ANTIOXIDANTS & REDOX SIGNALING
Volume 16, Number 3, 2012

DOI:10.1089/ars.2011.3980

ORIGINAL RESEARCH COMMUNICATION

Comparative Genomics of Thiol Oxidoreductases Reveals Widespread and Essential Functions of Thiol-based Redox Control of Cellular Processes

Dmitri E. Fomenko[1] and Vadim N. Gladyshev[2]

Abstract

Aims: Redox regulation of cellular processes is an important mechanism that operates in organisms from bacteria to mammals. Much of the redox control is provided by thiol oxidoreductases: proteins that employ cysteine residues for redox catalysis. We wanted to identify thiol oxidoreductases on a genome-wide scale and use this information to obtain insights into the general principles of thiol-based redox control. ***Results:*** Thiol oxidoreductases were identified by three independent methods that took advantage of the occurrence of selenocysteine homologs of these proteins and functional linkages among thiol oxidoreductases revealed by comparative genomics. Based on these searches, we describe thioredoxomes, which are sets of thiol oxidoreductases in organisms. Their analyses revealed that these proteins are present in all living organisms, generally account for 0.5%–1% of the proteome and that their use correlates with proteome size, distinguishing these proteins from those involved in core metabolic functions. We further describe thioredoxomes of *Saccharomyces cerevisiae* and humans, including proteins which have not been characterized previously. Thiol oxidoreductases occur in various cellular compartments and are enriched in the endoplasmic reticulum and cytosol. ***Innovation:*** We developed bioinformatics methods and used them to characterize thioredoxomes on a genome-wide scale, which in turn revealed properties of thioredoxomes. ***Conclusion:*** These data provide information about organization and properties of thiol-based redox control, whose use is increased with the increase in complexity of organisms. Our data also show an essential combined function of a set of thiol oxidoreductases, and of thiol-based redox regulation in general, in all living organisms. *Antioxid. Redox Signal.* 16, 193–201.

Introduction

Redox regulation of cellular processes is an important and widespread regulatory mechanism (8). Much of the redox control is provided by proteins that utilize catalytic cysteine (Cys) residues for oxidation or reduction of their substrates (11). These proteins are collectively known as thiol oxidoreductases. While mechanistically similar, thiol oxidoreductases involve structurally distinct families of enzymes, which can act on proteins, lipids, and compounds. The past several years have seen growing evidence for the important roles of thiol-based redox regulation in various metabolic pathways, as well as in aging, cancer (22), neurodegenerative diseases (2, 3, 24, 27), diabetes (13), and other pathophysiological processes. At the molecular level, thiol oxidoreductases regulate a variety of biological functions and are involved in antioxidant defense, signal transduction, metabolic processes, and protein folding and modification.

Many thiol oxidoreductases have previously been identified and characterized; however, understanding of the overall sets of such proteins in organisms and of their combined and protein-specific functions is quite limited (11, 12). This is because these proteins are difficult to identify by available protein function prediction methods (5–7, 10). Two thiol-based redox

Innovation

In this study, we introduced the concept of thioredoxomes (*i.e.*, sets of thiol oxidoreductases in organisms). We also developed computational tools for the identification of these proteins on a genome-wide scale. Using these tools, we characterized thioredoxomes in various organisms, including those of yeast and humans, which in turn revealed the composition and properties of thioredoxomes and, therefore, of thiol-based redox control in general.

[1]Department of Biochemistry and Redox Biology Center, University of Nebraska-Lincoln, Lincoln, Nebraska.
[2]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts.

systems, thioredoxin and glutaredoxin systems, have been particularly well characterized (14, 15), with many thousands of publications on each of their components, including thioredoxin, thioredoxin reductase, glutathione reductase, glutaredoxin, glutathione peroxidase, and peroxiredoxin. Several additional systems that are based on thiol oxidoreductases include the prokaryotic disulfide bond formation (Dsb) system (16) and the protein disulfide isomerase-based system located in the endoplasmic reticulum (ER) of eukaryotes (4, 25). Methionine sulfoxide reductases, which are proteins that repair oxidatively damaged methionines, also received much attention (17, 18, 20), and several other thiol-based redox regulatory processes have also been described (12).

Catalytic redox-active Cys residues in thiol oxidoreductases are highly conserved, and many such proteins have homologs, in which Cys is replaced with selenocysteine (Sec) (12). In these homologous selenoproteins, Sec is also the catalytic redox-active residue. However, while Cys residues may have many functions, Sec is always a catalytic redox residue. Thus, identification of a Sec residue in the position corresponding to Cys in a protein indicates a redox function of this protein and points to the exact location of the catalytic redox-active Cys residue. Although selenoproteins are relatively rare, recent dramatic increase in sequence information derived from various genome and other sequencing projects allows efficient identification of both selenoproteins and the corresponding Cys-containing thiol oxidoreductases.

In addition, many thiol oxidoreductases form protein complexes, in which reducing equivalents are transferred from one thiol oxidoreductase to another, both containing conserved catalytic Cys residues. If one such protein is a known thiol oxidoreductase, establishing a functional link between the two proteins may identify the second protein as an additional thiol oxidoreductase. For example, many proteins that form thiol oxidoreductase pathways transfer reducing equivalents from one protein to another to show fusion events. These proteins also show similar transcriptional regulation, and in bacteria, they are often located in the same operons. These considerations suggest a second approach to identify thiol oxidoreductases. Thiol oxidoreductases can also be identified by extensive homology analyses, as many of these proteins cluster within several protein folds, but diverge to the extent that standard BLAST analyses do not detect homology (12). Thus, exhaustive homology analyses within the sequence universe offer a third independent approach to detect thiol oxidoreductases.

In this work, we used these three strategies to carry out genome-wide searches for thiol oxidoreductases and further characterized the overall functions of thioredoxomes in organisms. These data illustrate the universal use of thiol oxidoreductases in cellular life and show the essential and widespread functions of thiol-based redox control in regulation of cellular processes.

