Divergence of Function in the Thioredoxin Fold Suprafamily: Evidence for Evolution of Peroxiredoxins from a Thioredoxin-like Ancestor†

Shelley D. Copley,*,‡ Walter R. P. Novak,§ and Patricia C. Babbitt*,§

Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309, and Department of Biopharmaceutical Sciences and California Institute for Quantitative Biomedical Research, University of California, San Francisco, California 94143

Received May 24, 2004; Revised Manuscript Received August 12, 2004

ABSTRACT: The thioredoxin fold is found in proteins that serve a wide variety of functions. Among these are peroxiredoxins, which catalyze the reduction of hydrogen peroxide and alkyl peroxides. Although the common structural fold shared by thioredoxins and peroxiredoxins suggests the possibility that they have evolved from a common progenitor, it has been difficult to examine this hypothesis in depth because pairwise sequence identities between proteins in these two superfamilies are statistically insignificant. Using the Shotgun program, we have found that sequences of reductases involved in maturation of cytochromes in certain bacteria bridge the sequences of thioredoxins and peroxiredoxins. Analysis of motifs found in a divergent set of thioredoxins, cytochrome maturation proteins, and peroxiredoxins provides further support for an evolutionary relationship between these proteins. Within the conserved motifs are specific residues that are characteristic of individual protein classes, and therefore are likely to be involved in the specific functions of those classes. We have used this information, in combination with existing structural and functional information, to gain new insight into the structure−function relationships in these proteins and to construct a model for the emergence of peroxiredoxins from a thioredoxin-like ancestor.

In recent years, it has become clear that the number of protein sequences vastly exceeds the number of distinct structural folds, and that nature often utilizes the same structural fold for multiple purposes (*1*). Understanding the divergence of an ancestral structure to serve many new functions requires insight into the architectural design principles embedded in the protein scaffold. Thus, we might expect constraints locked into the fundamental structure−function relationship in an ancestral structure to guide evolutionary changes leading to new functions.

The thioredoxin (Trx)[1] fold is a common structural fold found in a large number of proteins that carry out disparate functions. These include Trxs, protein disulfide isomerases (*2*), DsbAs (*3*), and glutaredoxins (*3*), all of which catalyze thiol−disulfide exchange reactions, glutathione *S*-transferases (*3*), which catalyze formation of glutathione conjugates from electrophilic substrates, calsequestrins (*4*), which bind calcium ions, and glutathione peroxidases (*3*) and peroxiredoxins (*5*), which catalyze reduction of peroxides. Three-dimensional structural similarities, and, in some cases, functional similarities, among these different superfamilies have been noted previously by many investigators. Although structural information can suggest very distant evolutionary relationships that are difficult to detect using sequence information alone, structural information is often insufficient for in-depth analysis of evolutionary relationships. Given the variety of different proteins in each of these superfamilies, an analysis that integrates sequence and functional/mechanistic information with structural observations is required to understand the evolution of these proteins and to provide more detailed insight into similarities and differences in their structure−function relationships. Yet, this is a daunting challenge. Except for protein disulfide isomerases, previous analyses have failed to identify statistically significant pairwise sequence identities between Trxs and other proteins of the Trx fold class. Thus, the Trx fold proteins provide an intriguing and challenging system in which to investigate questions of convergent or divergent evolution leading to disparate functions in a common structural scaffold.

Using tools designed to identify sequences that link very distantly related proteins, we have performed sequence and motif analyses on a broader set of sequences than has previously been investigated and here present evidence of sequence links between the Trxs and peroxiredoxins (Prxs). Using this set of linking sequences, we have been able to identify sequence motifs conserved among these proteins and deduce explicit correlations between similarities and differences in their functions and similarities and differences in their sequences and structures. These data support and extend the structural evidence that these proteins diverged from a common ancestor. Taken together, the sequence, structural,

† Supported by NSF Grant MCB-0077569 (to S.D.C.) and NIH Grant GM60595 (to P.C.B.). W.R.P.N. is a Burroughs Wellcome predoctoral fellow.

