*Chlamydomonas reinhardtii* 61 LNNRVLLVNVVASKUGITAAANYKEFATLLGKYPATDLTIWAFPCNQFGGQEPGTNAEIKFA  
*Homo sapiens* Gpx4 59 YRGFVCLIVNVASKUGITEVNYTQLVLDLHARYAECCGLRILAFPCNQFGKQEPGSDNEEIKR  
*Schistosoma mansoni* 29 YRGHVCLIVNVASKUGATDKNYRQLQEMHTRLVGGKGLRILAFPCNQFGQEPWAAEIEIKK  
*Drosophila melanogaster* 31 YKGVVLLVNVVASKCGLTTKNNYEKLTDLKEKYGGERGLVILNFPNQFGSQMPEADGEAMV

*Saccharomyces cerevisiae* HYR1 81 FCQLN--YGVTFPIMKKIDVNGGNEDEPVYKFLKSQKSCM-LGLRGIKWNFEKFLVDKKGK  
*Clostridium perfringens* 80 FCKLN--FGVTFPMFEKIDVNGENESLLYSYLKEQKSCM-FGSK-IKWNFTKFLVDREGN  
***Treponema denticola*** 80 FAQSK--YGVSFPTMAKIEVNGENTETIFSFLLKASNG-----EDIKWNFAKFLVDKKTGE  
*Arabidopsis thaliana* 87 FVCTR-FK-SEFPTFNKIEVNGENASPLYKFLKKGKWCIFG-DDIQWNFAKFLVDKNGQ  
*Chlamydomonas reinhardtii* 121 FASARGFSGAGALLMDKVDVNGANASPVYNFLKVAAGDT-S---DIQWNFGKFLVLRPDGT  
*Homo sapiens* Gpx4 119 FAAG--YNVKF-DMFSKICVNGDDAHPLEKWKIKIQPKGKGLGNAIKWNFTKFLVDKNGC  
*Schistosoma mansoni* 89 FVTEK-YGVQF-DMFSKIKVNGSDADDLYKFLKSRQHC--TLTNNIKWNFSKFLVDROGQ  
*Drosophila melanogaster* 91 CHLRD-SKADIGEVFAKVDVNGDNAAFLYKYLKAKQTC--TLGSGIKWNFTKFLVNVKEGV

*Saccharomyces cerevisiae* HYR1 138 VYERYSSLTKPSSISSETTEELLKEVE  
*Clostridium perfringens* 136 VIKRESFQTTPKSIEKDIEELLA---  
***Treponema denticola*** 133 RVTAYAPTVAPEDLKKDIEKLLN---  
*Arabidopsis thaliana* 143 AVQRYPTTSPLEHDIKNNLLNIS-  
*Chlamydomonas reinhardtii* 177 VFGRYAPTTPLEKYLVELINSR-  
*Homo sapiens* Gpx4 176 VVKRYGPMEEPVLKEDLPHYF---  
*Schistosoma mansoni* 145 PVKRYSPPTAPYDIEGDIMELLEKK-  
*Drosophila melanogaster* 148 PINRYAPTTPMDIAKDIEKLL----

## Supplementary Figure 6 - Glutaredoxin

```

Escherichia coli Grx3      1 -----MANVEIYTKETCPYCHRAKA
Nostoc sp PCC 7120       1 -----MSNLFNQLFGRSPAKIKANVEIYTWQTCPYCIHRAKL
Brucella suis            1 -----MVDVLIYTFPGCPYCARAKA
Magnetospirillum magnetotacticum 1 -----MAETEIYTTVCPCYVRAKK
Arabidopsis thaliana     1 MTMFRSISMVLLVALVTFISMVSSAASSPEADVFVKKTISSHKIVIFSKSYCPYCNRAKS
Geobacter sulfurreducens 1 -MMVRSLTAMLVLAATVALTPALLHSAPDKPGRTAES--RNPSVVIIVGEGUPYCDVEVER

```

```

Escherichia coli Grx3      21 LLSS-KGVSEQ-ELPTDGNAAKREEMIKRS-GRTVVPQIFIDA--QHIGGDDLYALDAR
Nostoc sp PCC 7120       37 LLWW-KGVQET-EYKIDGDEAARANMAERANGRRVVPQIFINN--QHIGGDDLYELDTK
Brucella suis            21 LLAR-KGAEEN-EIDASATPELRAEMQERS-GRNTEFPQIFIGS--VHVGCCDDLYALEDE
Magnetospirillum magnetotacticum 21 LFAK-KGVDTT-EINVSTDDCLRQYMTNRAGCRRSVPOIFIDG--VHVGCCDDLYALDKD
Arabidopsis thaliana     61 VFRELDQVPMVVVELDEREDGWSIQTALGEIVGRRTVPOVFING--KHLGGSDDTVDAYES
Geobacter sulfurreducens 58 FFTE-KGIPYT-CRDIRRDRAAFREWRERY-GGETVPMVVLDCGKKVLDGCD-HPALER-

```

```

Escherichia coli Grx3      76 GGLDPLLK-----
Nostoc sp PCC 7120       93 GQLDPLLVQPA-----
Brucella suis            76 GKLLSLLKTGKLI----
Magnetospirillum magnetotacticum 77 GKLDPMLAGAR-----
Arabidopsis thaliana     119 GELAKLLGVSGNKEAEL
Geobacter sulfurreducens 113 -ALADIRSSRP-----

