Identity and Functions of CxxC-Derived Motifs†

Dmitri E. Fomenko and Vadim N. Gladyshev*

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

Received March 21, 2003; Revised Manuscript Received June 17, 2003

ABSTRACT: Two cysteines separated by two other residues (the CxxC motif) are employed by many redox proteins for formation, isomerization, and reduction of disulfide bonds and for other redox functions. The place of the C-terminal cysteine in this motif may be occupied by serine (the CxxS motif), modifying the functional repertoire of redox proteins. Here we found that the CxxC motif may also give rise to a motif, in which the C-terminal cysteine is replaced with threonine (the CxxT motif). Moreover, in contrast to a view that the N-terminal cysteine in the CxxC motif always serves as a nucleophilic attacking group, this residue could also be replaced with threonine (the TxxC motif), serine (the SxxC motif), or other residues. In each of these CxxC-derived motifs, the presence of a downstream α-helix was strongly favored. A search for conserved CxxC-derived motif/helix patterns in four complete genomes representing bacteria, archaea, and eukaryotes identified known redox proteins and suggested possible redox functions for several additional proteins. Catalytic sites in peroxiredoxins were major representatives of the TxxC motif, whereas those in glutathione peroxidases represented the CxxT motif. Structural assessments indicated that threonines in these enzymes could stabilize catalytic thiolates, suggesting revisions to previously proposed catalytic triads. Each of the CxxC-derived motifs was also observed in natural selenium-containing proteins, in which selenocysteine was present in place of a catalytic cysteine.

Cysteine (Cys) is one of the least abundant amino acid residues in proteins, but it is frequently found in catalytic sites of enzymes. In particular, this residue is employed for redox function through reversible oxidation to disulfide or sulfenic acid. One of the best studied redox motifs involving cysteines is the CxxC motif, in which two cysteines are separated by two other residues (*1*, *2*). This motif is employed by thioredoxins and glutaredoxins for reduction of intermolecular and intramolecular disulfide bonds and other forms of oxidized cysteines, by protein disulfide isomerases and DsbA for disulfide bond formation, and by DsbC and other redox proteins for disulfide bond isomerization (*3*). The CxxC motif has been called a "rheostat at the active center", because changes in residues that separate the two cysteines dramatically influence redox potentials and pK_a values of cysteines, configuring proteins for a particular redox function (*4*, *5*).

Characterization of CxxC motifs in thioredoxins, glutaredoxins, DsbA, and other enzymes revealed that the N-terminal cysteine has a decreased pK_a and serves as a nucleophilic attacking group in redox reactions catalyzed by CxxC-containing redox enzymes (*6*). The attack by this cysteine on an oxidized cysteine in a substrate (e.g., sulfenic acid, disulfide, etc.) results in the formation of an intermolecular disulfide bond (*2*, *7*). The subsequent attack by the C-terminal cysteine in the CxxC motif on the intermolecular disulfide releases the reduced product and forms a disulfide in the redox enzyme.

This well-studied mechanism was employed to trap redox intermediates and identify targets of CxxC-containing proteins. Indeed, mutation of the C-terminal Cys in the CxxC motif allows generation of an intermolecular disulfide but prevents completion of the redox reaction, thus trapping the intermolecular intermediate (*8*, *9*). Subsequent isolation of the disulfide-bonded complex may be used to identify redox targets for thioredoxins and other redox proteins.

Recent studies revealed the presence of many natural homologues of CxxC-containing proteins, in which the C-terminal Cys in the CxxC motif is replaced with serine (the CxxS motif) (*10*). Some of these enzymes were biochemically characterized revealing an expanded repertoire of redox functions, such as the roles of CxxS-containing proteins in methionine sulfoxide reduction (*11*) and protein retention through transient formation of intermolecular disulfide bonds (*12*). Genomic analyses determined that CxxS motifs are highly conserved and present in structurally distinct proteins, suggesting that CxxS is a new fold-independent redox motif (*10*).

Interestingly, further homology and threading analyses of CxxC-containing proteins revealed proteins, in which the N-terminal Cys in the CxxC motif was replaced with other residues (*13*). However, possible functions of these proteins were not clear as the lack of the attacking cysteine was thought to disrupt redox functions in these proteins (*13*).

In this work, we identified new CxxC-derived redox motifs, in which N-terminal or C-terminal CxxC cysteines were replaced with serine and threonine. Further genomic analyses revealed major representatives of these redox motifs and proposed redox functions for several additional proteins. Finally, molecular modeling experiments suggested revisions

† This work was supported by NIH Grants GM061603 and RR017675.

* Corresponding author. Tel. (402) 472-4948. Fax (402) 472-7842. E-mail: vgladyshev1@unl.edu.

10.1021/bi034459s CCC: $25.00
Published on Web 08/30/2003

Table 1: Identification of Proteins Containing CxxS and SxxC Motifs in *Campylobacter jejuni, Methanococcus jannaschii, Escherichia coli,* and *Saccharomyces cerevisiae* Genomes

organism	total proteins	proteins with CxxS and SxxC		proteins with conserved CxxS and SxxC		proteins with CxxS and SxxC, for which conservation profile is not available		proteins with conserved CxxS and SxxC flanked by an α-helix		redox proteins (true positives/ candidates)	
		CxxS	SxxC	CxxS	SxxC	CxxS	SxxC	CxxS	SxxC	CxxS	SxxC
Escherichia coli	4289	707	694	62	68	71	67	33	37	10 (6/4)	7(2/5)
Campylobacter jejuni	1634	307	295	17	19	32	28	13	13	2 (0/2)	0(0/0)
Methanococcus jannaschii	1770	264	209	29	20	41	42	23	10	1 (1/0)	0(0/0)
Saccharomyces cerevisiae	6334	1946	2050	96	100	498	532	54	58	18 (9/9)	4(3/1)

to catalytic triads in glutathione peroxidases and peroxiredoxins.

MATERIALS AND METHODS

Collections of protein sequences derived from completely sequenced genomes of *Saccharomyces cerevisiae* (February 2002 release), *Campylobacter jejuni* (June 2001 release), *Methanococcus jannaschii* (April 2001 release), and *Escherichia coli* (April 2001 release) were obtained from NCBI. Bovine glutathione peroxidase 1 (GPx1) (gi 229946, PDB 1GP1) and human peroxiredoxin 5 (gi 15826629, PDB 1HD2) structures were obtained from PDB.

Perl scripts were developed for identification of candidate redox proteins containing CxxS, SxxC, CxxT, and TxxC motifs. The search algorithm utilized the following strategy, applied separately to each of the analyzed genomes: (i) the identification of all proteins containing CxxS, SxxC, CxxT, or TxxC; (ii) filtering out proteins that employ cysteines in these motifs for metal binding; (iii) analysis of conservation profiles for cysteines, serines, and threonines in CxxS, SxxC, CxxT, and TxxC motifs; and (iv) secondary structure prediction for the remaining proteins containing conserved CxxS, SxxC, CxxT, and TxxC sequences. Perl scripts were used in steps 1 and 2 for identification of CxxS-, SxxC-, CxxT-, or TxxC-containing and metal-binding proteins, in steps 3 and 4 for interaction between external programs by input/output formatting, and for final output visualization.

Metal-binding proteins were filtered out by identifying available Prosite patterns for metal binding sites (*14*). Conservation of CxxS, SxxC, CxxT, or TxxC sequences was assessed using a position-specific iterated (PSI) BLAST output (*15*). Parameters of the conservation analysis were optimized using a set of 947 thioredoxins and glutaredoxins, 500 peroxiredoxins, and 250 glutathione peroxidases that were automatically extracted from a nonredundant (NR) database at NCBI with the BLAST and confirmed by manual inspection. The final set of parameters was as follows: expectation value, 0.0001; optimal number of iterations, 3; and conservation cutoff limit, 65% identity in cysteines, serines, and threonines in CxxS, SxxC, CxxT, or TxxC motifs among homologues found in the NR database. In addition, the following replacements among homologues were allowed: Ser to Cys, Ser to Thr, Thr to Cys, Thr to Ser. Only proteins satisfying these criteria were considered further.

For the secondary structure prediction, PSIPRED was used (*16*). Accuracy of the PSIPRED prediction was estimated to be ~78−82% in PSI BLAST outputs. The database composed of thioredoxins, glutaredoxins, peroxiredoxins, and glutathione peroxidases was used for optimization of secondary structure analysis and the following sets of parameters were chosen: β-strand, 3 or more amino acid residues between −1 and −10 residues; and α-helix, 7 or more amino acid residues between +1 and +20 residues (with respect to an N-terminal amino acid in CxxS, SxxC, CxxT or TxxC motifs).

Proteins that contained a conserved redox motif in the context of a β-C/S/TxxC/S/T-α secondary structure or contained an α-helix downstream of the redox motif were manually analyzed for sequence homology to proteins with known function with PSI−BLAST, for solvent accessibility of a redox motif by inspecting three-dimensional structures, for identity of protein domains with CDD, Pfam, and clusters of orthologous groups (COG), and for gene context by COG.