Results

Genome-wide identification of thiol oxidoreductases

Thiol oxidoreductases are involved in a variety of biological functions (Fig. 1) and are characterized by different folds, protein length, and location of their catalytic redox-active Cys residues (12, 21). Various classes of thiol oxidoreductases show no similarity with regard to patterns of occurrence in organisms. To examine the use of thiol oxidoreduction in organisms across the three domains of life, we sought to identify these enzymes on a genome-wide scale. We used three independent methods: (i) identification of Cys/Sec pairs in homologous sequences; (ii) comparative genomics approach that functionally links known and unknown thiol oxidoreductases through domain fusions, and (iii) exhaustive homology searches starting from known thiol oxidoreductases.

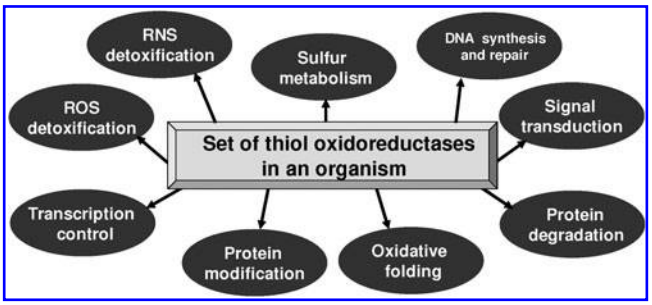

FIG. 1. Functions of thiol oxidoreductases. Thiol oxidoreductases are involved in various processes, such as sulfur metabolism, DNA synthesis and repair, signaling, protein degradation, oxidative folding, protein modification, regulation of gene expression, and ROS/RNS detoxification. These proteins likely also have other, currently unknown, functions in biology.

The first method (Fig. 2A) is based on the observation that the majority of known thiol oxidoreductases have homologs, in which the catalytic redox Cys is replaced with Sec. Whereas only some Cys residues serve thiol oxidoreductase functions in proteins (11), Sec always serves redox function. Thus, identification of a pair of homologs, wherein one is a Sec-containing protein and another a Cys-containing protein, suggests a catalytic redox function for the Cys that aligns with Sec and is flanked by homologous regions. This method is relatively simple, efficient, and is independent of protein family, protein structure, and organism of origin (12). To identify thiol oxidoreductases on a genome-wide scale, we used Cys-containing protein sequences from the non-redundant NCBI database. We also utilized a nucleotide sequence database that included all completely sequenced genomes, expressed sequence tags (ESTs), and environmental genome projects from NCBI. These datasets were then cross-analyzed, with TBLASTN, to identify Cys/Sec pairs.

The second method (Figs. 2B–2D) took advantage of the observation that thiol oxidoreductases often form functional complexes/modules, wherein one thiol oxidoreductase acts on another. For example, thioredoxin reductase provides reducing equivalents to thioredoxin, which in turn reduces peroxiredoxin. All these proteins are thiol oxidoreductases containing catalytic redox-active Cys residues. These linkages can be captured by comparative genomics methods (*i.e.*, these proteins often form fusions or are located in the same operon in prokaryotes). Continuing with the example shown above, both thioredoxin reductase-thioredoxin and thioredoxin-peroxiredoxin fusion proteins are known, and, sometimes, these proteins cluster in operons. Thus, to identify novel thiol oxidoreductases, we employed known thiol oxidoreductases, searched for domains fused with these proteins in representative organisms, and further filtered out the set to select the fused domains that contained conserved Cys residues.

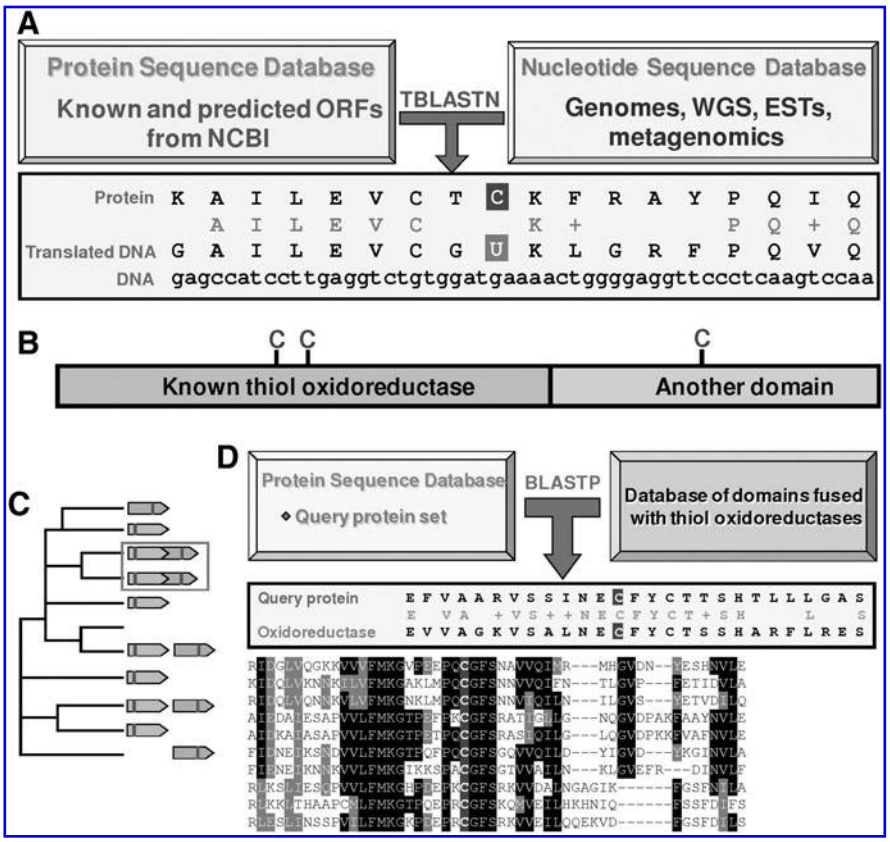

FIG. 2. Methods for high-throughput identification of thiol oxidoreductases. (A) The Sec/Cys approach. Thiol oxidoreductases are predicted by homology to selenoproteins, such that the candidate redox-active Cys in a thiol oxidoreductase aligns with Sec, and the pair is flanked by homologous sequences. The search is based on TBLASTN searches that analyze a large protein database that is searched against a large nucleotide sequence database. **(B)** A domain with a conserved cysteine, which is fused to a known thiol oxidoreductase, is often a thiol oxidoreductase itself. **(C)** A domain fusion event involving a thiol oxidoreductase and another domain with a conserved Cys indicates a functional relationship between these proteins even in organisms in which such fusion is not observed. **(D)** The approach involving domain fusion is illustrated, which analyzes two protein databases to identify new thiol oxidoreductases.