```

## Supplementary Figure 7 - Protein homologous to N-terminal domain of Prx reductases

```

Thermoplasma acidophilum          1  MANLIRKEDREYLKGFEEKYLKNDVDLVVFTSN--DEN-CRYCKEIVQLATEVSEIN---
Ferroplasma acidarmanus         1  MS-LIRDEDAKFLREDDFDKLKNNTVDLVVFSSE--SSD-CRYCKDTISLVDEVGALS---
Thermotoga maritima             1  MG-ILSDKDIAYLKLDFGKELKRKVKIVFFKTE--DKTRCQYCEIIECVLEELVSVDF---
Aquifex aeolicus               1  --MLLNLDVRMQLKELAQKEEKEPVSIKLIFS----QAIGCESCQTAEELLKETEVEVIGEA
Carboxydotherrnus hydrogenoformans 1  -MAFLQEKDIOFLKEKFAKEMVNDVTHFFFTKSPVLAQDCPYCDHTEKOLLEELAATS---
Chloroflexus aurantiacus       1  -MALMQEKDRKAVQGAFTG-LNRPVDIVLFTS----AESGMYSEVTOELLQEVVALH---

Thermoplasma acidophilum          55  ---PKIHLKVVNFEDDKEMVVKYGVKYPATIVSK-AGVEDGRIVVYGLPSGYEFGSLIE
Ferroplasma acidarmanus         54  ---DKINLVKVVYKDKEMVEKYGVKYPATIVAR-HDDKDGRITIIYGTIPSGYEFGLIE
Thermotoga maritima             55  ---PKLELEIHDLESDKEAVEKYQVEMVPATII LP-EDGKDYGIRFYGVPSGHEFGTIIQ
Aquifex aeolicus               55  VGQDKIKLDIYSPFTHKEETEKGVDRVPTIVIE---GDKDYGIRYIIGLPAGLEETTLIN
Carboxydotherrnus hydrogenoformans 57  ---EKIKLMVHTYPTKEAVEKYGIDKIPAIIVFE---GTEDVVGIRFYCTIPSGYEFSTVIE
Chloroflexus aurantiacus       52  ---PMLSLHVVYDLQDSAYAAELGVDKAPGIVFLVGEERQNHGLRFAGLPSGYEFASLIE

Thermoplasma acidophilum          111  DLKNVS-MGEADVSSKAABELLSKIDKPTITIKVYVTPTCPYCPRAVGTAHKFALLNPNIKG
Ferroplasma acidarmanus         110  DLENVS-VGEADVSKKAVDLLSKIDKPTITIKVYVTPTCQYCPKAVITAHKFALMKNKIKS
Thermotoga maritima             111  DIITVS-EGKPOLSEESIQLQSLLEPIRISVFTVPTCPYCPRAVLMAHNMAASDKIIC
Aquifex aeolicus               112  GIFHVS-QRKPOLSEKILELLQVVDIPIEIIWVFTVTS CGYCPASAAMAWDFALANDYITS
Carboxydotherrnus hydrogenoformans 111  TIIDIS-KGKPELPTNVLAEELAKVTSPTIKVFTVPTUPYCPRAAVANRFAMANANIRA
Chloroflexus aurantiacus       109  AIRLAGGAVQPDLPATMALLETIQS PMHLQVFTVPTCPYCPRAVVMAYRLALASPYISA

Thermoplasma acidophilum          170  EMIEALEFENEAEVGVSSVPHIVINNDVT--FTGAYPDDQFAEYVMEAYDHQ-----
Ferroplasma acidarmanus         169  EMIESLEFDKEASDAGVSAVPHIVINDVT--FGAYPDDQFAEYVMEAYNHQE-----
Thermotoga maritima             170  EMIEANEVWELSEKFGVSSVPHIVVNRDPSKFFVGAYPEKEFINEVRLAKG-----
Aquifex aeolicus               171  KVIDASENQDLAEQFQVVGVPKIVINKGVA-EFVGAQPENAFGLYIMAVYEKLRKREKQA
Carboxydotherrnus hydrogenoformans 170  EMIEVSEEPFELGDKYVNVFVGPKSVINETVE--IEGAAPETMEVQKVI EAVQ-----
Chloroflexus aurantiacus       169  EGIEVTEEPFELGDRYAVMGVPKIVIDDLVH--IEGAVPEGMMVNKLREAI AAAA-----

```

## Supplementary Figure 8 - Thiol:disulfide interchange protein

```

Geobacter metallireducens      1  -----
Aquifex aeolicus              1  -----MLAIRFFLIIFLMFLSV
Microbulbifer degradans       1  MHHSAFALQLLHRLISQQSENPMSPQAANTFFVKAAKSVGIFVLLVGAYTINVEIQSY
Magnetococcus sp. MC-1       1  -----MGGVGAWTLKKLISLILFALSMIMV

Geobacter metallireducens      1  MAQIEWITS-----LAEGLSMAGRENRLVLLDFFNPGUIGCKQMDAVAYPDAVMTFTN
Aquifex aeolicus             17  TFSQEWFAD-----FDKGVNTAKKEKKLVLYFYSDHCYCHQVEEFVFGEDVEKFTN
Microbulbifer degradans       61  LGRNAAVATGLPHTHTFEQALS LAKQSKPVLANFSAAWCPACRRLDKDV LAKFEVVKQRTE
Magnetococcus sp. MC-1       26  PSSQASEF IGESLDMDLPGEMANAGKEGKGLVVMFHYSGCPECDKMRQRVFPDPAVVAYYS

Geobacter metallireducens      55  DNLVPRIPADDP-----VLCPEFKVKWTPTIIVLDAQGDEHYRTLGFYPP
Aquifex aeolicus             71  KNFIVISVNNIS-----NLSEKEDVEGTPTFVIYDPLRG-----
Microbulbifer degradans       121 QHYIFTRIDYDTEEGQ-----TFMARYQAKGTPTL LILDAQGE-----
Magnetococcus sp. MC-1       86  QHFVLLLETNIRGLLEMVAPNGEGMSEKQWAHKMRIRATPEVFLFFDAEGK-----

Geobacter metallireducens     101 ADLIPSLLLGMGKARFNQPD RQAACQCFRRIISDYPKNSLAPEAIYLNGVARYIETHDVA
Aquifex aeolicus             105 -----KVLAKFFGSLDAQ
Microbulbifer degradans       159 -----QLKRINLTFAPA
Magnetococcus sp. MC-1       135 -----ERLKLTYQAPN

Geobacter metallireducens     161 NLIGHDRIAAEYDSPWLTTRADPYKLLKR
Aquifex aeolicus             118 TFLSMITRVCNKSS----VRR-----
Microbulbifer degradans       171 CFLTQL-----
Magnetococcus sp. MC-1       147 VFIQAGRYVQEKGWEEKGSFVRWLRQSGS--