* To whom correspondence should be addressed. S.D.C.: phone, (303) 492-6328; fax, (303) 492-1149; e-mail, shelley@cires.colorado.edu. P.C.B.: phone, (415) 476-3784; fax, (415) 514-4260; e-mail, babbitt@cgl.ucsf.edu.

‡ University of Colorado.

§ University of California.

[1] Abbreviations: CMP, cytochrome maturation protein; Prx, peroxiredoxin; Trx, thioredoxin.

10.1021/bi048947r CCC: $27.50
Published on Web 10/12/2004

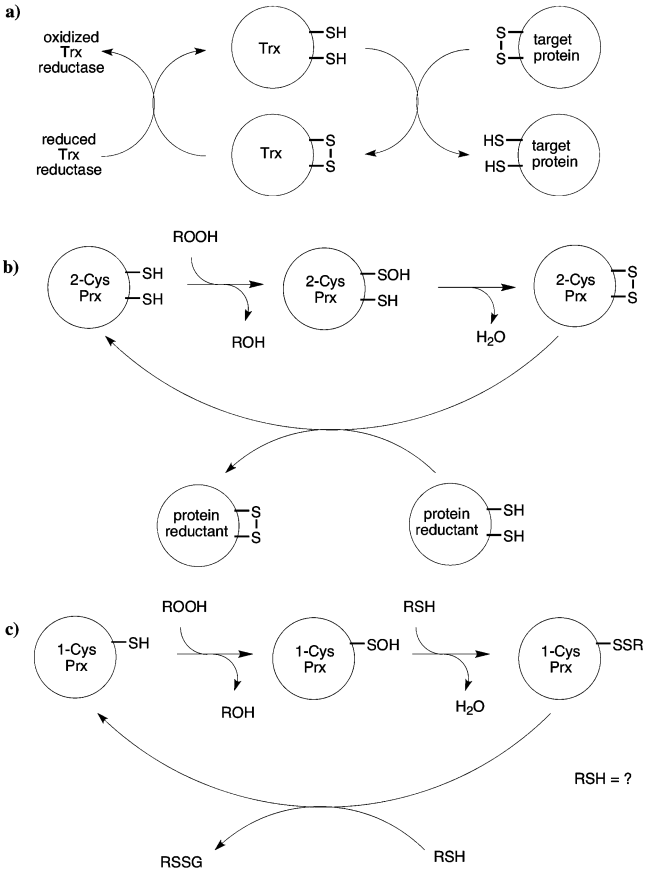


FIGURE 1: Mechanisms of (a) Trxs, (b) 2-Cys Prxs, and (c) 1-Cys Prxs. In 2-Cys Prxs, the second Cys may come from the same subunit or the other subunit in a dimeric enzyme. The protein reductants that regenerate reduced 2-Cys Prxs are cell-specific, and include Trx, glutaredoxin, AhpF, tryparedoxin, and AphD. Physiological reductants for 1-Cys Prxs are not known.

and functional links between the Trx and Prx superfamilies reveal, on a larger scale than has previously been attempted, fundamental design principles inherent in the Trx fold class and allow us to develop a model for how Prxs evolved from a functionally simpler Trx-like ancestor that carried out thiol−disulfide exchange reactions.

Trxs are small, simple, and ancient proteins that are found in all living organisms (see refs *6*−*9* for reviews). They serve a variety of functions, most of which are based upon their thiol oxidoreductase activity, which is endowed by a CXXC sequence that is partially exposed at the surface of the protein. The two cysteines in this motif (Figure 1a) are used to reduce disulfide bonds in at least 30 target proteins. The resulting Trx disulfide is then reduced by Trx reductases that transfer reducing equivalents derived from either NADPH or ferredoxins. Some of the target proteins are enzymes in which a disulfide bond is formed at the active site during substrate turnover, while others are enzymes or transcription factors whose function is regulated by the reduction of a disulfide bond. Trxs also play important roles in defense against oxidative stress. They can directly reduce hydrogen peroxide, dehydroascorbate, and certain radicals, and can serve as reductants for Prxs. In humans, Trx is secreted by both normal and cancer cells and acts as a growth factor (*7*). Some of the activities of Trxs do not depend on the active site CXXC. In *Escherichia coli*, inactive mutants of Trx are able to support assembly of F1 and M13 phages (*10*) and to activate phage T7 polymerase (*11*).