For comparison, we developed a third set of predicted thiol oxidoreductases by carrying out exhaustive PSI-BLAST searches starting with known thiol oxidoreductases. By this method, distant homologs of these proteins could be identified that conserve the catalytic redox Cys. The three approaches were then applied to various sets of open reading frames (ORFs), proteomes of model organisms, and other protein datasets (Fig. 3A). All three approaches were efficient in detecting thiol oxidoreductases. For example, the Cys/Sec method detected 27,701, domain fusion 17,799, and PSI-BLAST 20,367 proteins in 803 completely sequenced bacterial genomes (Fig. 3*B*). In addition, these sets of proteins were highly overlapping, with the domain fusion and PSI-BLAST protein sets being essentially a subset of the Cys/Sec set. Since the PSI-BLAST set included all known thiol oxidoreductases, the Cys/Sec approach was clearly highly efficient in identifying these proteins: it found essentially all such proteins (and also predicted additional candidate thiol oxidoreductases). Based on this information, we defined sets of thiol oxidoreductases in organisms as those included in either Cys/Sec or PSI-BLAST sets. We further designate these sets as thioredoxomes. It should be noted that, in this work, we did not analyze non-catalytic redox Cys, which are subject to posttranslational modification, such as glutathionylation and S-nitrosylation.

Properties of thioredoxomes

The basic properties of thioredoxomes were further examined. First, we observed a consistent increase in the number of thiol oxidoreductases with the increase in the proteome size. The dependence was approximately linear (Fig. 4A). This pattern contrasted with the changes in the abundance of basic metabolic enzymes, such as glycolytic enzymes (Fig. 4A), whose numbers only slightly increased in organisms with the larger proteomes. This observation held for thiol oxidoreductases identified with any of the three methods (Fig. 4B). The thioredoxin fold dominated thioredoxomes, accounting for approximately 50% of thiol oxidoreductases in organisms (Supplementary Fig. S1A; Supplementary data are available online at www.liebertonline/ars). In addition, we found that thiol oxidoreductases occurred in all organisms, indicating an essential function of thiol oxidoreduction (even though many proteins in this system are not essential).

We also found a strong correlation between the number of thiol oxidoreductases and the number of catalytic redox-active Cys residues (Supplementary Fig. S1B). The latter was slightly larger: the majority of thiol oxidoreductases had a single catalytic Cys, but some of these enzymes were composed of two or more thiol oxidoreductase domains, each having the catalytic redox Cys. The largest number of catalytic redox-active Cys in a thiol oxidoreductase was four (in protein disulfide isomerase). We observed a correlation with the proteome size for both thioredoxin-fold and non-thioredoxin-fold thiol oxidoreductases (Supplementary Figs. S1C and S1D), suggesting an even distribution of these protein groups in organisms.

Analysis of thioredoxomes of various bacterial phyla revealed a highly variable occurrence of thiol oxidoreductases (Fig. 4C). This distribution correlated with the genome/proteome size rather than with being a part of a particular

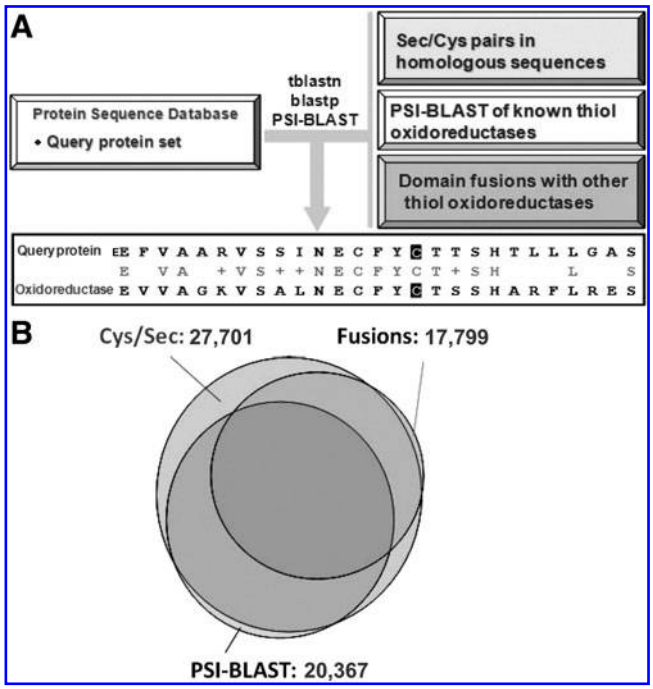

FIG. 3. High-throughput identification of thiol oxidoreductases. (A) Identification of thiol oxidoreductases in protein databases and proteomes of individual organisms. Three approaches were used: (i) Sec/Cys pairs in homologous sequences; (ii) Domain fusion events involving thiol oxidoreductases; and (iii) PSI-BLAST-based searches for distant homologs of known thiol oxidoreductases with conserved catalytic redox-active Cys. **(B)** Predicted sets of thiol oxidoreductases in 803 completely sequenced bacterial genomes. Cys/Sec, Fusion, and PSI-BLAST refer to the three approaches used to predict thiol oxidoreductases. Total numbers of thiol oxidoreductases predicted by each approach are listed and their overlap is shown in the figure.

bacterial group or phylum. These properties of prokaryotic thioredoxomes were also evident in eukaryotes. Eukaryotes had larger thioredoxomes, but only because their proteomes were larger. When prokaryotic and eukaryotic thioredoxomes were normalized to the proteome size or compared in organisms with equivalent proteome size, the prokaryotic thioredoxomes were actually slightly larger (Fig. 4D).

The smallest thioredoxomes corresponded to the archaeon *Nanoarchaeum equitans* and seven bacteria from the *Borrelia* taxon (*B. afzelii, B. burgdorferi, B. burgdorferi, B. duttonii, B. garinii, B. recurrentis,* and *B. turicatae*). *N. equitans* has three thiol oxidoreductases: thioredoxin reductase, thioredoxin and peroxiredoxin (Fig. 5A), whereas Borrelias have thioredoxin reductase, thioredoxin, and coenzyme A disulfide reductase. Interestingly, *B. hermsii* has an additional glutathione peroxidase. The largest detected thioredoxome, that of *Arabidopsis thaliana*, consisted of 194 proteins (Fig. 5B).