```

## Supplementary Figure 9 - DsbG-like

*Chloroflexus aurantiacus* (Sec) 1 -----MARPLILVVVIMLVACIASAPVADMPSPPTAVVPTTTPSTSPST-----  
*Chloroflexus aurantiacus* (Cys) 1 -----MRVRLLVLTMLLSLTACIAATPVP-SPTAIPTPTVTPEPLSQS-----  
*Thermobifida fusca* 1 MVRPSCCCGWSRRCVFEWRVVSQVVGEVGRGSGRGWGRGYVVVAVVLLLVAVAAAGVW----  
*Streptomyces coelicolor* 1 -----MTGSTPSSSASPSSSRPPSASRRSRRTVAALAVLAAAAVTAFA-----  
*Archaeoglobus fulgidus* 1 MLMRDLAVGVVIGLIIGATAVYAFTALQPSCEEVCPANPEPESLDQKTTLESVSKKLNLIH  
*Sinorhizobium meliloti* 1 ---MSASQTSLPKRLLSGVAVAAALAMALAACSDEKKEAASTTPAETTASTDATTASTSQ

*Chloroflexus aurantiacus* (Sec) 44 TTAFTVAAPTAALPTVAVPATAVPVVTYRG-----  
*Chloroflexus aurantiacus* (Cys) 44 TPTERTIIRLSESPLPVFTPTPIPTPVISAQCIT-----AS  
*Thermobifida fusca* 57 VGWGGRAESDSRPGDCAVAGSASFPVVGAP-----  
*Streptomyces coelicolor* 46 LTLDDADNREEKAGEPAAVTASAAEAPADEG-----L  
*Archaeoglobus fulgidus* 61 QNNEDVSVRISDYSPTYEVYRVKVEFYNDNCTLESYDMFLTANGSLLFTNYVDLTKLSEE  
*Sinorhizobium meliloti* 58 APAAASVAKPATEVAQASTPAAKVVEIPKSEGSVD-----M

*Chloroflexus aurantiacus* (Sec) 75 -ALVGRDANGAYTLGCPAAPITLTDYSDFLUTVCRRHVLTVEPALIEQVVTGRVLYVFR  
*Chloroflexus aurantiacus* (Cys) 80 MAQVIGISAEFYAILGDPNAPVTIIEFTDFGCTFCRRHHVLTFPALREEFISGGQVFYVVR  
*Thermobifida fusca* 88 ---VPPQVDPELVLGRSDAPVTVVVFSDYQCPCYCARFALEQQPVLVERYVETGOVRLVWR  
*Streptomyces coelicolor* 78 LALARRDASDELAIGRADAPVLLIEYSDFCCPFCGRFARETKEPELLRSYVVKGTLRIEWR  
*Archaeoglobus fulgidus* 121 EVRINVSIDDFPFKGAEDAKVIVVEFSNYACGHCADFAETETPKILEKYG--DKVKIVFR  
*Sinorhizobium meliloti* 93 AKLLEPGALPEMALGTEANAPVTIIVEYMSMTCCHCANFHNDFDALKAKYIDSGKVRFTVR

*Chloroflexus aurantiacus* (Sec) 134 PVLNHCAASLITTAAFECAGEQDAEWPMHELLFERQGEVAATRSDLPALMRSYAADLGL  
*Chloroflexus aurantiacus* (Cys) 140 QLPVTSPHGDQAALAAALCAGEQGYWEMHDQLFA-AGDAWYSDATTARRRITIALATDLGL  
*Thermobifida fusca* 145 DYPYLGEESVRAAFAARAAGRQGRYWDYHEALYE-SSEVWRAAC-ASRESLVEVAATIGL  
*Streptomyces coelicolor* 138 NFPTFGEESQAALACWAAGRONKFEWFHDVAYG--KPRERNTCAFDAENLVAMAREAGI  
*Archaeoglobus fulgidus* 179 DFPGFGEISYFAAEFAANCAGEQGYWFEHDLLE-----NQREWISNNSKYDYAEQLGL  
*Sinorhizobium meliloti* 153 EFP-EDPRAAAAFMLARCAPE-GQYFPMVSMLEK--QQEQWAAQNGRDALLQLSKLAGF