Prxs contain a canonical Trx fold, but have an N-terminal extension, an insertion between β2 and α2 of the Trx fold, and, in some cases, a C-terminal extension. Prxs catalyze the reduction of hydroperoxides and alkyl peroxides formed by oxidative damage to cellular molecules, especially lipids (*5*, *12*). Recent studies in which Prx genes have been disrupted have confirmed the physiological role of these enzymes in protecting cells from oxidative stress (*13*−*15*). In eukaryotes, Prxs may play an additional role in cell signaling by modulating hydrogen peroxide levels (*16*). Reduction of peroxides begins with attack of an active site Cys (or, in rare cases, a Se-Cys) on an oxygen atom of the peroxide, leading to cleavage of the O−O bond (Figure 1b,c). The sulfenic acid formed in the first step is subsequently attacked by a second Cys (termed the resolving Cys), forming a disulfide and releasing hydroxide. The resolving Cys may come from the same protein chain or the other monomer in a dimer (Figure 1b), or a different molecule such as Trx or glutaredoxin (Figure 1c). Finally, the free Prx is regenerated by reduction of the disulfide; the physiological reductant has not been identified in all cases.

The structural similarities between Trxs and Prxs have been recognized for some time (*17*), leading to suggestions that Prxs might have evolved from Trxs. In addition, before the first structure of a Prx was available, Schröder and Ponting (*18*) generated a multiple-sequence alignment of Trxs and Prxs using ClustalW, secondary structure predictions for Prxs, and structural alignments of several Trx fold proteins to support such a relationship. However, the low pairwise sequence identities between members of these superfamilies made it difficult to assess this hypothesis in detail. Moreover, their alignment fails to align functionally important residues that we now know superimpose in the active sites of the Trxs and Prxs (see below).

A valuable approach for detection of relationships among divergent sequences such as the Trxs and Prxs is to look for proteins whose sequences show a convincing relationship to both types of proteins. The transitive property of sequence relationships suggests that, for consonant domains, if A is related to B, and B is related to C, then A must be related to C. Systematic analysis of the effectiveness of this approach has been documented (*19*). We have used the Shotgun program (*20*) to identify proteins whose sequences link those of Trxs and Prxs, and thus provide support for an evolutionary relationship between these superfamilies. Analysis of sequence motifs found in a set of 105 divergent proteins further supports the proposed evolutionary relationship and, in the context of the available structural and functional information, allows us to postulate a model for the evolutionary divergence of the ancient Trx fold to provide new catalytic capabilities in Prxs.

METHODS

The analysis of the relationship between Trxs and Prxs was begun using PSI-BLAST (*21*) to generate a divergent set of proteins related to structurally characterized Trxs and Prxs in which no pairwise identity was greater than 45%. Additional proteins that linked the Trxs and Prxs were found using the Shotgun program (*20*). After several iterations of

Table 1: Summary of Structural Features in Trxs, CMPs, and Four Classes of Prxs

Family	N-terminal extension	Central insertion[a]	C-terminal extension	Active site consensus sequence[b]	*cis*-Pro or Arg	Position of resolving Cys in Prxs when present	Quaternary structure
Trx	-	-	–	FWADWCGPCKMIA	*cis*-Pro	NA	monomer
CMP	α–ββ3_{10} or Loop-ββα	αβ	–	FWATWCGPCRxEA	*cis*-Pro	NA	monomer
Prx1	+ (structure unknown)	+ (structure unknown)	–	FYPLAFTPVCTKEAC	Arg	5 residues from active site Cys	monomer
Prx2	ββ–Loopββ3_{10}	αβ	–	FPSIDTGVCAAQ/SVR	Arg	Central insertion	dimer (face-to-edge)
Prx3	Loop-ββα	αβ	–	VPGAFTPTCSANHL	Arg	C-terminus	dimer (face-to-edge)
Prx4	ββ–Loop-ββ3_{10}	αβ	+	YPADFTFVCPTEL	Arg	C-terminal extension	dimer (edge-to-edge)

[a] Central insertion located between β2 and α2 of the canonical Trx fold. [b] Consensus sequences were determined using the MEME algorithm for each protein family.