An example of the search is illustrated below using the analysis of the *S. cerevisiae* genome for proteins containing candidate CxxS redox motifs (Table 1). Among 6334 proteins predicted in the yeast genome, 1946 proteins contained one or more CxxS sequences, and only 96 proteins contained one or more conserved CxxS sequences (that is, N-terminal and C-terminal amino acids (Cys, Thr, and Ser) were conserved among close homologues). To determine conservation of these residues, we analyzed the PSI−BLAST output file. Analysis of secondary structure revealed that 54 of the 96 proteins had a CxxS motif that was flanked by an α-helix. The final step in our search strategy included a manual analysis of our automatic output containing the 54 sequences. The main contamination appeared to be metal-binding proteins that were not filtered out using Prosite patterns in the previous steps. For manual filtration, we used homology analyses against known metal-binding proteins using PSI−BLAST, solvent accessibility of redox motifs by inspecting three-dimensional structures (only for proteins with known structures and their homologues), identity of protein domains with CDD, Pfam, and COG, and gene context by COG. This last step filtered out 36 out of the 54 proteins since they employed cysteines for metal coordination or had known structural disulfide bonds. The final set had 18 CxxS-containing proteins, including 9 that were "true positives" (known thiol/disulfide oxidoreductases).

Alignments for thioredoxins, glutaredoxins, peroxiredoxins, and glutathione peroxidases were performed using PSI−BLAST with an expectation value of 10 using representatives from a database of thioredoxins, glutaredoxins, peroxiredoxins, and glutathione peroxidases. Separately, ClustalW alignments were generated.

To model glutathione peroxidase and peroxiredoxin structures, SYBIL 6.6 was used. The PDB glutathione peroxidase structure contained the selenocysteine residue in the protonated selenenic form, which appeared to preclude formation of a hydrogen bond between the hydroxyl group of threonine

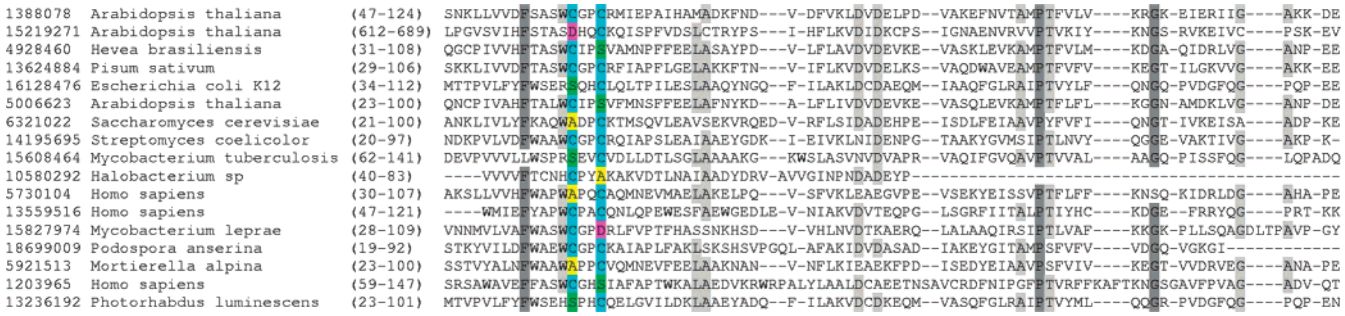

FIGURE 1: Alignment of thioredoxin homologues. Randomly chosen proteins with significant similarity to thioredoxin and that either conserve CxxC or replace one of the Cys with other residues are shown. CxxC-derived cysteines are shown in blue, serines in green, alanines in yellow, and aspartates in purple. Residues conserved in more than 90% sequences are highlighted with dark gray and residues conserved in more than 60% sequences are highlighted with light gray. GI numbers for each sequence used in the alignment are indicated on the left.

and selenocysteine (Sec) (selenolate in the native protein). Replacement of the selenenic group with selenolate allowed formation of the hydrogen bond between Sec and Thr of the CxxT motif without disrupting other structural interactions in the active site. The PDB human peroxiredoxin 5 structure contained a catalytic cysteine in the reduced state and was used directly to access hydrogen bond formation with threonine and previously proposed catalytic residues.

Yeast microarray gene expression analyses were performed on gene clusters obtained by the fuzzy k-means clustering method as described by Gasch and Eisen (*30*). This method allowed simultaneous participation of each gene in several clusters. NifU-like Isu2 gene was a member of cluster #4, which primarily included targets of the Yap1 transcription factor. For visualization of the data, FuzzyExplorer program was used via http://rana.lbl.gov/FuzzyK.

RESULTS

Presence of Serine and Threonine in the C-terminal Position of the CxxC Motif in Thioredoxin-Fold Proteins. Thioredoxin-fold proteins (thioredoxins, glutaredoxins, protein disulfide isomerases, and nucleoredoxins) are major representatives of the CxxC redox motif. In some of these proteins, the C-terminal CxxC cysteine is replaced with serine (*10*). Through homology searches of the NR database, we identified over 900 proteins exhibiting sequence homology to thioredoxin and containing CxxC and CxxC-derived motifs. Although the majority of these proteins possessed a CxxC motif, a number of proteins were identified that replaced the C-terminal CxxC Cys (as well as N-terminal Cys, see below) with serine, and additional proteins were also identified that had valine, alanine, and several other residues in these positions (Figure 1). Thus, the C-terminal position in the CxxC motif in thioredoxin homologues tolerates several residues.

We further used PSI−BLAST-based screens to analyze sequences of thioredoxin-fold proteins that exhibit distant sequence similarity to thioredoxins, including glutathione-*S*-transferases (GSTs), glutathione peroxidases (GPxs) (Figure 2A), and peroxiredoxins (Figure 2B). Interestingly, these analyses revealed that GPxs possessed a conserved threonine residue that aligned with the C-terminal CxxC cysteine in thioredoxins (Figures 2A and 3). Identification of natural serine-for-cysteine replacements in glutaredoxin homologues (*10*) and threonine-for-cysteine replacements in glutathione peroxidases (Figure 3) were of particular interest. Structurally, serine and threonine are highly similar to Cys: cysteine and serine differ by a single atom (sulfur versus oxygen), whereas a side chain of threonine has an additional methyl group. Since the thiol of the C-terminal CxxC cysteine in thioredoxins is known to share the proton with the N-terminal CxxC thiolate (*17*), and since serine and threonine at this position could also form a hydrogen bond with the N-terminal CxxC thiolate, natural substitutions of Cys with Ser or Thr might be consistent with the initial step (attack of the N-terminal CxxC thiolate on an oxidized form of sulfur to form intermolecular disulfide) in the catalytic mechanism of thioredoxin and its homologues.

Occurrence of Thioredoxin Homologues that Lack the N-terminal Cysteine in the CxxC Motif. PSI−BLAST homology analyses also identified additional thioredoxin-fold proteins in which the N-terminal Cys in the CxxC motif was replaced with serine (e.g., gi 15604262, 15892467, and 6320303) (Figures 1 and 3). While substitution of the C-terminal CxxC cysteine with Ser is employed in various redox proteins, the lack of the N-terminal Cys is thought to contradict the redox function of thioredoxin-fold proteins (*13*). Interestingly, those proteins that lacked the N-terminal Cys still conserved the C-terminal Cys, consistent with a possible redox function for the remaining Cys. Moreover, the presence of Ser in place of the N-terminal Cys suggested that these residues could potentially form a hydrogen bond with the thiolate, if the latter is formed on the C-terminal CxxC Cys.

Identification of CxxC-Derived Motifs on a Genomic Scale. Initial analysis of proteins, which exhibited a significant sequence similarity to thioredoxin and lacked the N-terminal CxxC Cys, revealed no examples of functionally characterized proteins. To identify proteins in which either first or second Cys was replaced with Ser or Thr, we searched several completely sequenced genomes for such CxxC-derived redox motifs. Specifically, complete sets of predicted proteins in representatives of the three domains of life, *S. cerevisiae* (eukaryotes), *M. jannaschii* (archaea), *C. jejuni*, and *E. coli* (bacteria), were analyzed to identify proteins containing conserved CxxS, SxxC, CxxT, and TxxC motifs. Secondary structures flanking the motifs were also analyzed. All searches were performed using Perl scripts developed as described in Materials and Methods.