*Chloroflexus aurantiacus* (Sec) 194 A-LEPFEDACMNDGAAQRLAETLDAE-QRQRCIRVQVVEIG-----DIRLVGLQTLLE  
*Chloroflexus aurantiacus* (Cys) 199 D-SAVLQRCMEHPATQATLARHVSE-AHALRVFGTPTFFIN-----NQLFAGAQPFA  
*Thermobifida fusca* 203 D-TDQFAVDLADPVLREAVEEDFAF-ALGLGVPGTPEFLID-----GEAFFGAQPVE  
*Streptomyces coelicolor* 196 ADIERFQADMASDEARGAVRADQEE-GYTLGVTSTPEFLVN-----GRPILGAQFTD  
*Archaeoglobus fulgidus* 234 N-VDEEKACLESVKYREVDKDYKD-GISYGVGTGPTFFIGTPNGTFVNGKVVAGALNFE  
*Sinorhizobium meliloti* 209 T-QESFEACINQKLLLDVNAVVMQRGAKEFGVKSTPTFFVN-----GEHYSGDMSVD

*Chloroflexus aurantiacus* (Sec) 244 RFASLIERQP-----  
*Chloroflexus aurantiacus* (Cys) 249 RWRDVLLEGGR-----  
*Thermobifida fusca* 253 RFAERLDEALGKRG-----  
*Streptomyces coelicolor* 247 TFEAVETAATAAKTANTEGAGR  
*Archaeoglobus fulgidus* 292 QFAALIEQELQQAS-----  
*Sinorhizobium meliloti* 260 VMSALIDSKL-----