Shotgun to add new homologues, a set of sequences filtered, with a few exceptions, at a 45% redundance level was assembled for motif analysis. While it is ideal to perform motif analysis using only divergent sequences, in some cases it was necessary to include a few sequences with higher pairwise identities in order to ensure adequate representation of sequence features characteristic of some groups of proteins. MEME (*22*; http://meme.sdsc.edu/meme/website/intro.html) was used to identify motifs present in the entire divergent set, as well as subsets of the set.

Structural superpositions were generated using MinRMS (*23*; http://www.cgl.ucsf.edu/Research/minrms/). The superposition of CMP (PDB entry 1KNG) and Trx (PDB entry 1XOA) aligns 97 α-carbons with an rmsd of 1.87 Å. Superpositions of CMP (PDB entry 1KNG) with members of Prx classes 2−4 aligned as follows (number of α-carbons aligned and rmsd): 115 and 2.00 Å for 1PSQ, 116 and 1.97 Å for 1HD2, and 127 and 1.99 Å for 1PRX, respectively. Superpositions were visualized using Chimera and the MinRMS Plot extension (*24*; http://www.cgl.ucsf.edu/chimera/). Figures 3, 4b, and 5−7 were generated with Chimera (*24*). In the case of Figures 3 and 5, the structures were first superposed, and then individual structures from the superposition were used to make the individual panels.

RESULTS

Cytochrome Maturation Protein Sequences Link the Sequences of Trxs and Prxs. We have identified proteins whose sequences link those of Trxs and Prxs, lending support to the hypothesis that the Prxs diverged from a Trx-like ancestor. We began our analysis with a query set comprised of Trxs and Prxs, none of which had more than 45% identity to any other member of the set. Shotgun (*20*) was used to identify potential relatives of both Trxs and Prxs (i.e., "bridging" proteins) by congruence analysis of BLAST outputs for the entire query set. The Shotgun analysis was iterated for several cycles, with addition of new bridging proteins to the query set at each iteration. Ultimately, a divergent set of 105 Trxs, Prxs, and bridging proteins was assembled in which no pairwise sequence identity was greater than 72% (and was generally less than 35%) within each group, or greater than 26% between members of these three groups. The bridging proteins for which functions are known are involved in maturation of cytochromes [e.g., CcmG from *E. coli* (*25*) and TlpA from *Bradyrhizobium japonicum* (*26*)]. [It has previously been noted that the structure of TlpA is similar to those of Trxs and Prxs (*26*).] These cytochrome maturation proteins (CMPs) catalyze thiol oxidoreductase reactions, just as Trxs do, lending further support to the suggestion that these proteins are homologous. However, while Trxs generally interact with a large number of protein substrates, CMPs are dedicated reductases involved in specific pathways. Thus, CcmG and TlpA are believed to be specific for targets in the pathways for biogenesis of cytochrome *c* (*27*) and cytochrome aa_3 (*28*), respectively. There are additional functional and structural differences between CMPs and Trxs, as well. CMPs are located in the periplasm, while Trxs are generally found in the cytoplasm. The redox potentials of their active sites are expected to differ to allow activity in the differing redox environments of these two compartments. Finally, CMPs contain a large N-terminal extension and an insertion after β2 of the canonical Trx fold (see Table 1 and further discussion below).

Motifs Found in Trxs and CMPs. We carried out motif analysis using the MEME algorithm (*22*) to further investigate the possible evolutionary relationship among Trxs, CMPs, and Prxs. Because of the extreme divergence of the proteins in the set, this analysis was divided into two parts. First, we identified motifs for the subset of sequences consisting of Trxs, CMPs, and homologous proteins of unknown function. Second, we carried out a comparable analysis for the subset consisting of CMPs, Prxs, and homologous proteins of unknown function.[2] The same set of CMPs was used in both analyses.