Statistics of the searches are shown in Tables 1 and 2. The number of identified proteins was relatively low for any of the four motifs. Most true positives (defined as homo-

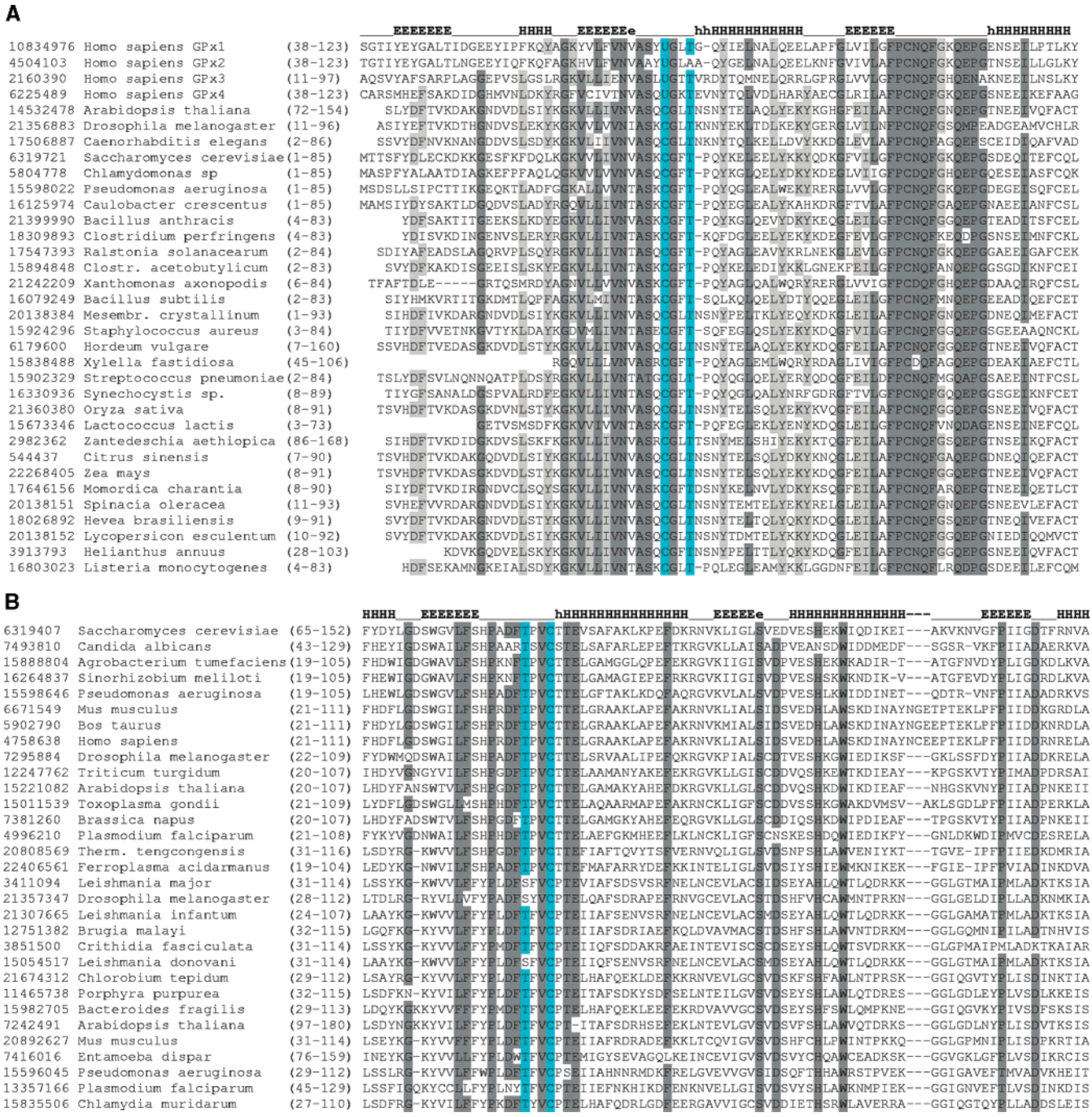

FIGURE 2: Alignment of glutathione peroxidases and peroxiredoxins. (A) Glutathione peroxidases. (B) Peroxiredoxins. CxxC-derived cysteines (or selenocysteines, U) and threonines are shown in blue. Secondary structure consensus is indicated above the sequences. α-Helices predicted in more than 90% sequences are shown as "H", α-helices predicted in more than 60% sequences are shown as "h", *β* structures predicted in more than 90% sequences are shown as "E", and *β* structures predicted in more than 60% are shown sequences as "e". Predicted coil structures are shown as "−". Residues conserved in more than 90% sequences are highlighted with dark gray and residues conserved in more than 70% sequences are highlighted with light gray. GI numbers for each sequence used in the alignment are indicated on the left.

logues of functionally characterized redox proteins containing CxxC motifs) had an α-helix immediately following the redox motifs, consistent with the previous observation for the CxxS motif (*10*). Tables 3−6 list proteins in the four analyzed genomes that contained conserved CxxS, SxxC, CxxT, and TxxC motifs immediately upstream of an α-helix. The overall protein set consists of an approximately equal number of true positives and other proteins. The large number of homologues of redox proteins among the hits produced by our algorithm suggests that at least some of the candidate proteins may also be involved in redox processes through their CxxC-derived motifs or be homologues of redox proteins. However, further experimental analyses are necessary to determine their participation in redox processes and examine relevant redox functions.

Analyses of true positives also revealed major representatives of the four CxxC-derived motifs. Consistent with the previous data, many glutaredoxin homologues employed the CxxS motif, including five present in the *S. cerevisiae* genome (*10*). Only three of these were previously experi-

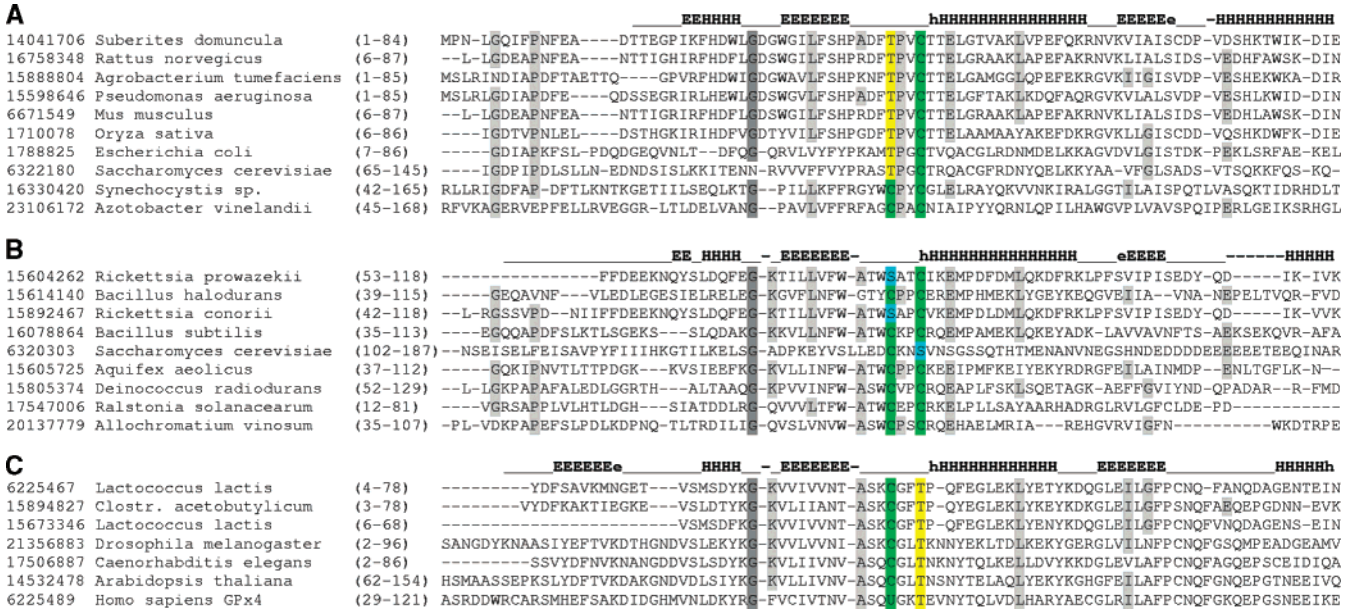

FIGURE 3: Alignment of peroxiredoxins, thioredoxins, and glutaredoxins and glutathione peroxidases. A single alignment was generated with PSI−BLAST and divided into three groups to show sequences that flank CxxC and CxxC-derived motifs in peroxiredoxins (A), thioredoxins and glutaredoxins (B), and glutathione peroxidases (C). CxxC-derived cysteines are shown in green, serines are in blue, and threonines are in yellow. Secondary structure consensus is shown above the sequences separately for each group of sequences. α-Helices predicted in more than 90% sequences are shown as "H", α-helices predicted in more than 60% sequences are shown as "h", β structures predicted in more than 90% sequences are shown as "E", and β structures predicted in more than 60% sequences are shown as "e". Predicted coil structures are shown as "−". Residues conserved in more than 90% sequences are highlighted with dark gray and residues conserved in more then 50% sequences with light gray. GI numbers for each sequence used in the alignment are indicated on the left.

mentally characterized. GPxs were major representatives of the CxxT motif, whereas the TxxC motif was primarily represented by peroxiredoxins.