## Supplementary Figure 10 - Fe-S oxidoreductase

```

Archaeoglobus fulgidus      1  --MVEEKVVEKGERAVEVTKWRRVFDYRAMSDEILRPD----EPRKEKFLEAMRKYL
Desulfovibrio desulfuricans 1  MGIADRKTEDEGLKRGTAALTP----DRIQSVIQSVIQGE----TQAR-----
Aquifex aeolicus          1  --MVKLFHKEKVSQEDIKKFYN-----FCKSSINSE----VMAAY-----
Magnetococcus sp MC 1     1  --MADDIHVPEIGDDIVTPAPVVGTSHLKPLPAQAKHMEPLDFPCEFRVENWQQAGIEKF
Pyrobaculum aerophilum    1  --MHKLEFRLGDTSNLKPLKAPEGLVVKLQKKAGIESPALKYSYVEKR-----L
Desulfovibrio vulgaris    1  --MSDDTLKPHQEPGRFTKDRVMEVLPDG-----KYSYVEKR-----

Archaeoglobus fulgidus      55  KQNWPFLLPYKLTLEACTKCGTCAEACTMYLGSGR----KKIYSPVYRS--DMLRKTYYKHF
Desulfovibrio desulfuricans 41  -----LKVYAEETCMRCGMCAPACHYYLSHDG----DESYSPVG-----
Aquifex aeolicus          34  -----LEACVRCGLCAEACHFYMGENISGKIDPILTPAYKADLLRETYKENY
Magnetococcus sp MC 1     59  GDLLSKYRSLQVFMDCVCKGSCCTDKCHYYLGTQD----PNNMPVQRAELMREVVYRYE
Pyrobaculum aerophilum    48  REEYGRNRNRIKTAVDTCVHCACIDACPTYLTKD----LRNSPVG-----
Desulfovibrio vulgaris    28  -----GNLNLCLTCGACSAACFPATGLEDMD-----

Archaeoglobus fulgidus      111  TLTGKLEGPLIGAKDPTEDDINALAESAY-RCTVCRRCALACPFGLDNLGITRETRKIFA
Desulfovibrio desulfuricans 75  KVEQTMWKILRSEGRILTPDDIYLMQAQIAYTECNLCRRCHYCPVGDITGYIMSTVRRICH
Aquifex aeolicus          81  TLWGRILKILFGFVVKIKPDPDIYEQVRLAYYTCTMCDRCIKLCPMGIDTPEMLVGVIRGAT
Magnetococcus sp MC 1     114  TPGGKLEPSLVKASDFN--EETLEKWFYFHCQSCQRRCSVFCPYGIDTAEVMAARELMD
Pyrobaculum aerophilum    90  --RAELLRDIIKRRKKNVDDVLELLYIYYWQCLTCRRCGYVCPFGIDQADITRVVVRGVLF
Desulfovibrio vulgaris    53  --PRKFLRMAALG-----MPEEVTTTPWVWVWCTMCMRCMYVCPMQIINIPQLVYHARASWP

Archaeoglobus fulgidus      170  DLGIVPEDEIKDNGVENQLKYGNAPKTPYEAFMDLLEFKEDIEDEKGV--EVEIPVDKKG
Desulfovibrio desulfuricans 135  RLGVTPELYIQDT-AHSHSGIMNQMWVKEDEWIDSLQWQEEBARDEIP---TVRFPTEKEG
Aquifex aeolicus          141  QIGLTPEDLVEA-TNRAIEQGSPLGVDTKTFLQRIDFISDEW-----EVEPVVDQP-
Magnetococcus sp MC 1     173  SIVGV-HKYSAEIIVGKVHDLGNLGLPKPALKGTLDFLEDDILETTGH--EVRLPLDQKG
Pyrobaculum aerophilum    148  EAGLVSRYAAMT-IDKHIDIGNNMGITPAATINILTYFANEIKSEKGV--DIEYVYVRHD
Desulfovibrio vulgaris    106  REKRPRGIVNSCDAALKTESNSAMGASPDDEAYVVEDVLEEVVRSVQPGQEKLTAPVDKKG

Archaeoglobus fulgidus      228  AKYLIMNAGDYI-----AFTETVQGIVEIMNYVGEDWTLNSPK
Desulfovibrio desulfuricans 191  ADIMYSVIAPEPK-----FRTOILYQAGVTFDQAGVNWTL--PQ
Aquifex aeolicus          191  AEYLYIPSSIELM-----KYPESVAAAAILNKSGVKTLS---
Magnetococcus sp MC 1     230  AEVLLPPSADFFSA-----PHVDSLMGYAKVPHQAGISVTIS---
Pyrobaculum aerophilum    205  QKMMLKFKGAELVGEVERSQWPEAIMYVSSADLFLNTEITLKGCLAFTHAIKLPVIVN---
Desulfovibrio vulgaris    166  AMYFLNQNSREPVTI-----PDEMVLWKILDMAGADWTYG---

Archaeoglobus fulgidus      267  TGVNDIVNYGLFYSDEDLV-RVMKAHVETAKKLDVEYLVVGECGHAMDSIAHFAKDLIPP
Desulfovibrio desulfuricans 228  TPGWNSDMAMFSGDYEMGRVKKRAEETAQRKVKRIVMGECGHAFRSIYDVGNRWLGW
Aquifex aeolicus          227  TVAHEATNFGMVFQDKKVKELLERILKGAKELGITKTISSPCGHAMQAIRFVAPNVVFP
Magnetococcus sp MC 1     227  SVASEANFGMEIGNFDQMKIAKRISQARELGVKRVIVGECGHWVAYAFWNTLNGP
Pyrobaculum aerophilum    262  TRSVEANFGMLT-HEKLMKLIHQHYIDAAKELGAKMAFGCGHWRAFKNVAVPVER
Desulfovibrio vulgaris    202  SVGWAAENYCMFAADDEAWETIVRNKVKAVEDLGCKVWLNTEUGHELYAIRSGLQKFNIK

Archaeoglobus fulgidus      326  EERPFK----VTSWMLLDQFTREGKIKLDKEKNPE---PVTIHDSCKWKRT-----G
Desulfovibrio desulfuricans 288  KWHVPV---VVSVEFFWELFTQGKIKLAK-QFEE---PVTIHDPCNIIRG-----R
Aquifex aeolicus          286  KEWDFE---VINVVEFTIKMLDEGRIKLNK-KVVD---VVTILHDSCQIGRR-----G
Magnetococcus sp MC 1     328  FDYLDPRYPVPHICEFTNDLYNRGATMDRSANDDK--ITTFHDSCNVARASRMGNPG
Pyrobaculum aerophilum    321  EGIKTY-----HIIHLLVRAIREGRIKLNPEANGDI--LYMYQDPCQYSRG-----G
Desulfovibrio vulgaris    262  PKFEIES-----ILRLYARWIREGRIPVSSEWNRERKVKFTVQDPCQLVRKSFG----D

Archaeoglobus fulgidus      372  GIYEEPRRILOACK--DREMYPN--REWNYCCGGGGGFAMAKDDFLKFRMETYGMM
Desulfovibrio desulfuricans 333  GLTEKLEVVVSFTCP--NVVEMTPN--REHNLCCAGGG-VINCGPPFKNVRVEGN-RAK
Aquifex aeolicus          331  GVLREPREILTHISQ--KFVDSVEF--PEKNICCGGGGGVVVVHEADEHRRKAFIT---K
Magnetococcus sp MC 1     386  QQFEIPRALIRASAN--RFVMDPDTIHEKTFCCGGGGGLLTDELMDLRIKGVMPRVTAL
Pyrobaculum aerophilum    366  DLIDEPRFIMNHVVK--KWVECPQN--ROLNWCCGGQAGMLADELKPLRLOYAKLWYECA
Desulfovibrio vulgaris    312  PVADDLRFVAKAVCGEENVLEMWPN--RSNNYCCGGGGGFLQSGYPARRRYGRLK----

Archaeoglobus fulgidus      428  VQQLKQFGAKIITATICSNCKAQFREMINYYNLDMRFSG-VSELVANALVYE-----
Desulfovibrio desulfuricans 387  AEQLKRELVGVITAPCHNCHGGLEDIVHHYELGMSLKF-IGDLIEYCMEKPAE-----
Aquifex aeolicus          384  MELFDKAGTKNVACYCANCLALEKSSKELNRDYKFLS-LVELVAEAEDE-----
Magnetococcus sp MC 1     444  KKVMEEDNVNFLALCAICKAQLFTVLVPYYGIPMDTVGVHQLVSNALQLGAKKGIV---
Pyrobaculum aerophilum    422  INAGAQHVVRPCMCKGTLNGVIEELNKMYKSLTFGG-VMDLVYKALVF-----
Desulfovibrio vulgaris    366  NEQIVAFGAPYVITAPCHNCHSQISDLSDHYGAYRVVH-LWTLIALSLGILCENERYLG

```

## Supplementary Figure 11 - DsrE-like

```

Methanococcus jannaschii      1 MK-----FTVITTEAPY-CKERAMTSALRFALTTALLEGIEVNIFFLENGVYVAKKEQNP
Methanopyrus kandleri        1 MGTPEGIDVITVVIHSEAPY-CQERANTALRFALTTALVEGEEVKIFLLIEDGVFLGKKGQNP
Magnetococcus sp MC 1        1 MT-----IILILFNQEPYNGSDLTWNGLRLAASLLDAECEVRIFFLNDAVDMARDACKP
Desulfovibrio vulgaris      1 -----MQILLIILSSS-DPEIKWNAVRFQNVLLGEGDDVTIFLNGPAVDLAAGDSAT
Methanosarcina mazei         1 -----MIIGIIINTS-EPETVWNAFRFCTTSLINDHEVKIFLLGRGVESENIRDEK