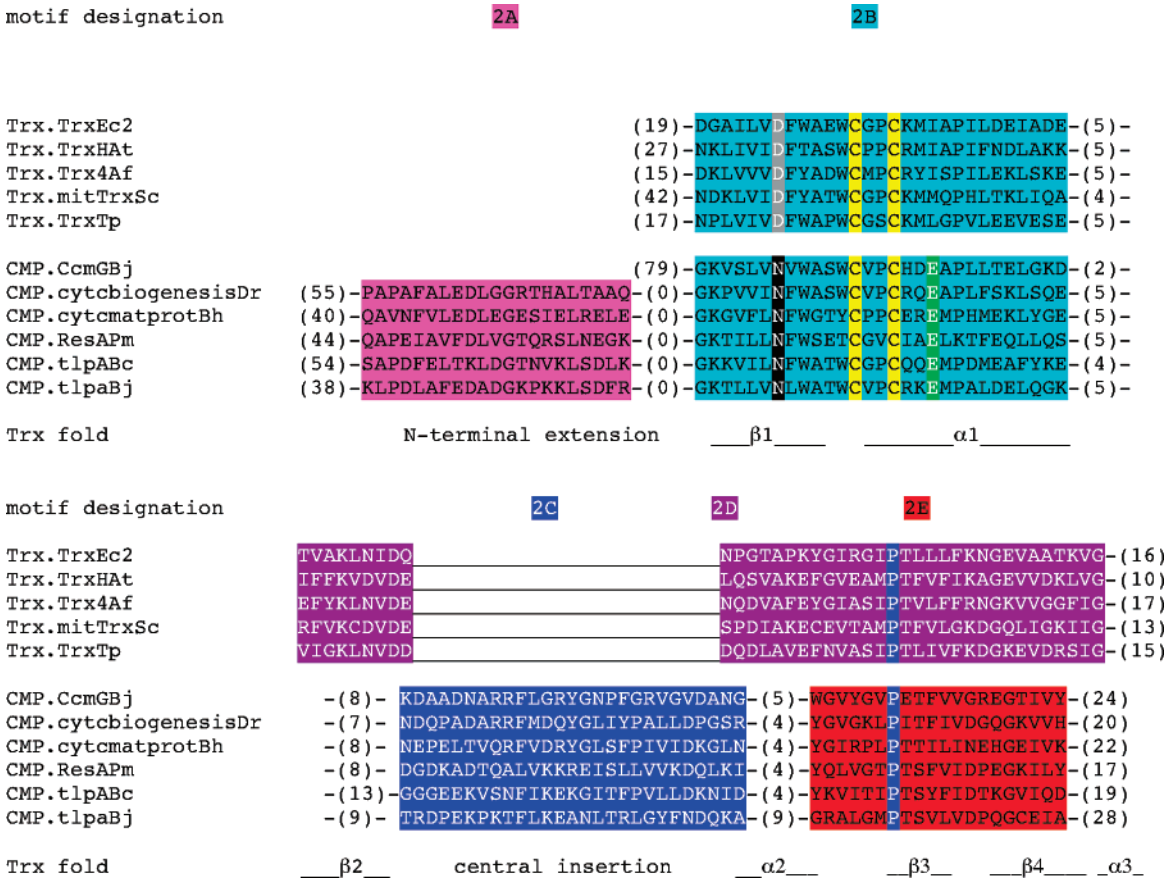


FIGURE 2: Motifs found in a set of 56 Trxs, CMPs, and homologous proteins of unknown function by MEME. Only a subset of the 56 sequences is shown. The secondary structural elements of the canonical Trx fold and the positions of the N-terminal extension and central insertion are shown below the motifs. Certain residues in the motifs are highlighted in colors corresponding to those used in Figure 3. On the basis of motif patterns, 34 proteins were assigned as Trxs and 22 as CMPs. The *E* values and number of occurrences for each motif are as follows: magenta motif (2A), 3.8e-039, 19 CMPs; cyan motif (2B), 1.5e-676, 33 Trxs and 22 CMPs; blue motif (2C), 4.6e-27, 22 CMPs; purple motif (2D), 4.1e-297, 30 Trxs; and red motif (2E), 5.2e-060, 21 CMPs and 2 Trxs. Trx.TrxEc2, Trx, *E. coli*, PDB entry 1XOA. Trx.TrxHAt, Trx, *Arabidopsis thaliana*, gi1388082. Trx.Trx4Af, Trx, *Archaeoglobus fulgidus*, gi11499727. Trx.mitTrxSc, Trx3, mitochondrial precursor, *Saccharomyces cerevisiae*, gi1907116. Trx.TrxTp, Trx, *Treponema pallidum*, gi3323237. CMP.CcmGBj, CcmG, *B. japonicum*, gi27375582. CMP.cytcbiogenesisDr, cytochrome *c* biogenesis protein, thiol−disulfide interchange protein, *Deinococcus radiodurans*, gi15805374. CMP.cytcmatprotBh, CMP.ResAPm, ResA, *Pasteurella multocida*, gi15602312. CMP.tlpABc, thiol−disulfide interchange protein tlpA, *Bacillus cereus*, gi30021713. CMP.tlpaBj, TlpA, *B. japonicum*, gi15988313, PDB entry 1JFU.