An NADPH-dependent thioredoxin system, consisting of thioredoxin reductase, thioredoxin, and peroxiredoxin, is the main system that controls the redox state of Cys in proteins, and we found that it was present in essentially all analyzed organisms. However, host-associated *Phytoplasma asteris* Onion Yellows (an organism with 1021 ORFs) lacked the thioredoxin system. Yet, this organism still had four thiol oxidoreductases: arsenate reductase, rhodanese-like protein, lipoamide dehydrogenase, and hydroperoxide reductase OsmC. Further analysis of bacterial habitats shown that host-associated organisms had fewer thiol oxidoreductases (Supplementary Fig. S2). We hypothesize that phytoplasma Onion Yellows and certain other host-associated organisms may access thiol oxidoreductases of the host for their metabolic/regulatory needs. In contrast to host-associated organisms, terrestrial organisms had greater numbers of thiol oxidoreductases (Supplementary Fig. S2).

Representative thioredoxomes

Next, we examined thioredoxomes of model organisms. The *S. cerevisiae* set of thiol oxidoreductases had 47 proteins covering 31 conserved domains, including 10 that previously have not been characterized with regard to thiol oxidoreductase function. It would be of interest to characterize these candidate thiol oxidoreductases. Subdivided by protein family, the largest families of thiol oxidoreductases in *Saccharomyces cerevisiae* were glutaredoxin-like proteins (five monothiol and three dithiol glutaredoxins), protein disulfide isomerases (six proteins), peroxiredoxins (five proteins), ubiquitin-activating-like proteins (three proteins), thioredoxins (three proteins), and sulfhydril oxidases and methionine sulfoxide reductases (Fig. 6). The majority of yeast thiol oxidoreductases localized to cytosol (20 proteins) and the ER (10 proteins).

The human thioredoxome consisted of 111 proteins containing catalytic redox-active Cys and 25 proteins containing Sec, for a total of 136 proteins (Fig. 7A). The largest protein families were thioredoxin-like (19 proteins) and protein disulfide isomerase-like proteins (12 proteins) (Fig. 7B). As in yeast, the majority of these proteins localized to the cytosol and the ER (42 and 34 proteins, respectively), whereas there was lower abundance of mitochondrial thiol oxidoreductases (based on the presence of a predicted signal peptide and available localization data) (Fig. 7C).

Roles of environmental factors in influencing thioredoxome composition

We examined a possible influence of environmental factors, such as growth temperature and oxygen content, on the size of thioredoxomes. Thermophilic and hyperthermophilic showed smaller thioredoxomes in comparison with psycophilic organisms (Supplementary Fig. S3). Oxygen content also did not influence thioredoxomes significantly, with the exception of the microaerobic group, which had fewer thiol oxidoreductases (Supplementary Fig. S4). It should be noted that our observations do not exclude a possible role of environmental factors in regulating thiol oxidoreductase gene expression.

Distribution of thiol oxidoreductase families in bacterial phyla. We analyzed the distribution of thiol oxidoreductase families in various bacterial phyla and found that common thioredoxin-fold proteins, such as thioredoxins and peroxiredoxins, were present in almost all searched organisms. However, other thiol oxidoreductases, such as DsbB, GST, DsrE, and HesB, had scattered patterns of occurrence. This observation suggested that some thioredoxin-fold oxidoreductases formed a core of thioredoxomes while other thiol oxidoreductases had more specialized functions.

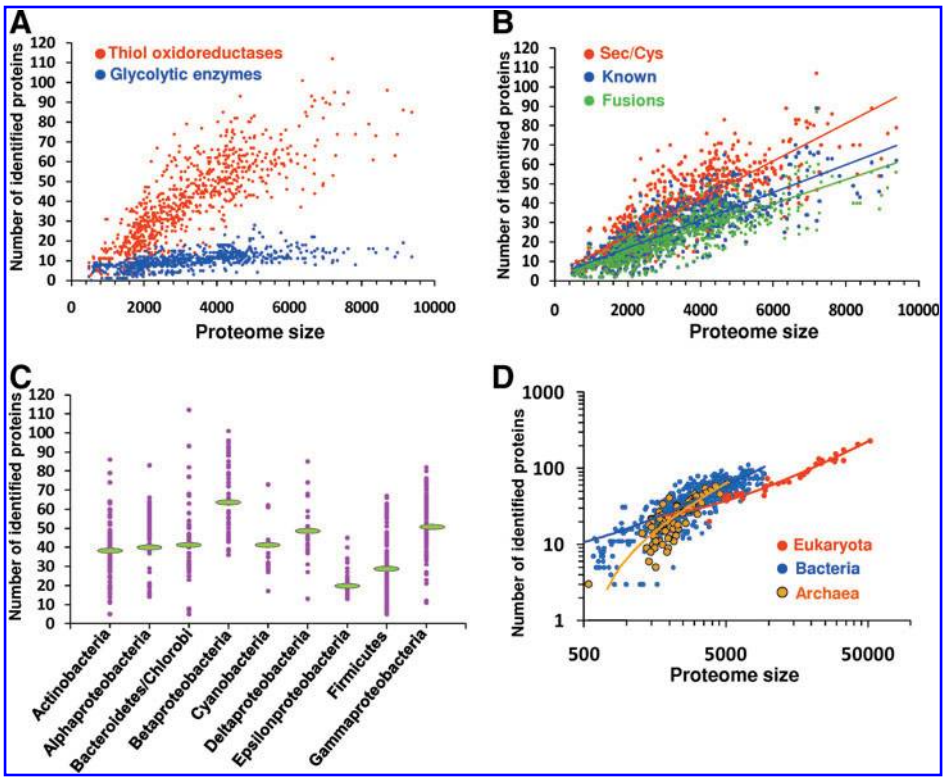

FIG. 4. Bacterial thioredoxomes. (A) Occurrence of thiol oxidoreductases in bacteria. Thiol oxidoreductases were predicted for 803 completely sequenced bacterial genomes. Glycolytic enzymes predicted for the same set of organisms are shown for comparison. **(B)** Identification of bacterial thiol oxidoreductases by the Sec/Cys, domain fusion and homology to known thiol oxidoreductases approaches. **(C)** Distribution of thioredoxomes in different phyletic groups of bacteria. An average number of thiol oxidoreductases in each group is marked with *green*. **(D)** Comparative distribution of thioredoxomes in eukaryota, bacteria, and archaea. (To see this illustration in color the reader is referred to the web version of this article at www.liebertonline.com/ars).