```

```

Methanococcus jannaschii      53 SE--VPNYLELLKNATELGAIVKVCGPCKKARGLKEED-LIEGAKLATMHDLIAFVKESD
Methanopyrus kandleri        60 DE--VPNYLELLEQCIEQGAEVKACGFCCKARGLSEED-FIEGVELATMHDLVNWKESD
Magnetococcus sp MC 1        54 AEGYDQDVVALLKLIARGVAVKVCCTCMTRCGTHKNQPYFDGAEQSTMAALAQWVSSD
Desulfovibrio vulgaris      51 FP-----IAEQAKLFLSLSEGVLAAGCKMGIHGVDAET---SLAPLSNMKFLTEQVRNAD
Methanosarcina mazei         51 FN-----VQEQIKLFMGNSCKIFACCTCLKAR-QMVGS---EVCPLSTMRDLIHIVEESN

```

```

Methanococcus jannaschii      110 NVVTF-
Methanopyrus kandleri        117 NVIFF-
Magnetococcus sp MC 1        114 RVITF-
Desulfovibrio vulgaris      103 RIINF-
Methanosarcina mazei         102 RILTFG

```

## Supplementary Figure 12 - NADH oxidase

*Deinococcus radiodurans* 1 -----MRIVIVGGVAAGMSAASRAKRFDPDAEVVVFERGDFISYGACGLP  
*Bacillus halodurans* 1 -----MKYVVIIGGDAAGMSAAMEIVRNEEAANITLTEMGSIYSYACGLP  
***Geobacter metallireducens*** 1 -----MAKELVVIIGGDAAGMSAASQAQRRRPELEIVVFERSPHTSFSAUGIP  
*Archaeoglobus fulgidus* 1 -----MRVVVIGGGAAGMSAASRVKALQPEWEVTVFETNFVSHAPCGIP  
*Methanococcus jannaschii* 1 MVNRKPNPNKNGEEMRAIIGSGAAGLITASTIRKYNKDMEIVVITKEKEIAYSPCAIP  
*Oceanobacillus iheyensis* 1 -----MSKKIIIVGGVGGGATVAAQIRRTNKTAELLVLERNGYVSFANCGMP  
*Giardia intestinalis* 1 -----GGTAVAKTIKQHPSTIEVSIYERNDNLSLSCGIA

*Deinococcus radiodurans* 46 YVLGGAVGEWDDLARTPAQMRGR-GIGVQLGHEVTGVDAAEARTVTVLDRACRVATEPY  
*Bacillus halodurans* 46 YVVGVIPTTEKLIARTIETFRNKYGDARTNHEVTKIDPDSKHVYGAN-----FEIPY  
***Geobacter metallireducens*** 48 YVVGVRVDKEEKLVIERSPEKFRKYGIDARALHEVVEIDAAQGVVVRVLRQGGSAWESY  
*Archaeoglobus fulgidus* 46 YVVEGLSDP-SHLMYYPPEFREFEKRGLDLHINAKVVEAG--DGFVRVIE--DGQEKTYEW  
*Methanococcus jannaschii* 61 YVTEGATKSFDDIIMHTPEDYKRERNIDLLTETTVIDVDSKNNKIKCVD-KDGNFEMNY  
*Oceanobacillus iheyensis* 48 YVVGGTITDRQKLIYPE-EKFAQKYDLTVQTHANVTKINREOKSVVYEK--YKTEHKADY  
*Giardia intestinalis* 36 LGVHGTIVKDMESLFYSNPDLAGL-GCVCHMQYNVTDIDFTTKKLTAVSLVTNETVTERY

*Deinococcus radiodurans* 105 DRLLVATGVSAPRPWAQTDLAGVHVLRETPDCQATADSL--KGAGR-VCIVGGGYIGLE  
*Bacillus halodurans* 100 DKLLIATGARPLVNPWPGRTAGIHTIKTIPDTEVLLADL--KGEIKNVVIVGGGYIGLE  
***Geobacter metallireducens*** 108 DQLLIATGAVPLCPDIPGSDAVDICGVNLTLESGLERRRLD-KGGMKGVVGGGYIGLE  
*Archaeoglobus fulgidus* 101 DKLVATGALPKTPPEFEGLELVNFTVRHPVQAALREAV---EKAENVVIVGAGYVGVVE  
*Methanococcus jannaschii* 120 DYLVLATGAEFFIPPEEGKDLGVFKVRTIEDGRATLKYIE-ENCKKVAVVGAGIIGLE  
*Oceanobacillus iheyensis* 105 DILLSPGASVLPDITAGLQSNTPFLRTIEDMDNTEQFIQ-SNNPQSAATIGGGHIGLE  
*Giardia intestinalis* 95 DKIVFATGSWPIIPDIPGKSDKVLKCNMHAKKIVETFSHEENTKHCVVIGAGYIGVE

*Deinococcus radiodurans* 162 LAENMCROGLSVVLLERNPDVGGRVLDPEYRPRLLDELGRHGVDVRCGTTVEGLIKGAGR  
*Bacillus halodurans* 158 MAENLALTCKNVTIVBANAQLAA-IFDQEMGEIHHQEAERKGVTLRLKEEVKGFEG-TDR  
***Geobacter metallireducens*** 167 MAEALVRHGLEVSLVNRAPQVVG-TLDYDMGAMVSQALRDVGVSLYLEETLTAFTKGGK  
*Archaeoglobus fulgidus* 158 MAEAAAARGKKVTVVEFLDQPLP-NLDRDVADLVKHKLEEK-VNLRRLGKVEAFEG-DGA  
*Methanococcus jannaschii* 179 MAYGLKCGLDVLLVEMAPQVLPFRFLDPMAEIVQKYLEKEGKVKVMSKPLEKIVG-KEK  
*Oceanobacillus iheyensis* 164 MAESLCHRGFSCSLVDRSEHVLK-RIDKEMAIHI DEHLQEKGTALYVNDGLKSFSD--NG  
*Giardia intestinalis* 155 LAEAFGLKQPCILLDGSGRINSRNFDKEFTDICEDEMRAHGVLQMGRLAEAFDE-DDG