A set of divergent sequences containing 56 Trxs, CMPs, and homologous proteins of unknown function was analyzed using MEME. The motifs identified for a subset of these sequences are illustrated in Figure 2, and conserved functionally important residues in these motifs are shown in the active sites of a Trx and a CMP in Figure 3. (Although the N-terminal and C-terminal regions of these proteins have been shown to be important for function, they are not depicted in Figure 2 because no statistically significant motifs were found in these regions.) MEME identified a very strong motif (2B, cyan in Figure 2) containing the active site CXXC in 55 of 56 proteins in the set. In both Trxs and CMPs, the other side of the active site (see Figure 3) contains a conserved *cis*-Pro required for binding of polypeptide substrates. MEME finds different motifs for the two families in the region of the conserved *cis*-Pro (2D, purple, and 2E, red, in Figure 2). Several residues in this region are involved in substrate specificity in Trxs; in the structures of fragments of NF-κB (*29*) and Ref-1 (*30*) liganded to *E. coli* Trx1,[3] nine and five residues, respectively, preceding the *cis*-Pro contact the target peptide. Since much of this region is involved in binding target peptides, it is not surprising that the sequences have diverged sufficiently that a single motif is no longer recognizable. MEME also identifies two motifs in CMPs that are located within the N-terminal extension (2A, magenta in Figure 2) and the central insertion found after β2 of the canonical Trx fold (2C, blue in Figure 2). Although only one sequence motif was found by MEME in both Trxs and CMPs, the extremely strong statistical significance of this motif (E = 1.5e-676), coupled with the previously recognized structural and functional similarities between these proteins, provides convincing evidence that these proteins are indeed related. It is possible that no additional motifs were found

[2] Analysis of the entire set of proteins identified similar motifs, but both the *E* values for each motif and the number of occurrences of each motif were lower in the composite analysis due to the extreme divergence of the proteins in the set. Thus, breaking the analysis into two parts produced results that were statistically stronger and more useful for identifying functionally important changes in sequence between the Trxs and CMPs in one case, and CMPs and Prxs in the second case.

[3] Our set did not include *E. coli* Trx1, but the homologous regions of *E. coli* Trx2 and the other Trxs are readily identifiable.