Functional TxxC Motif in Peroxiredoxins. Peroxiredoxins (Figure 2B) are thiol-dependent peroxidases containing the thioredoxin fold (*18*−*20*). Our finding of the TxxC motif in these proteins (Figure 2B) was unexpected because previous sequence and structure comparisons between peroxiredoxins and other thioredoxin-fold proteins suggested that the catalytic Cys in peroxiredoxins corresponds to the N-terminal CxxC Cys in thioredoxins and glutaredoxins, as well as to the catalytic Cys in GPxs (*18*). To clarify this issue, we reanalyzed multiple sequence alignments of thioredoxin-fold proteins and peroxiredoxins using PSI−BLAST and found that, consistent with our initial observations, the Cys in the TxxC motif in peroxiredoxins aligned with the C-terminal cysteine, and that the threonine corresponded to the N-terminal CxxC Cys (Figure 3). Moreover, we identified a new protein group in bacteria (last two sequences in Figure 3A) that were most homologous to peroxiredoxins yet had the CxxC motif that aligned with the TxxC motif in peroxiredoxins. This family has not been previously described nor functionally characterized. Thus, peroxiredoxins emerged as functionally characterized redox proteins (thiol-dependent peroxidases) containing a CxxC-derived redox motif, which lacked the N-terminal Cys, but employed the C-terminal CxxC Cys for redox function.

This catalytic cysteine in peroxiredoxins is known to be highly reactive and has a low pK_a (*19*). The thiolate is thought to be stabilized by a positively charged environment of the catalytic site. In particular, arginine (e.g., Arg127 in human peroxiredoxin 5) and histidine residues were proposed to stabilize the thiolate (*21*). However, our finding that the TxxC motif in peroxiredoxins is derived from the thioredoxin-fold CxxC motif suggested that the threonine may form a hydrogen bond with the catalytic thiolate, or alternatively both threonine and arginine could be involved. To test if the threonine could indeed interact with the thiolate, we analyzed a previously solved structure of human peroxiredoxin 5 (PDB number 1HD2, gi 15826629) (*21*). The Cys47 (S)−Thr44 (O) distance was 3.0 Å, whereas the distance between Cys47 (S) and Arg127 (N) was 3.76 Å. No other groups were identified in the vicinity of Cys47 that could form a hydrogen bond with the thiolate. Thus, the threonine in the TxxC motif appears to be a group that could form a hydrogen bond with the thiolate and its function might be assisted by arginine.

Functional CxxT Motif in Glutathione Peroxidases. PSI−BLAST was also sufficient to align the CxxT motif in glutathione peroxidases with the CxxC motif in thioredoxins (Figure 3). Thus, the conserved threonine corresponded to the C-terminal Cys in the CxxC motif. The role of this conserved Thr has not been previously addressed. The catalytic triad of GPxs is thought to consist of cysteine (or selenocysteine in some eukaryotic GPxs), glutamine, and arginine (e.g., Sec45, Trp158 and Gln80 in bovine GPx1, PDB number 1GP1), with the latter two residues directly involved in stabilization of the catalytic selenolate. Instead, our data suggested that the threonine might form a hydrogen bond with thiolate or selenolate. Although structures of two mammalian glutathione peroxidases are known (cytosolic GPx1 and plasma GPx3) (*22*, *23*), both proteins were crystallized in nonfunctional states that had the catalytic Sec in the form of selenenic acid. When we replaced selenenic acid in the GPx1 structure with selenolate, the latter could form a 3.19 Å hydrogen bond with Thr48 (O) without disrupting other interactions at the active site. In contrast, the distance between Sec45 (Se) and Trp158 (N) was

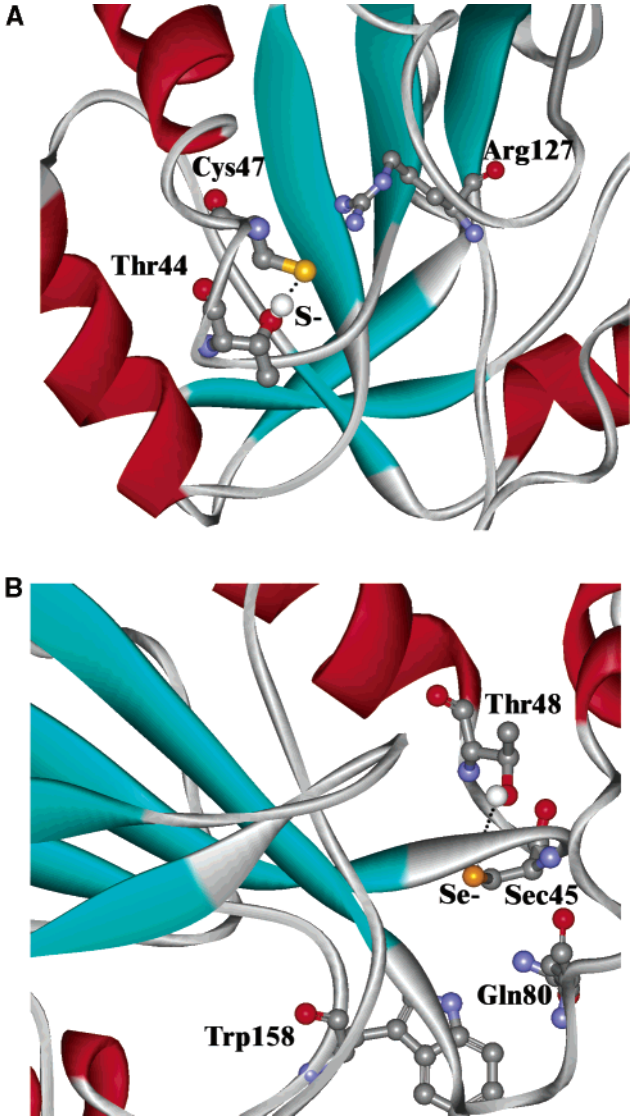

FIGURE 4: Molecular models of peroxiredoxins and glutathione peroxidases. (A) Human peroxiredoxin 5. The model illustrates stabilization of the reactive thiolate by formation of hydrogen bond between threonine-44 and ionized cysteine-47. The arginine previously proposed to stabilize the thiolate is also shown. (B) Bovine GPx1. The model illustrates stabilization of the reactive selenolate by formation of a hydrogen bond between ionized selenocysteine-45 and threonine-48. The tryptophan and glutamine previously proposed to stabilize the thiolate are also shown. α-Helices are shown in red, β structures are in blue, and coil structures are in gray.

6.23 Å, and between Sec45 (Se) and Gln80 (N) was 3.92 Å. Thus, the CxxC-derived threonine, rather than (or in addition to) residues previously proposed to constitute a catalytic triad, appeared to be the group that could interact with the selenolate in glutathione peroxidases.

Other True Positives Identified by Genomic Searches. Proteins containing the CxxS motif were previously described (*10*) and will not be further discussed in this article. Five true positives (homologues of known redox proteins) containing the SxxC motif were sequence homologues of thioredoxin, peroxiredoxin, and glutathione-*S*-transferase (GST) (Tables 3−6). In addition, all of these proteins exhibited a thioredoxin fold and the SxxC motif in these proteins occupied a position equivalent to the standard CxxC redox motif in thioredoxin-fold proteins. Of these five proteins, only peroxiredoxins (gi 6323138 and 6320661) were functionally characterized (in these proteins, SxxC motif replaced the TxxC motif that was found in most peroxiredoxins), but structural and functional interrelationships (*24*, *25*) suggest that the SxxC-containing thioredoxins and glutathione-*S*-transferases might also have a redox function. Possible electron donors and acceptors for these thioredoxin and GST homologues are not known. SxxC-containing GSTs form a separate cluster (e.g., cluster #0435 in COG), in which all family members conserve the SxxC motif except two yeast proteins that replace the serine with alanine. These proteins have a second conserved Cys in the C-terminal sequences, which might be involved in a disulfide bond with the SxxC cysteine. The genes for *E. coli* SxxC-containing GSTs are flanked by genes encoding membrane-bound proteins, which show a pattern of occurrence similar to that of SxxC-containing GSTs. A SxxC-containing thioredoxin is also a member of a separate cluster (COG #3118), which includes several additional SxxC- and CxxC-containing thioredoxin homologues.

NifU-like proteins, also called Isu1p (gi 6325122) and Isu2p (gi 6324800), contain the TxxC motif and are probably involved in the formation or repair of iron−sulfur clusters (*26*, *27*). Another NifU-like homologue, Nfu1p (gi 6322811), was identified by searching the yeast genome for the CxxS motif (Table 3). Specific functions of TxxC and CxxS motifs in NifU proteins are not known, but their location within the context of the secondary structure that corresponds to that found around CxxC-derived motifs suggests a possible redox function. The involvement of redox chemistry in iron−sulfur cluster assembly (*28*, *29*) is also consistent with this conclusion.