*Deinococcus radiodurans* 222 VTGVQTDGGLVLRADVVAAGVVKPNVDLRAAGARIGK-----TGAAAVDVVQQTNVVD  
*Bacillus halodurans* 216 VQAVVTSATVPAELVITAIAGVVPNTTFLEGQPFRRHE-----NSALKVNAYMETNLP  
***Geobacter metallireducens*** 226 VTGVVTDRTLPADIVILGLVGRPNTALASAAGLPIGE-----KSRVNRNEMQTGVA  
*Archaeoglobus fulgidus* 215 VRKVVTDKGEYPADVVIVATGVKANTATAEQIGCKIGE-----TGATWDSRMQTSVE  
*Methanococcus jannaschii* 238 VEAVYVDGKLYDVMVIMATGVRPNIELAKKAGCKIG-----KFAIEVNEKMQTSIP  
*Oceanobacillus iheyensis* 221 TTLHLSSDKTIQADMTIMATGKPNTELAIDAGLEIGE-----TGIKVNQVMQTTDP  
*Giardia intestinalis* 214 KIVVRTSKGVYSGDAAILCIGFRPVTEMILESAERHGVKLDVHHPKSAIITDECARTSLP

*Deinococcus radiodurans* 275 GVYAAGDNCESLRVTRRRVHTPLGLTANRMGRJAGINMAGG-DAKFPGVVGTSTFKVFG  
*Bacillus halodurans* 269 DTYAAGDCATQYHRKLLDDYIPLGTHANKQGRLAGLNMGK-RRAFAGVVGTSIIFKFFD  
***Geobacter metallireducens*** 279 GTWAAGDCAESFHLVSRKPVHIALGTVANRHGRVAGINLGGG-YATFPGVVGTAVTKICQ  
*Archaeoglobus fulgidus* 268 NVFAAGDCAETHTMTKKRWVPLAPPENKMGYVAGVNAAGG-NIEFPVGLGTQITKFFD  
*Methanococcus jannaschii* 290 NIYAVGDCVEVIDFITTEKTLSEFGTAAVRQGVKAKNIAQV-EAKEYPVLNSAVSKIGD  
*Oceanobacillus iheyensis* 274 TIFALGDAVEVTDFTTREPAAHALAWPAHROAFIISFSLSN-PISDDGLIGSSILRVFD  
*Giardia intestinalis* 274 GVYAIGDCATIHXYTWDDEDRYVPLATNAIRTGLAAAHILKHLRLIGTEGTSIRIFK

*Deinococcus radiodurans* 334 LGVARTGLTQGEAALGLN-AVSVDVTSTDHAGYYADARPIHVRITGERTGRLLGGQIV  
*Bacillus halodurans* 328 LSLGRTGLSEKETRDARLP-ASSITFDQRDIAGYYPGAEPKIKLVHSETNQLLGGQVI  
***Geobacter metallireducens*** 338 VEVARTGLQEEELRELIGIE-WISAVIKSRTRAGYEPGAGGITVKKVLAERSGRLLGGQIV  
*Archaeoglobus fulgidus* 327 LEIGATGLTEKAAKAEAGFE-VKTAVVKKTRVHYYPGAKDTFLKVVADASTKRLLGAQVL  
*Methanococcus jannaschii* 349 LEIGGTGLTAFSANLKRIP-IVIGRAKALTRARYYPGGKELEIKMIBND-GKVVLCQIV  
*Oceanobacillus iheyensis* 333 LITVAATGQNKQTLIDNEIA-FEETIMKCYSHAAYYPGSKELWVQIVYDKNTGQLGGSVI  
*Giardia intestinalis* 334 HHMASTGLTEEGALLAGIKNVKSVIINDTDRPCEMFTNANVMVKLVYDGDTHRVLGGQMM



```

Deinococcus radiodurans 393 AENPLSVKRVVLA A A H L H S R G K V E D L F Q M D L A Y A P P F S G V W D V L L V A A D R L N R A L R A ---
Bacillus halodurans    387 G T N - G V A K R V D V L A T A L Y Q Q L T L E E L L D L D L S Y A P P F N G V W D P I Q Q A A R R A -----
Geobacter metallireducens 397 G M E - G S A K R I D T L A T A L H A G F T V E E M I N L D L G Y A P P F S P V W D P V V I A A R E V A K E L -----
Archaeoglobus fulgidus 386 G A D - - V A M R V N V F A A M I Q G G F T T K D V F F A D L G Y A P P F T P T W D P I V V S A R I L K F -----
Methanococcus jannaschii 407 G G E - R V A E R I D A M S T A I F K R V S A E E L A N M E F C Y A P P V S M V H E P L S L A A E D A L K K L S N K --
Oceanobacillus iheyensis 392 G F D - G A D K R M A V L A T A I K G K L T V A D L A S L E L G Y S P T Y S G A R D P L N I L G Y K A M E Q L E D ---
Giardia intestinalis    394 S E A - S V N Q V M N A L A V C I Q N R Y T I E K M A V Q D F F E Q E H Y N N P W N T L N V A A I S A I N K K D L V M S