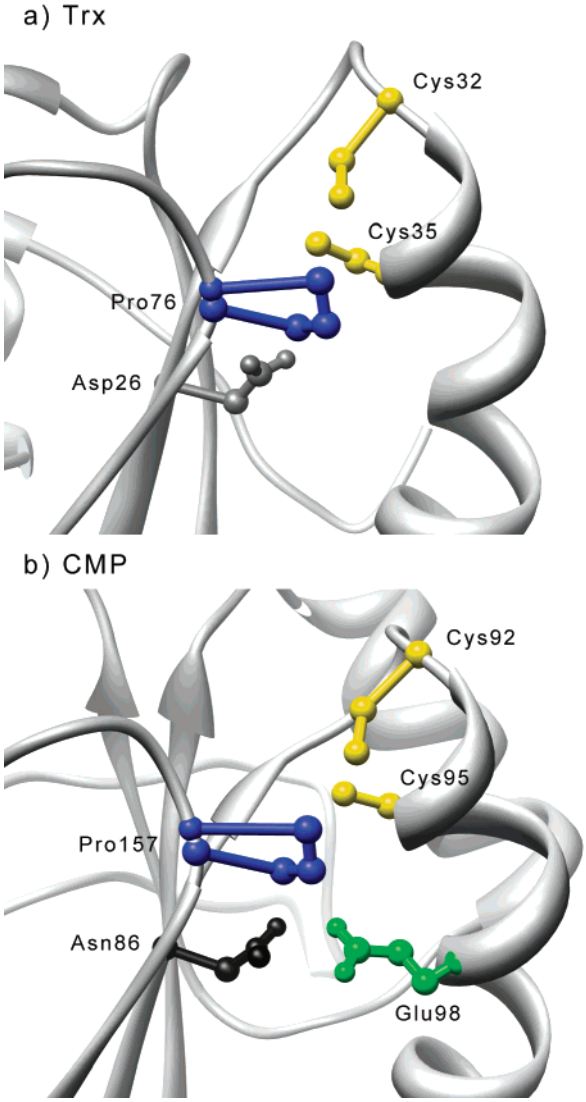


FIGURE 3: Close-up view of the active sites of (a) *E. coli* Trx (PDB entry 1XOA) and (b) a CMP, *B. japonicum* CcmG (PDB entry 1KNG).

because the sequences are extremely divergent, possibly because they diverged a very long time ago.

Motifs Found in CMPs and Prxs. A set of 67 CMPS, Prxs, and homologues of unknown function was analyzed by MEME. The motifs found are shown for a subset of the proteins in Figure 4a and mapped onto the backbone structures of a CMP and three Prxs in Figure 4b. (The rationale for division of Prxs into four classes designated Prx1−Prx4 will be described below in the Discussion.) Close-up views of the active sites of representative proteins are provided in Figure 5. Five motifs are strongly conserved across both protein families. These include three motifs located in the active site region (4C, cyan, 4D, brown, and 4G, red, in Figure 4), one in the N-terminal extension (4B, magenta in Figure 4), and one located partially within the central insertion in the Trx fold (4F, green in Figure 4). Three additional motifs (4A, black, 4E, blue, and 4H, yellow, in Figure 4) are conserved only in Prxs or in a subset of Prxs.

The most strongly conserved motif (4C, cyan in Figure 4) contains the active site cysteine(s) (one in Prxs and two in CMPs). A second motif (4D, brown in Figure 4) extends beyond this region, and contains additional residues near the active site. (The distinctive differences in this region in one subset of Prxs prevent this entire region from being recognized as a single motif.) A third highly conserved motif (4G, red in Figure 4) contains the conserved *cis*-Pro in CMPs. Notably, Prxs lack this *cis*-Pro, having an Arg in the homologous position. Nevertheless, there is sufficient conservation of sequence in this region that a motif can be recognized. This finding is particularly significant given that MEME finds two different motifs for this region in Trxs and CMPs, which have in common the structurally aligned *cis*-Pro and would appear to be more closely related to each other based upon functional considerations. Lending support to this alignment and to the assertion that these proteins are homologous, we note that this *cis*-Pro in CMPs and the Arg in Prxs also align in the structural comparisons shown in Figure 5.