To gain further insight into possible redox functions of NifU-like proteins, we analyzed changes in their gene expression in response to environmental changes in *S. cerevisiae*. Exploration of 95 previously published yeast microarray data sets through fuzzy k-means clustering (*30*) revealed that Isu2 was a top member of the oxidative stress response cluster. Expression of genes in this cluster is regulated by transcription factor Yap1, and the Isu2 gene has a perfect Yap1 binding site. Other members of the cluster are primarily redox proteins, such as thioredoxin reductase Trr1, peroxiredoxin Ahp1, and glutathione peroxidase Gpx2. Two other NifU-like yeast proteins also appeared to have a redox component in their gene expression profiles as these were assigned to oxidative stress or general stress clusters (*30*). Although when considered alone, gene expression analyses are generally insufficient in predicting redox function, the combination of the redox motif and the strong response to redox stress favor a redox function.

Candidate Redox Proteins. Several proteins were identified that contained a conserved SxxC motif flanked by a downstream helix. Yeast aminopeptidase (gi 6324090) was first identified as a eukaryotic DNA-binding cysteine protease involved in inactivation of bleomycin. The enzyme is a tetramer of 48-kDa polypeptides (*31*). The SxxC motif is strictly conserved in this protein, and available structural information indicates that this motif is located on the protein surface. Putative amidase (gi 1787155) and its two homologues (hypothetical proteins, gi 1787968 and 1788299) are involved in peptidoglycan binding and cell wall degradation, whereas an uncharacterized protein (gi 1787437) has homol-

FIGURE 5: Predicted redox motifs in mammalian selenoproteins. Sequences flanking selenocysteine are shown for mouse proteins. The predicted redox motifs are shown in red. U is selenocysteine. Predicted secondary structures (H, α-helix; E, *β*-strand) are shown above the sequences.

Table 2: Identification of Proteins Containing CxxT and TxxC Motifs in *Campylobacter jejuni*, *Methanococcus jannaschii*, *Escherichia coli*, and *Saccharomyces cerevisiae* Genomes

organism	total proteins	proteins with CxxT and TxxC		proteins with conserved CxxT and TxxC		proteins with CxxT and TxxC, for which conservation profile is not available		proteins with conserved CxxT and TxxC flanked by an α-helix		redox proteins (true positives/ candidates)	
		CxxT	TxxC	CxxT	TxxC	CxxT	TxxC	CxxT	TxxC	CxxT	TxxC
Escherichia coli	4289	661	678	41	53	52	59	26	35	5(1/4)	11(4/7)
Campylobacter jejuni	1634	181	196	6	16	16	14	4	9	1(0/1)	5(2/3)
Methanococcus jannaschii	1770	212	191	20	19	32	29	12	4	1(0/1)	3 (1/2)
Saccharomyces cerevisiae	6334	1619	1494	60	53	342	310	36	19	5(3/2)	7 (5/2)

ogy to sporulation proteins. SxxC motifs are strictly conserved among bacterial homologues of these proteins, but functional information is largely lacking.

Among CxxT-containing proteins, yeast protein Vps53p (gi 6322431) is thought to be required for retrieval of late Golgi proteins and to form a stable complex with Vps52p and Vps54p. The Vps52p/Vps53p/Vps54p complex binds GTPase Ypt6p and Tlg1p protein (*32*). Though molecular function of Vps53p is not known, Vsp53p-containing complex is involved in fusion of endosome-derived vesicles. A second identified yeast protein is glycogen debranching enzyme (gi 6325442). This is a well-characterized protein involved in glycogen degradation and possessing two independent catalytic activities: glucosyltransferase and amylo-1,6-glucosidase (*33*, *34*).

E. coli (gi 1786212) and *C. jejuni* (gi 6968332) proteins containing the TxxC motif belong to a LytB family and are thought to be involved in the mevalonate-independent 2-C-

Table 3: CxxS-, SxxC-, CxxT-, and TxxC-Containing Proteins Identified in the *Saccharomyces cerevisiae* Genome

gi number	protein length[a]	C/S/TxxC/S/T position[b]	C/S/TxxC/S/T and flanking secondary structure[c]	description[d]
			CxxS	
6319488	**203**	**108**	**β-CPYS-α**	**glutaredoxin homologue**
6320193	**231**	**136**	**β-CSYS-α**	**glutaredoxin homologue**
6320303	**285**	**211**	**β-CGFS-α**	**glutaredoxin homologue**
6321022	**244**	**171**	**β-CGFS-α**	**glutaredoxin homologue**
6325198	**150**	**60**	**β-CGFS-α**	**glutaredoxin homologue**
6320726	**517**	**62**	**β-CLHS-α**	**protein thiol-disulfide isomerase homologue, ER-resident protein**
		405	**β-CIHS-α**	
6321642	**148**	**90**	**β-CMLS-α**	**required for arsenate resistance**
6325458	**130**	**76**	**β-CTGS-α**	**required for arsenate resistance**
6322811	**256**	**199**	**β-CSSS-α**	**NifU-like protein**
6320882	449	198	CRES-α	*S*-adenosyl-L-homocysteine hydrolase
6324041	373	181	α-CVAS-α	hypothetical ORF; Caf40p
6324588	757	583	α-CISS-α	B-type regulatory subunit of protein phosphatase 2A
6321792	472	3	CTVS-α	component of pheromone response pathway
6319497	1165	448	CIAS-α	required for chitin synthesis
6321005	801	565	α-CIVS-α	secretion (Golgi retention) deficient
6322745	717	405	β-CGTS-α	required for biosynthesis of cell wall
6324247	904	848	α-CRSS-α	required for mismatch repair in mitosis and meiosis
6325025	1887	1305	CATS-α	trifunctional enzyme
			SxxC	
6323138	**176**	**58**	**β-SPTC-α**	**peroxiredoxin**
6320661	**196**	**44**	**β-SFVC-α**	**peroxiredoxin**
6322929	**370**	**42**	**β-SLAC-α**	**glutathione *S*-transferase**
6324090	483	93	SGRC-α	aminopeptidase of cysteine protease family
			CxxT	
6319721	**162**	**36**	**β-CGFT-α**	**glutathione peroxidase Gpx2p**
		79	**α-TEFC**	
6322228	**163**	**35**	**β-CGFT-α**	**glutathione peroxidase**
6322826	**167**	**35**	**β-CAFT-α**	**glutathione peroxidase Gpx1p**
6322431	822	458	β-CATT-α	required for vacuolar protein sorting
6325442	1536	536	CHST-α	glycogen debranching enzyme
			TxxC	
6319407	**261**	**87**	**β-TPVC-α**	**peroxiredoxin**
6322180	**215**	**103**	**β-TPGC-α**	**peroxiredoxin**
6323613	**196**	**44**	**β-TFVC-α**	**peroxiredoxin**
6324800	**156**	**84**	**β-TFGC-α**	**nifU-like protein**
6325122	**165**	**92**	**β-TFGC-α**	**nifU-like protein**
6323067	83	23	TKHC-α	essential during assembly for full cytochrome C oxidase activity
6320240	370	340	β-TDAC-α	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase

[a] Number of amino acids in an ORF. [b] Amino acid position corresponding to the first amino acid in the C/S/TxxC/S/T sequence. [c] C/S/TxxC/S/T primary sequences and flanking secondary structure (α, α-helix; β, β-strand). [d] Predicted function or information obtained by database searches. Bold shows true positives.

methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis occurring in bacteria, plants, and parasites (*35*, *36*). A thiamin biosynthesis protein ThiC (gi 1790427) is involved in the formation of 4-amino-5-hydroxymethyl-2-methylpyrimidine, an intermediate in pyrimidine biosynthesis (*37*, *38*). In addition to the conserved CxxT motif, this protein contains a CxxC motif 90 amino acids downstream of CxxT. Additional *E. coli* proteins identified in our searches include a protein component (gi 1786601) of LIV−II membrane fraction that is associated with the branched chain amino acid transport system and belongs to a transportase family (*39*) and a protein (gi 1790450) homologous to homocysteine *S*-methyltransferase (vitamin B12-dependent) enzymes that may be involved in methionine biosynthesis (*40*). Finally, a single identified CxxT-containing archaeal protein (gi 1591525) has not been functionally characterized.

Among two identified yeast TxxC-containing proteins, cytochrome oxidase C subunit Vib (gi 6323067) is one of the 13 subunits in the cytochrome C oxidase complex, a terminal oxidase in bacterial and mitochondrial electron transport pathways. This protein may bind a heme (*41*, *42*). A second yeast protein, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (gi 6320240), catalyzes the first step in aromatic amino acid biosynthesis from chorismate by stereospecific condensation of phosphoenolpyruvate and D-erythrose-4-phosphate to form 3-deoxy-D-arabino-heptulosonate-7-phosphate. The protein is a tetramer consisting of two tight dimers (*43*). The TxxC motif in this protein is conserved and located on the protein surface. Three homologues of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase were also identified by our algorithm in the *E. coli* genome (gi 1787996, 1786969, and 1788953).