```

```

Deinococcus radiodurans -----
Bacillus halodurans -----
Geobacter metallireducens -----
Archaeoglobus fulgidus -----
Methanococcus jannaschii -----
Oceanobacillus iheyensis -----
Giardia intestinalis 453 KYLAQALCGCALQAILRG

```

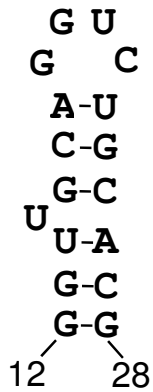
## Supplementary Figure 13 - Distant homolog of peroxidase/peroxynitrite reductase system component AhpD

*Neisseria meningitides* 1 MFKDWKEHTALVKKSFGLGKAHPKMLQAYGALQAAAA-EALDAKTRELIATAVAVTTR  
*Haemophilus influenzae* 1 MFTDWKEHTSHVKKSFGLGKQYPKMLQAYQALGAAAEGNVLDAKTRELIATAVAVTTR  
*Pseudomonas aeruginosa* 1 MLNNWTEFVPAVKKAFGLGKQHPKMLAAYGALQAAAAE-CALDAKTRELISAVAVTTR  
*Methanothermobacter thermautotrophicus* 1 -----MKTGADRFLLEELPEVAESFKNFREAVRSECKLTEREKILLISVACSVAVR  
*Methanosarcina mazei* 1 -----MTEKIDMDERAKMAEETFGVMKALMGLHSEVVKDGALESAKTEELMMVGIAVAIR  
*Geobacter sulfurreducens* 1 -----MLGHEVHCKAMAMKIRKKILDFEY----EEVLDARTRELIRVGCVAVAVG

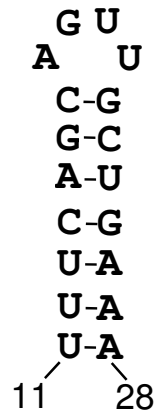
*Neisseria meningitides* 60 CES<sup>C</sup>ISVHAAAATKAGATDSEIAGALATAIALNAGAAYTY-SLRALAEAVETOK-  
*Haemophilus influenzae* 61 CES<sup>C</sup>ISAHAEAAVKAGASEAEVAALATAIALNAGAAYTY-SLRALAEAYSVOKA  
*Pseudomonas aeruginosa* 60 CD<sup>C</sup>IGVHTEAAKAGASEAIACTLATAISLNAGAAYVY-SLRALAEAYDQFKK  
*Methanothermobacter thermautotrophicus* 50 CD<sup>C</sup>ACTRRHAEAAEAGITEGELAEAAVAALTRAGSAMNT-SAIIFRD-----  
*Methanosarcina mazei* 56 CEYCLWKHVPKMGATREEILEAVSTAIMMSSGGPQVAYGSSVVLKILDELNV  
*Geobacter sulfurreducens* 47 CPT<sup>C</sup>ULKKHFAAAKEAGATDAELKEALAYGIIAPSGRAKNE-VLNMAGELLELGD-

## Known selenoproteins

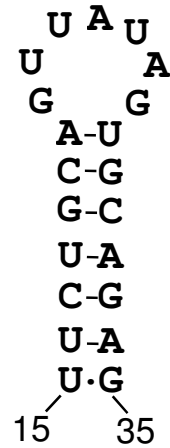
Formate dehydrogenase  
*Escherichia coli*



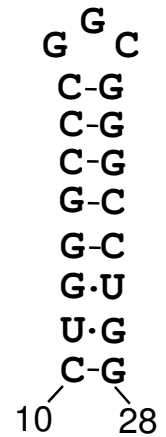
Peroxioredoxin  
*Eubacterium acidaminophilum*



Proline reductase  
*Clostridium sticklandii*



Selenophosphate synthetase  
*Geobacter sulfurreducens*

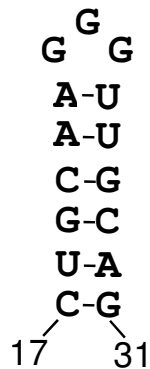


## New selenoproteins

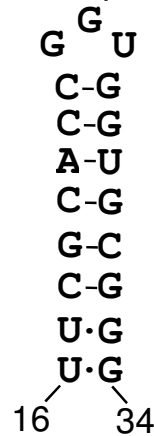
HesB-like protein  
*Desulfovibrio vulgaris*



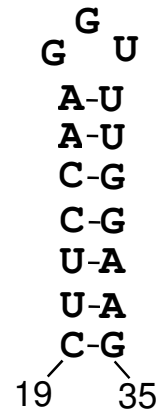
SelW-like protein  
*Campylobacter jejuni*



Thioredoxin  
*Geobacter sulfurreducens*

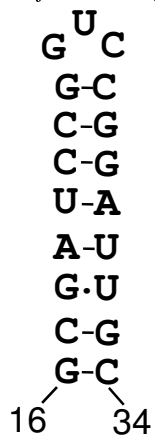


Glutathione peroxidase  
*Treponema denticola*

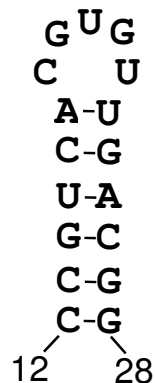


## Candidate selenoproteins

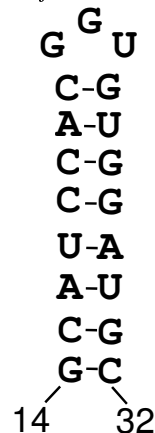
Fe-S oxidoreductase  
*Desulfovibrio vulgaris*



DsbG protein  
*Chloroflexus aurantiacus*



DsrE-like  
*Desulfovibrio vulgaris*



Distant AhpD homolog  
*Geobacter sulfurreducens*