Database and computational resources

As a result of our work, several web tools were developed (located at two mirror sites: http://genomics.unl.edu/THIOREDOXOME and http://gladyshevlab.org/THIOREDOXOME) to facilitate identification and classification of thiol oxidoreductases. They utilize a non-redundant dataset of known selenoproteins and their Cys homologs (1065 proteins), a database of thiol oxidoreductase domain fusions events (513 proteins), a database of known thiol oxidoreductases (597 proteins), and a non-redundant database of thiol oxidoreductases (1185 proteins). A query can be entered as a separate protein or a set of proteins, and the output identifies the type of thiol oxidoreductase and highlights its catalytic redox Cys. In addition, we provide a database of identified thiol oxidoreductases in various organisms.

Discussion

Thiol oxidoreductases form a group of functionally-related proteins that utilize redox chemistry of their catalytic Cys for redox regulation of cellular processes. These proteins are represented by several protein families and folds (2, 3, 11–13, 22, 24, 27) and the extent of their utilization in biology has been unclear. Thiol oxidoreductases are also important because they could serve as a tool to understand organization and mechanisms of thiol-based redox control. Nearly every cellular process is now known to be regulated by the redox status of catalytic Cys residues in proteins, as well as by post-translational modifications, such as S-nitrosylation, glutathionylation, and disulfide bonding. However, Cys residues in proteins have many functions, such as metal coordination, structural stabilization, membrane targeting, post-translational modifications, and redox and non-redox catalysis (11). Thus, sequence- and structure similarity-based

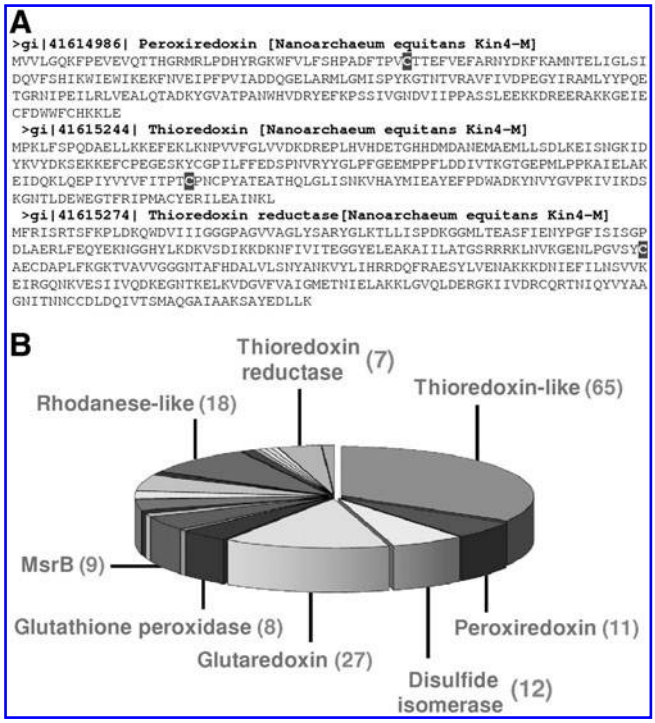

FIG. 5. Smallest and largest thioredoxomes detected. (A) One of smallest thioredoxomes. An archaeon *Nanoarchaeum equitans* has 535 ORFs, among which three are thiol oxidoreductases, each having a single catalytic redox Cys. **(B)** The largest thioredoxome. 194 thiol oxidoreductases were identified in *Arabidopsis thaliana*.

Domain	Protein name	Protein Function	Localization	Redox motif
cd03016	PRX1	Peroxiredoxin	Mitochondria	TPVC
cd03015	TSA1	Peroxiredoxin	Cytosol	TFVC
cd03015	TSA2	Peroxiredoxin	Cytosol	SFVC
cd03013	AHP1	Peroxiredoxin	Cytosol	CTVS
cd03017	DOT5	Peroxiredoxin	Nucleus	TPGC
cd00340	GPX1	Glutathione peroxidase	Cytosol	CAFT
cd00340	GPX2	Glutathione peroxidase	Cytosol	CGFT
cd00340	GPX3	Glutathione peroxidase	Cytosol	CGFT
cd03419	GRX1	Glutaredoxin	Cytosol, nucleus	CPYC
cd03419	GRX2	Glutaredoxin	Cytosol	CPYC
cd03028	GRX3	Monothiol glutaredoxin	Cytosol, nucleus	CGFS
cd03028	GRX4	Monothiol glutaredoxin	Nucleus	CGFS
cd03028	GRX5	Monothiol glutaredoxin	Mitochondria	CGFS
cd03419	GRX6	Monothiol glutaredoxin	Vacuole	CPYS
cd03419	GRX7	Monothiol glutaredoxin	Vacuole	CSYS
cl00388	GRX8	Glutaredoxin-like	Cytosol	CPDC
cl09162	TRR1	Thioredoxin reductase	Cytosol	C
cl09162	TRR2	Thioredoxin reductase	Mitochondria	C
pfam04756	OST3	Thiol oxidoreductase from OST complex	ER	CSLC
pfam04756	OST6	Thiol oxidoreductase from OST complex	ER	CQLC
cd02947	TRX1	Thioredoxin	Cytosol	CGPC
cd02947	TRX2	Thioredoxin	Cytosol	CGPC
cd02947	TRX3	Thioredoxin	Mitochondria	CGPC
cd02989	PLP1	Phosducin-like protein	Cytosol	C
cd02995	PDI1	Protein disulfide isomerase	ER	CGHC, CGHC
cd02995	EUG1	Protein disulfide isomerase	ER	CLHS, CIHS
cd03002	MPD1	Protein disulfide isomerase	ER	CGHC
cd02961	MPD2	Protein disulfide isomerase	ER	CQHC
cd02961	ESP1	Protein disulfide isomerase	ER	CPHC
cd01490	UBA1	Ubiquitin activating enzyme (E1)	Cytosol, nucleus	C
cd01489	UBA2	SUMO activating enzyme complex	Nucleus	C
cl00287	UBA3	Ubiquitin activating enzyme	Cytosol	C
cd00757, cd01526	UBA4	URM1 activating enzyme	Cytosol	C, C
cd01449	YOR251C	Thiosulfate sulfurtransferase	Cytosol	C
cd01519	YOR285W	Rhodanese	ER, Mitochondria	C
cd01519	AIM42	Rhodanese-like	Mitochondria	C
cl00366	MSRA	Methionine sulfoxide reductase A	Cytosol	C
cl00369	MSRB	Methionine sulfoxide reductase B	Mitochondria	CVNS
cl00853	MSRC	Free methionine sulfoxide reductase	Cytosol, Mitochondria	C
cl00400	ISA1	Isa1p (HesB-like)	Mitochondria	C
cl01497	MGT1	DNA repair methyltransferase	Nucleus	C
PRK06327	LPD1	Lipoamide dehydrogenase	Mitochondria	C
COG5061	ERO1	ERO1	ER	C
cl02107	ERV1	ERV1	Mitochondria	C
cl02107	ERV2	ERV2	ER	C
PRK06116	GLR1	Glutathione reductase	Cytosol, mitochondria, nucleus	C
cl00428	YPL222W	Selenoprotein O	Mitochondria	C