Among bacterial TxxC-containing proteins, three were carbon starvation proteins (CstA homologues, gi 1786814, 1786814, and 6968354) that are regulated by levels of carbon substrates. CstAs are hydrophobic (probably membrane associated) proteins that may be involved in peptide utilization (*44*). Glyceraldehyde 3-phosphate dehydrogenase (gi 1787686) is a glycolysis/gluconeogenesis enzyme that catalyzes reversible oxidation and phosphorylation of D-glyceraldehyde-3-phosphate to 1,3-diphospho-glycerate. The enzyme also plays a role in cytoskeleton formation (*45*, *46*). The glyceraldehyde-3-phosphate dehydrogenase TxxC motif is conserved and located on the protein surface, and the enzyme is known to be redox regulated (*47*−*49*). No functional information is available for the two remaining

Table 4: CxxS-, SxxC-, CxxT-, and TxxC-Containing Proteins Identified in the *Escherichia coli K12* Genome

gi number	protein length[a]	C/S/TxxC/S/T position[b]	C/S/TxxC/S/T and flanking secondary structure[c]	description[d]
				SxxC
1786700	**296**	**46**	***β*-SQHC-α**	**thioredoxin homologue**
1789489	**328**	**59**	***β*-SLAC-α**	**putative glutathione *S*-transferase**
1787155	615	524	α-SSGC-α	putative amidase
1787356	320	213	α-SSGC-α	hypothetical protein
1787437	510	99	*β*-SNPC-α	putative sporulation protein
1787968	334	205	SSGC-α	hypothetical protein
1788299	310	203	SQGC-α	hypothetical protein
				CxxS
1787943	**115**	**30**	***β*-CGFS-α**	**glutaredoxin homologue**
1787757	**143**	**125**	**CPVS-α**	**osmotically inducible protein**
1788841	**119**	**13**	***β*-CSKS-α**	**arsenate reductase**
1788991	**104**	**14**	***β*-CGTS-α**	**arsenate reductase**
1789918	**141**	**15**	***β*-CGTS-α**	**arsenate reductase**
1789804	**294**	**232**	***β*-CTCS-α**	**disulfide bond chaperones of the HSP33 family**
1789475	495	264	*β*-CGGS-α	altronate hydrolase
1789516	523	288	*β*-CGGS-α	putative altronate hydrolase
1787577	322	41	*β*-CGKS-α	putative ATP-binding component of a transport system
1788683	436	99	*β*-CATS-α	putative acyltransferase
				CxxT
1788003	**183**	**36**	***β*-CGLT-α**	**glutathione peroxidase**
1786212	316	196	CYAT-α	involved in penicillin tolerance
1786601	439	290	α-CLVT-a	branched chain amino acid transport system II carrier protein
1790450	1227	309	*β*-CCGT-α	B12-dependent homocysteine-N5-methyltetrahydrofolate transmethylase, repressor of metE and metF
		310	*β*-CGTT-α	
1790427	631	493	α-CYVT-α	thiamin biosynthesis
		580	CSMC-α	
				TxxC
1786822	**187**	**43**	***β*-TFVC-α**	**peroxiredoxin**
1787584	**168**	**57**	***β*-TGVC-α**	**peroxiredoxin**
1788825	**156**	**42**	***β*-TPGC-α**	**peroxiredoxin**
1788878	**128**	**59**	***β*-TYGC-α**	**NifU-like protein**
1786814	701	331	α-TIAC-α	carbon starvation protein
1786969	350	324	*β*-TDAC-α	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase
1787686	134	34	TTNC-α	glyceraldehyde 3-phosphate dehydrogenase C
1787996	348	323	*β*-TDPC-α	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase
1788953	356	326	*β*-TDAC-α	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase
1790242	275	214	*β*-TQAC-α	diaminopimelate epimerase
1790814	721	338	α-TIAC-α	carbon starvation protein

[a] Number of amino acids in an ORF. [b] Amino acid position corresponding to the first amino acid in the C/S/TxxC/S/T sequence. [c] C/S/TxxC/S/T primary sequences and flanking secondary structure (α, α-helix; *β*, *β*-strand). [d] Predicted function or information obtained by database searches. Bold shows true positives.

Table 5: CxxS-, CxxT-, SxxC-, and TxxC-Containing Proteins Identified in the *Campylobacter jejuni* Genome

gi number	protein length[a]	C/S/TxxC/S/T position[b]	C/S/TxxC/S/T and flanking secondary structure[c]	description[d]
			CxxS	
6967954	390	155	*β*-CGGS-α	altronate hydrolase
6968729	435	194	*β*-CHPS-α	hypothetical protein
			CxxT	
6968332	277	185	CDAT-α	lytB homologue
			TxxC	
6967808	**198**	**45**	***β*-TFVC-α**	**peroxiredoxin**
6968225	**175**	**55**	***β*-TPVC-α**	**peroxiredoxin**
6968354	703	332	α-TIAC-α	carbon starvation protein A homologue
6968950	249	188	*β*-TLAC-α	putative diaminopimelate epimerase
6968344	325	170	TRGC-α	putative periplasmic protein

[a] Number of amino acids in an ORF. [b] Amino acid position corresponding to first amino acid in the C/TxxC/S/T sequence. [c] C/TxxC/S/T primary sequences and flanking secondary structure (α, α-helix; *β*, *β*-strand). [d] Predicted function or information obtained by database searches. Bold shows true positives.

prokaryotic proteins (gi 6968344 and 1500567) containing the TxxC motif.

CxxC-Derived Motifs Are Represented in Selenocysteine-Containing Proteins. Previously characterized selenocysteine-containing proteins (mammalian GPxs, thioredoxin reductases, methionine-R-sulfoxide reductases, etc.) catalyze redox reactions using selenolate as the attacking group (*50*). In these enzymes, location of selenocysteine is indicative of a redox center. In addition, most selenoproteins have homologues in other organisms, which contain cysteine in place of selenocysteine (and which are less efficient catalytically than the corresponding selenoproteins). In these proteins, the

Table 6: CxxS-, CxxT-, SxxT-, and TxxC-Containing Proteins Identified in the *Methanococcus jannaschii* Genome

gi number	protein length[a]	C/S/TxxC/S/T position[b]	C/S/TxxC/S/T and flanking secondary structure[c]	description[d]
			CxxS	
1499579	**275**	**235**	***β*-CELS-α**	**conserved hypothetical protein DsrE-like protein**
			CxxT	
1591525	223	159	*β*-CGST-α	conserved hypothetical protein
			TxxC	
1591451	**222**	**47**	***β*-TPVC-α**	**peroxiredoxin**
1591758	295	233	*β*-TTAC-α	diaminopimelate epimerase (dapF)
1500567	458	364	α-TCVC-α	predicted coding region MJ1665

[a] Number of amino acids in an ORF. [b] Amino acid position corresponding to first amino acid in the C/TxxC/S/T sequence. [c] C/TxxC/S/T primary sequences and flanking secondary structure (α, α-helix; *β*, *β*-strand). [d] Predicted function or information obtained by database searches. Bold shows true positives.

cysteine that aligns with selenocysteine likely plays a redox function. These observations suggest that homology analyses of selenoproteins and their cysteine homologues might reveal the location of redox groups and predict redox motifs employed by these proteins.

We analyzed 18 known mammalian selenoproteins for the presence of CxxC-derived motifs, in which a cysteine is replaced with selenocysteine (U), using multiple sequence alignments of proteins present in the NCBI nonredundant database. Interestingly, of these 18 proteins, 12 possessed the CxxC-derived motifs, including four UxxT (GPx1, GPx2, GPx3, and GPx4), three SxxU (thyroid hormone deiodinases 1, 2 and 3), one UxxS (methionine-R-sulfoxide reductase SelR), one UxxC (Selenoprotein P) and three CxxU (selenoproteins M, T, and W) motifs. In addition, except for methionine-R-sulfoxide reductase, these selenoproteins had a predicted α-helix immediately downstream of the CxxC-derived motifs. Of the six remaining selenoproteins, four had selenocysteine that was flanked by Cys (thioredoxin reductases 1, 2, and 3 and selenoprotein N), and two had selenocysteine that was separated from a cysteine by a single residue (Sep15 and selenophosphate synthetase 2). We also analyzed known selenoproteins in other organisms and identified CxxC-derived motifs in several additional proteins, including Drosophila BthD, archaeal heterodisulfide reductase, and bacterial selenoprotein A of glycine reductase complex (CxxU motifs), bacterial proline reductase, and selenoproteins B of glycine, betaine, and sarcosine reductase complexes (UxxC motifs) and *Eubacterium acidaminophilum* peroxiredoxin PrxU (TxxU motif).