FIG. 6. The *Saccharomyces cerevisiae* thioredoxome. This thioredoxome has 47 proteins, which are listed in the figure with regard to domain name, protein name, protein function, cellular localization, and the redox motif that includes a catalytic redox-active Cys residue.

algorithms alone cannot be used for the identification of thiol oxidoreductases and characterization of sets of these proteins in organisms (5–7, 9, 10).

In this work, we defined the overall use of thiol oxidoreductases by organisms in the three domains of life. Much of the previous work focused on the characterization of known thiol oxidoreductases, especially proteins of the thioredoxin fold, such as thioredoxins, glutaredoxins, protein disulfide isomerases, peroxiredoxins, and glutathione peroxidases. However, how many such proteins as well as other (especially nonthioredoxin fold) thiol oxidoreductases are present in organisms was not clear. We applied three independent methods for genome-wide identification of thiol oxidoreductases. The most efficient approach was the one that identified Sec/Cys pairs flanked by conserved sequences (12, 19). This method is based on the observation that Sec residues in proteins exclusively serve redox functions and that these redox Sec are replaceable with Cys only in the catalytic sites of thiol oxidoreductases. Therefore, identification of Sec/Cys pairs indicates the identity of redox-active catalytic Cys residues in proteins. This approach had the best sensitivity of all tested approaches and detected nearly all known thiol oxidoreductase families. Two other methods analyzed fusions of thiol oxidoreductases to other domains containing conserved Cys residues and carried out exhaustive searches for distant homologs of known thiol oxidoreductases. These two methods were somewhat less sensitive than the Sec/Cys approach; however, their selectivity was still sufficient to use them as alternative methods for thiol oxidoreductase prediction (Fig. 4B). A combination of the three methods then offered an opportunity to describe the sets of thiol oxidoreductases in organisms.

With this approach, we showed that thiol oxidoreductases are present in all living organisms (for which genome sequences are available) and generally account for 0.5%–1% of the proteome, establishing thioredoxomes as essential and unexpectedly abundant enzymatic systems in cells. We observed that host-associated organisms have smaller thioredoxomes, but surprisingly, we did not find strong influence of environmental factors. However, environmental factors

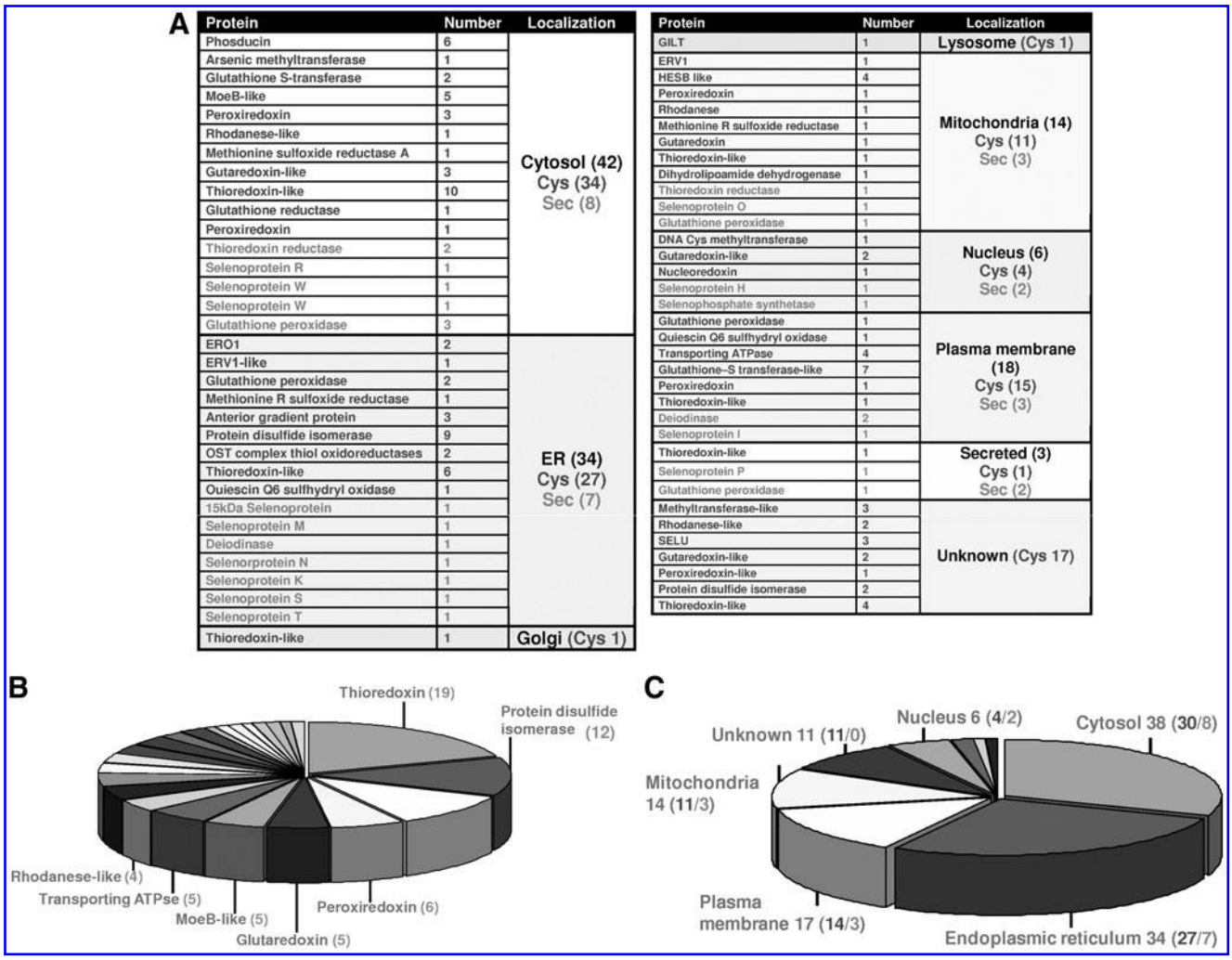

FIG. 7. The human thioredoxome. **(A)** The human thioredoxome consists of 136 proteins, including 111 proteins with catalytic Cys residues and 25 selenoproteins. **(B)** Occurrence of thiol oxidoreductase families in the human thioredoxome. **(C)** Subcellular localization of human thiol oxidoreductases.

could regulate gene expression of thiol oxidoreductases and thus influence their function. An interesting observation was that the size of thioredoxomes correlated with the proteome size. Such correlation was observed in all branches of life, and it was especially pronounced in prokaryotes.

Many thiol oxidoreductases are strategically located in between electron donors (*e.g.*, NADPH) and acceptors (*e.g.*, hydrogen peroxide, oxygen, reactive nitrogen species) and establish a flow of reducing equivalents that is linked to diverse cellular processes, such as antioxidant defense, repair of oxidative protein modifications, protein folding, DNA synthesis and repair, and other processes. A correlation of the number of proteins in a system with the proteome size is not unique for thiol oxidoreductases. Similar trends were observed for signaling proteins, such as kinases, phosphatases, and transcription factors (1, 23, 26). It appears that the extent of the use of such proteins supports organismal complexity. The role of thiol oxidoreductases in regulation and signaling is an interesting, yet poorly studied area of redox biology. Known signaling functions are limited to a group of thioredoxin-fold proteins, such as thioredoxins, glutaredoxins, and thiol peroxidases. The observed correlation between proteomes and thioredoxomes suggests a general role of thioredoxomes in biology, which provide the backbone of thiol interaction and redox regulation in the cell.

The sets of thiol oxidoreductases also illustrate the complexity of thiol-based redox control in cells and organisms and provide tools for the analyses of these processes. We believe it would be important to further examine the combined function of thiol oxidoreductases (*i.e.*, the thioredoxome). In this regard, this research area could benefit from the previous studies that characterized systems, such as kinomes and phosphoproteomes, in cells. Currently, identification of redox-regulated Cys residues in proteins is difficult and there are no known sequence- or structure-based motifs that could describe these residues. Therefore, an ability to efficiently identify proteins that act on these redox-regulated Cys should be of great value, and these proteins could be used as tools for comparative genomic analyses of thiol-based redox regulation. From our work, it is already clear that since thiol oxidoreductases are present in all living organisms, the thiol-based redox control is a widespread, basic, essential process in all life forms.

Materials and Methods

Thiol oxidoreductases were identified by three independent methods. First, these proteins were found through sequence similarity to selenoproteins and their catalytic redox-active Cys were defined as the amino acid corresponding in sequence alignments to Sec. In this method, Cys-containing proteins from NCBI non-redundant database and completely sequenced genomes were searched against a local non-redundant dataset of known selenoproteins and their Cys-containing homologs, which was separately compiled based on our previous research, literature searches, and additional homology analyses that identified distant homologs of known selenoproteins. A standalone TBLASTN program was used with the expectation value of 1e-3. Proteins with Cys corresponding to Sec in multiple alignments were considered as thiol oxidoreductases.

Second, a collection of protein domains that are fused to known thiol oxidoreductases was generated using STRING (http://string.embl.de/, accessed July 7, 2011) NCBI CDART, and standalone RPS BLAST from NCBI BLAST package with the expectation value 1e-5. Domains fused to known thiol oxidoreductases were analyzed for occurrence of conserved Cys using a standalone BLAST program with the expectation value 1e-5. Domains fused with thiol oxidoreductases and containing at least one conserved Cys were considered as candidate thiol oxidoreductases.

Third, all predicted ORFs from NCBI were searched against a database of known thiol oxidoreductases using NCBI BLAST program with the expectation value of 1e-3 to identify distant homologs of thiol oxidoreductases. These analyses were supplemented with exhaustive PSI BLAST searches starting involving known thiol oxidoreductases.

Thiol oxidoreductases identified in these searches were analyzed separately and were also used to create a combined database of thiol oxidoreductases. In addition, associated web tools were prepared, including the Sec/Cys pair method, a tool to search for homologs of known thiol oxidoreductases, a tool to search for thiol oxidoreductase domain fusion events, and a method combining all three search strategies. These search tools are located at two mirror sites: http://genomics.unl.edu/ THIOREDOXOME and http://gladyshevlab.org/THIOREDOXOME. Glycolytic enzymes in completely sequenced bacteria were identified using *E. coli* glycolytic enzymes as seeds. NCBI BLAST standalone program with the expectation value 1e-5 was used for automatic genome parsing.

Acknowledgments

This research was completed in part utilizing the Prairie-Fire Beowulf cluster from Research Computing Facility of the University of Nebraska–Lincoln. Supported by NIH Grant GM065204.

Author Disclosure Statement

The authors declare no competing interests.