All mammalian selenoproteins containing the UxxT, SxxU, and UxxS motifs required thioredoxin or glutathione for their redox activity. However, while the redox reactions catalyzed by GPxs and methionine-R-sulfoxide reductases are known in detail, the deiodinase reaction was only partially characterized. In particular, mutation of selenocysteine completely inactivated the enzyme, whereas mutation of a residue within the SxxU motif had little effect on catalytic activity (*51*). Our analysis suggests that the serine in the SxxU motif might form a hydrogen bond with selenocysteine and that its mutation might decrease or abolish deiodinase catalytic activity.

While the presence of selenocysteine in the N-terminal position of the CxxC motif in some selenoproteins was not surprising, the C-terminal CxxC location of selenocysteine in selenoproteins M, T, and W will require further studies as to whether this residue serves a role of a nucleophlic attacking redox group.

DISCUSSION

CxxC is the major redox motif utilized for formation, reduction, and isomerization of disulfide bonds. In this study, we found a remarkable flexibility in cysteine (and selenocysteine) occurrences that this motif tolerates. In particular, one of the cysteines could be replaced with serine or threonine, the residues that may assist the remaining Cys in its redox function.

Reaction mechanisms of CxxC-containing redox proteins were previously thoroughly studied using thioredoxin as a model protein (*1*−*3*). These studies revealed that the N-terminal Cys in the CxxC motif was the nucleophilic attacking group, and subsequent studies with thioredoxin-like proteins, such as protein disulfide isomerase and DsbA, confirmed this finding. However, curious identification of thioredoxin homologues, in which the N-terminal CxxC Cys was absent (*13*), posed a dilemma. How could a redox function be maintained if the attacking redox group is lacking? Our study suggests that both cysteines in the CxxC motif potentially could serve as attacking redox groups, and the identity of the attacking nucleophilic thiolate in the CxxC motif is not known a priori. In addition, the cysteine in a CxxC-derived motif (SxxC, TxxC, CxxS, and CxxT), whether it aligns with the N-terminal or C-terminal CxxC cysteine, might be a redox residue.

Analyses of mammalian selenoproteins support these conclusions. Selenocysteine serves as a redox group in selenoproteins with known reaction mechanisms (*50*). We found that in the majority of selenoproteins, selenocysteine occurs in the context of CxxC-derived motifs. In fact, all of the CxxC-derived motifs discussed in this paper were represented in selenoproteins.

We also observed rare cases in which CxxC-derived motifs were reversed within a protein family. For example, in addition to the majority of thioredoxins that had the CxxC motif, several thioredoxin homologues possessed a SxxC motif (gi 16123259, 13236192, 22965617, and 2145977), whereas several other homologues had a CxxS motif (gi 10120659, 24644807, and 31220174). Similarly, in addition to many of the glutaredoxins with the CxxC motif, a large number of homologues were of the CxxS type, whereas one glutaredoxin homologue (gi 27716651) was detected that had the SxxC motif. We also identified a number of thioredoxin-fold proteins, in which both cysteines were replaced with

serine or threonine (for example gi 19551267, 16765802, 16273645, 19552441, and 19075082). These proteins likely lost their redox function.

Analyses of completely sequenced genomes identified peroxiredoxins as major representatives of the TxxC motif. However, previous alignments involving these proteins (*18*) suggested that the catalytic TxxC cysteine corresponded to the N-terminal Cys in the CxxC motif. We reevaluated alignments involving peroxiredoxins, thioredoxins, and other thioredoxin-fold proteins and found that, in agreement with our genomic searches, the catalytic Cys corresponded to the C-terminal Cys in the CxxC motif. Thus, peroxiredoxins provide an example of functionally characterized redox proteins, in which the N-terminal Cys of the CxxC motif is replaced with a nonredox residue, whereas the redox function is maintained by the C-terminal Cys. Molecular modeling experiments confirmed that threonine and cysteine in the TxxC motif could form a hydrogen bond.

In view of the conservation of the TxxC motif in peroxiredoxins and the proximity of the threonine to the catalytic cysteine in these enzymes, it is not clear why the role of the threonine has not been addressed in numerous functional and structural studies reported for these proteins. Nevertheless, some recent experimental data are consistent with our conclusions. Flohe and co-workers reported a mutational analysis of *Leishmania donovani* tryparedoxin peroxidase, which is a thioredoxin peroxidase homologue (*52*). The authors found that the replacement of the threonine in the TxxC motif with valine disrupted catalytic activity of the enzyme, whereas a serine-for-threonine mutation was tolerated, suggesting the involvement of the threonine in thiolate stabilization (*52*).

Major representatives of the CxxT motif were glutathione peroxidases. We could find no reference in the literature to the possible role of the conserved threonine in forming a hydrogen bond with the catalytic thiolate (or selenolate in selenocysteine-containing glutathione peroxidases), and this residue has not been subjected to mutational analysis. Tryptophan and glutamine were previously suggested to stabilize the thiolate (or selenolate) and, together with Cys (or selenocysteine), to serve as a catalytic triad. However, our molecular modeling experiments revealed that the hydrogen bond between selenocysteine and threonine in bovine GPx1 is feasible and might be preferred over that involving tryptophan and glutamine. The finding that the CxxT motif in glutathione peroxidases is derived from the CxxC motif was also consistent with the role of the threonine in stabilization of the thiolate.

Our observations could help explain some findings of redox proteins that lacked predicted active site redox groups, as well as to reevaluate claims of redox-independent functions of some thioredoxin-fold proteins, especially when these studies were based on mutations of only one of the CxxC cysteines. Thus, information on the identity of the attacking cysteine might be needed prior to interpreting the data, or alternatively, both cysteines have to be mutated to truly exclude redox function.

It should be pointed out that not all proteins that contain CxxC or CxxC-derived motifs followed by α-helices have a redox function. Indeed, the presence of only two residues in the motif inevitably generates false positives in genomic searches. However, it is clear that our search strategy allows dramatic enrichment of redox proteins and generates predictions of redox groups in these proteins, providing a basis for further experimental approaches. The usefulness of this method is illustrated by the following example. In our initial paper on the CxxS motif (*10*), we predicted a redox function for *E. coli* and *M. jannaschii* OsmC proteins and proposed that the cysteine in the CxxS motif in these proteins is redox active. Recently, two studies independently described this protein as thiol-dependent peroxidase and identified its CxxS cysteine as one of two redox-active residues (*53, 54*).

REFERENCES

1. Holmgren, A. (2000) Antioxidant function of thioredoxin and glutaredoxin systems. *Antioxid. Redox. Signal. 2*, 811−820.
2. Martin, J. L. (1995) Thioredoxin−a fold for all reasons. *Structure 3*, 245−250.
3. Ritz, D., and Beckwith, J. (2001) Roles of thiol-redox pathways in bacteria. *Annu. Rev. Microbiol. 55*, 21−48.
4. Chivers, P. T., Laboissiere, M. C., and Raines, R. T. (1996) The CXXC motif: imperatives for the formation of native disulfide bonds in the cell. *EMBO J. 15*, 2659−2667.
5. Chivers, P. T., Prehoda, K. E., and Raines, R. T. (1997) The CXXC motif: a rheostat in the active site. *Biochemistry 36*, 4061−4066.
6. Kortemme, T., and Creighton, T. E. (1995) Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pK_a values in proteins of the thioredoxin family. *J. Mol. Biol. 253*, 799−812.
7. Holmgren, A. (1989) Thioredoxin and glutaredoxin systems. *J. Biol. Chem. 264*, 13963−13966.
8. Verdoucq, L., Vignols, F., Jacquot, J. P., Chartier, Y., and Meyer, Y. (1999) In vivo characterization of a thioredoxin h target protein defines a new peroxiredoxin family. *J. Biol. Chem. 274*, 19714−19722.
9. Motohashi, K., Kondoh, A., Stumpp, M. T., and Hisabori, T. (2001) Comprehensive survey of proteins targeted by chloroplast thioredoxin. *Proc. Natl. Acad. Sci. U.S.A. 98*, 11224−11229.
10. Fomenko, D. E., and Gladyshev, V. N. (2002) CxxS: Fold-independent redox motif revealed by genome-wide searches for thiol/disulfide oxidoreductase function. *Protein Sci. 11*, 2285−2296.
11. Kumar, R. A., Koc, A., Cerny, R. L., and Gladyshev, V. N. (2002) Reaction Mechanism, Evolutionary Analysis, and Role of Zinc in Drosophila Methionine-R-sulfoxide Reductase. *J. Biol. Chem. 277*, 37527−37535.
12. Anelli, T., Alessio, M., Mezghrani, A., Simmen, T., Talamo, F., Bachi, A., and Sitia, R. (2002) ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. *EMBO J. 21*, 835−844.
13. Fetrow, J. S., Siew, N., Di Gennaro, J. A., Martinez-Yamout, M., Dyson, H. J., and Skolnick, J. (2001) Genomic-scale comparison of sequence- and structure-based methods of function prediction: does structure provide additional insight? *Protein Sci. 10*, 1005−1014.
14. Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res. 27*, 215−219.
15. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. (1997) Gapped BLAST and PSI−BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389−3402.
16. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrixes. *J. Mol. Biol. 292*, 195−202.
17. Jeng, M. F., Holmgren, A., and Dyson, H. J. (1995) Proton sharing between cysteine thiols in *Escherichia coli* thioredoxin: implications for the mechanism of protein disulfide reduction. *Biochemistry 34*, 10101−10105.
18. Schroder, E., and Ponting, C. P. (1998) Evidence that peroxiredoxins are novel members of the thioredoxin fold superfamily. *Protein Sci. 7*, 2465−2468.
19. Rhee, S. G., Kang, S. W., Netto, L. E., Seo, M. S., and Stadtman, E. R. (1999) A family of novel peroxidases, peroxiredoxins. *Biofactors 10*, 207−209.
20. Hofmann, B, Hecht, H. J., and Flohe, L. (2002) Peroxiredoxins. *Biol. Chem. 383*, 347−364.