References

1. Anantharaman V, Iyer LM, and Aravind L. Comparative genomics of protists: New insights into the evolution of eukaryotic signal transduction and gene regulation. *Annu Rev Microbiol* 61: 453–475, 2007.
2. Andres-Mateos E, Perier C, Zhang L, Blanchard-Fillion B, Greco TM, Thomas B, Ko HS, Sasaki M, Ischiropoulos H, Przedborski S, Dawson TM, and Dawson VL. DJ-1 gene deletion reveals that DJ-1 is an atypical peroxiredoxin-like peroxidase. *Proc Natl Acad Sci USA* 104: 14807–14812, 2007.
3. Calabrese V, Sultana R, Scapagnini G, Guagliano E, Sapienza M, Bella R, Kanski J, Pennisi G, Mancuso C, Stella AM, and Butterfield DA. Nitrosative stress, cellular stress response, and thiol homeostasis in patients with Alzheimer's disease. *Antioxid Redox Signal* 8: 1975–1986, 2006.
4. Fassio A, and Sitia R. Formation, isomerisation and reduction of disulphide bonds during protein quality control in the endoplasmic reticulum. *Histochem Cell Biol* 117: 151–157, 2002.
5. Fetrow JS, Godzik A, and Skolnick J. Functional analysis of the *Escherichia coli* genome using the sequence-to-structure-to-function paradigm: Identification of proteins exhibiting the glutaredoxin/thioredoxin disulfide oxidoreductase activity. *J Mol Biol* 282: 703–711, 1998.
6. Fetrow JS, Siew N, Di Gennaro JA, Martinez-Yamout M, Dyson HJ, and Skolnick J. Genomic-scale comparison of sequence- and structure-based methods of function prediction: Does structure provide additional insight? *Protein Sci* 10: 1005–1014, 2001.
7. Fetrow JS and Skolnick J. Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and T1 ribonucleases. *J Mol Biol* 281: 949–968, 1998.
8. Finkel T. Oxidant signals and oxidative stress. *Curr Opin Cell Biol* 15: 247–254, 1996.
9. Fomenko DE and Gladyshev VN. CxxS: Fold-independent redox motif revealed by genome-wide searches for thiol/disulfide oxidoreductase function. *Protein Sci* 11: 2285–2296, 2002.
10. Fomenko DE and Gladyshev VN. Identity and functions of CxxC-derived motifs. *Biochemistry* 42: 11214–11225, 2003.
11. Fomenko DE, Marino SM, and Gladyshev VN. Functional diversity of cysteine residues in proteins and unique features of catalytic redox-active cysteines in thiol oxidoreductases. *Mol Cells* 26: 228–235, 2008.
12. Fomenko DE, Xing W, Adair BM, Thomas DJ, and Gladyshev VN. High-throughput identification of catalytic redox-active cysteine residues. *Science* 315: 387–389, 2007.
13. Gil-del Valle L, de la C Milian L, Toledo A, Vilaró N, Tápanes R, and Otero MA. Altered redox status in patients with diabetes mellitus type I. *Pharmacol Res* 51: 375–380, 2005.
14. Holmgren A. Antioxidant function of thioredoxin and glutaredoxin systems. *Antioxid Redox Signal* 2: 811–820, 2000.
15. Holmgren A, Johansson C, Berndt C, Lönn ME, Hudemann C, and Lillig CH. Thiol redox control via thioredoxin and glutaredoxin systems. *Biochem Soc Trans* 33: 1375–1377, 2005.
16. Kadokura H and Beckwith J. Mechanisms of oxidative protein folding in the bacterial cell envelope. *Antioxid Redox Signal* 13: 1231–1246, 2010.
17. Kim HY and Gladyshev VN. Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol* 3: e375, 2005.
18. Kim HY, Fomenko DE, Yoon YE, and Gladyshev VN. Catalytic advantages rrovided by selenocysteine in methionine-S-sulfoxide reductases. *Biochemistry* 45: 13697–13704, 2006.

19. Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigo R, and Gladyshev VN. Characterization of mammalian selenoproteomes. *Science* 300: 1439–1443, 2003.
20. Lee BC, Dikiy A, Kim HY, and Gladyshev VN. Functions and evolution of selenoprotein methionine sulfoxide reductases. *Biochim Biophys Acta* 1790: 1471–1477, 2009.
21. Martin JL. Thioredoxin—Fold for all reasons. *Structure* 3: 245–250, 1995.
22. Matés JM, Segura JA, Alonso FJ, and Márquez J. Intracellular redox status and oxidative stress: Implications for cell proliferation, apoptosis, and carcinogenesis. *Arch Toxicol* 82: 273–299, 2008.
23. Manning G, Plowman GD, Hunter T, and Sudarsanam S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 27: 514–520, 2002.
24. Newman SF, Sultana R, Perluigi M, Coccia R, Cai J, Pierce WM, Klein JB, Turner DM, and Butterfield DA. An increase in S-glutathionylated proteins in the Alzheimer's disease inferior parietal lobule, a proteomics approach. *J Neurosci Res* 85: 1506–1514, 2007.
25. Riemer J, Bulleid N, and Herrmann JM. Disulfide formation in the ER and mitochondria: Two solutions to a common process. *Science* 324: 1284–1287, 2009.
26. Templeton TJ, Iyer LM, Anantharaman V, Enomoto S, and Abrahante JE. Comparative analysis of Apicomplexa and genomic diversity in eukaryotes. *Genome Res* 14: 1686–1695, 2004.
27. Wassef R, Haenold R, Hansel A, Brot N, Heinemann SH, and Hoshi T. Methionine sulfoxide reductase A and a dietary supplement S-methyl-L-cysteine prevent Parkinson's-like symptoms. *J Neurosci* 27: 12808–12816, 2007.

Address correspondence to:
Prof. Dmitri E. Fomenko
Department of Biochemistry and Redox Biology Center
University of Nebraska-Lincoln
Lincoln, NE 68588

E-mail: dfomenko@genomics.unl.edu

Prof. Vadim N. Gladyshev
Division of Genetics, Brigham and Women's Hospital and Harvard Medical School
77 Avenue Louis Pasteur
Boston, MA 02115

E-mail: vgladyshev@rics.bwh.harvard.edu

Date of first submission to ARS Central, March 9, 2011; date of final revised submission, August 13, 2011; date of acceptance, September 1, 2011.

Abbreviations Used

BLAST = basic local alignment search tool
Cys = cysteine
ER = endoplasmic reticulum
Grx = glutaredoxin
NADPH = nicotinamide adenine dinucleotide phosphate
ORF = open reading frame
Sec = selenocysteine
Trx = thioredoxin
TR = thioredoxin reductase

This article has been cited by:

1. Vyacheslav M. Labunskyy , Vadim N. Gladyshev . Role of Reactive Oxygen Species-Mediated Signaling in Aging. *Antioxidants & Redox Signaling*, ahead of print. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]