21. Declercq, J. P., Evrard, C., Clippe, A., Stricht, D. V., Knoops, A., and Bernar, B. (2001) Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 Å resolution. *J. Mol. Biol. 311*, 751−759.
22. Epp, O., Ladenstein, R., and Wendel, A. (1983) The refined structure of the selenoenzyme glutathione peroxidase at 0.2-nm resolution. *Eur. J. Biochem. 133*, 51−69.
23. Ren, B., Huang, W., Akesson, B., and Ladenstein, R. (1997) The crystal structure of seleno-glutathione peroxidase from human plasma at 2.9 Å resolution. *J. Mol. Biol. 268*, 869−885.
24. Dirr, H., Reinemer, P., and Huber, R. (1994) X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *Eur. J. Biochem. 220*, 645−661.
25. Allardyce, C. S., McDonagh, P. D., Lian, L. Y., Wolf, C. R., and Roberts, G. C. (1999) The role of tyrosine-9 and the C-terminal helix in the catalytic mechanism of Alpha-class glutathione S-transferases. *Biochem J. 3*, 525−531.
26. Olson, J. W., Agar, J. N., Johnson, M. K., and Maier, R. J. (2000) Characterization of the NifU and NifS Fe−S cluster formation proteins essential for viability in *Helicobacter pylori. Biochemistry 39*, 16213−16219.
27. Seidler, A., Jaschkowitz, K., and Wollenberg, M. (2001) Incorporation of iron-sulphur clusters in membrane-bound proteins. *Biochem. Soc. Trans. 29*, 418−421.
28. Krebs, C., Agar, J. N., Smith, A. D., Frazzon, J., Dean, D. R., Huynh, B. H., and Johnson, M. K. (2001) IscA, an alternate scaffold for Fe−S cluster biosynthesis. *Biochemistry 40*, 14069−14080.
29. Ollagnier-de-Choudens, S., Mattioli, T., Takahashi, Y., and Fontecave, M. (2001) Iron−sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin. *J. Biol. Chem. 276*, 22604−22607.
30. Gasch, A. P., and Eisen, M. B. (2002) Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biology 3*, 1−22.
31. Xu, H. E., and Johnston, S. A. (1994) Yeast bleomycin hydrolase is a DNA-binding cysteine protease. Identification, purification, biochemical characterization. *J. Biol. Chem. 269*, 21177−21183.
32. Siniossoglou, S., and Pelham, H. R. (2001) An effector of Ypt6p binds the SNARE Tlg1p and mediates selective fusion of vesicles with late Golgi membranes. *EMBO J. 20*, 5991−5998.
33. Bao, Y., Yang, B. Z., Dawson, T. L., Jr., and Chen, Y. T. (1997) Isolation and nucleotide sequence of human liver glycogen debranching enzyme mRNA: identification of multiple tissue-specific isoforms. *Gene 197*, 389−398.
34. Nakayama, A., Yamamoto, K., and Tabata, S. (2000) High expression of glycogen-debranching enzyme in *Escherichia coli* and its competent purification method. *Protein Expr. Purif. 19*, 298−303.
35. McAteer, S., Coulson, A., McLennan, N., and Masters, M. (2001) The lytB gene of *Escherichia coli* is essential and specifies a product needed for isoprenoid biosynthesis. *J. Bacteriol. 183*, 7403−7407.
36. Rohdich, F., Hecht, S., Gartner, K., Adam, P., Krieger, C., Amslinger, S., Arigoni, D., Bacher, A., and Eisenreich, W. (2002) Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *Proc. Natl. Acad. Sci. U.S.A. 99*, 1158−1163.
37. Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. (1999) Thiamin biosynthesis in prokaryotes. *Arch. Microbiol. 171*, 293−300.
38. Zhang, Y., Taylor, S. V., Chiu, H. J., and Begley, T. P. (1997) Characterization of the *Bacillus subtilis* thiC operon involved in thiamine biosynthesis. *J. Bacteriol. 179*, 3030−3035.
39. Quay, S. C., and Oxender, D. L. (1976) Regulation of branched-chain amino acid transport in *Escherichia coli*. *J. Bacteriol. 127*, 1225−1238.
40. Drummond, J. T., Loo, R. R., and Matthews, R. G. (1993) Electrospray mass spectrometric analysis of the domains of a large enzyme: observation of the occupied cobalamin-binding domain and redefinition of the carboxyl terminus of methionine synthase. *Biochemistry 32*, 9282−9289.
41. Lamarche, A. E., Abate, M. I., Chan, S. H., and Trumpower, B. L. (1992) Isolation and characterization of COX12, the nuclear gene for a previously unrecognized subunit of *Saccharomyces cerevisiae* cytochrome *c* oxidase. *J. Biol. Chem. 267*, 22473−22480.
42. Taanman, J. W., and Capaldi, R. A. (1992) Purification of yeast cytochrome *c* oxidase with a subunit composition resembling the mammalian enzyme. *J. Biol. Chem. 267*, 22481−22485.
43. Shumilin, I. A., Kretsinger, R. H., and Bauerle, R. H. (1999) Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*. *Structure 7*, 865−875.
44. Schultz, J. E., and Matin, A. (1991) Molecular and functional characterization of a carbon starvation gene of *Escherichia coli*. *J. Mol. Biol. 218*, 129−140.
45. Dugaiczyk, A., Haron, J. A., Stone, E. M., Dennison, O. E., Rothblum, K. N., and Schwartz, R. J. (1983) Cloning and sequencing of a deoxyribonucleic acid copy of glyceraldehyde-3-phosphate dehydrogenase messenger ribonucleic acid isolated from chicken muscle. *Biochemistry 22*, 1605−1613.
46. Kim, H., Feil, I. K., Verlinde, C. L., Petra, P. H., and Hol, W. G. (1995) Crystal structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from Leishmania mexicana: implications for structure-based drug design and a new position for the inorganic phosphate binding site. *Biochemistry 34*, 14975−14986.
47. Brune, B., and Mohr, S. (2001) Protein Thiol Modification of Glyceraldehyde-3-phosphate Dehydrogenase and Caspase-3 by Nitric Oxide. *Curr. Protein. Pept. Sci. 2*, 61−72.
48. Cotgreave, I. A., Gerdes, R., Schuppe-Koistinen, I., and Lind, C. (2002) S-glutathionylation of glyceraldehyde-3-phosphate dehydrogenase: role of thiol oxidation and catalysis by glutaredoxin. *Methods Enzymol. 348*, 175−182.
49. Sparla, F., Pupillo, P., and Trost, P. (2002) The C-terminal extension of Glyceraldehyde-3-phosphate dehydrogenase subunit B acts as an autoinhibitory domain regulated by thioredoxins and nicotinamide adenine dinucleotide. *J. Biol. Chem. 277*, 44946−44952.
50. Hatfield, D. L., and Gladyshev, V. N. (2002) How selenium has altered our understanding of the genetic code. *Mol Cell Biol. 22*, 3565−3576.
51. Sun, B. C., Harney, J. W., Berry, M. J., and Larsen, P. R. (1997) The role of the active site cysteine in catalysis by type 1 iodothyronine deiodinase. *Endocrinology 138*, 5452−5458.
52. Flohe, L., Budde, H., Bruns, K., Castro, H., Clos, J., Hofmann, B., Kansal-Kalavar, S., Krumme, D., Menge, U., Plank-Schumacher, K., Sztajer, H., Wissing, J., Wylegalla, C., and Hecht, H. J. (2002) Tryparedoxin peroxidase of Leishmania donovani: molecular cloning, heterologous expression, specificity, and catalytic mechanism. *Arch. Biochem. Biophys. 397*, 324−335.
53. Lesniak, J., Barton, W. A., and Nikolov, D. B. (2002) Structural and functional characterization of the Pseudomonas hydroperoxide resistance protein Ohr. *EMBO J. 21*, 6649−6659.
54. Cussiol, J. R., Alves, S. V., Oliveira, M. A., and Netto, L. E. (2003) Organic hydroperoxide resistance gene encodes a thiol-dependent peroxidase. *J. Biol. Chem. 278*, 11570−11578.

BI034459S