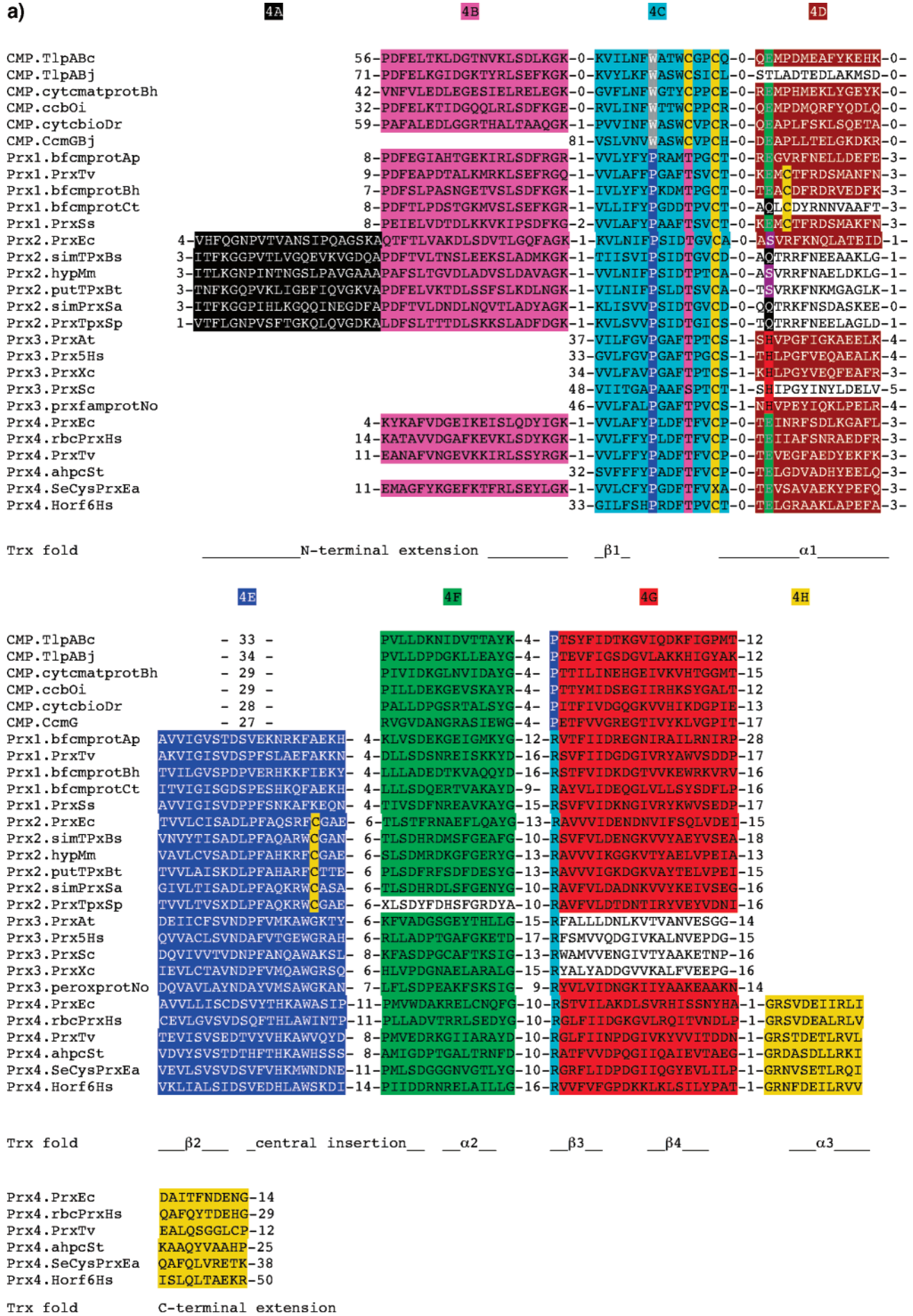
Two motifs found by MEME are located in the N-terminal extension. The magenta motif (4B) shown in Figure 4 is found in the N-terminal extensions of CMPs and many Prxs. In the structurally characterized proteins in the set that have this motif, it forms a $\beta\beta\alpha$ module on the side of the protein opposite to the active site (Figure 4b). We speculate that this module may have arisen in CMPs to facilitate formation of protein complexes involved in biogenesis of cytochromes, and been retained in the Prxs even though it has no obvious function. The black motif (4A in Figure 4) found in a subset of bacterial Prxs forms an additional $\beta\beta$-loop structure that buttresses a loop between $\alpha2$ and $\beta3$ of the canonical Trx fold that is involved in the dimeric interface (Figure 4b, structure b).

Two additional motifs found by MEME involve the central insertion between $\beta2$ and $\alpha2$ of the canonical Trx fold. The blue motif (4E) shown in Figure 4 is found only in Prxs, while the green motif (4F) is found in both CMPs and Prxs. Both of these motifs contain loops involved in the dimer interface in *E. coli* thiol peroxidase and human Prx5, although the residues involved are different in these two proteins. In addition, parts of both motifs are involved in the interface between adjacent dimers in the decameric structure that appears to form when some Prxs are oxidized (*31*, *32*). However, not all Prxs form decamers when they are oxidized (*33*).

The yellow motif (4H) shown in Figure 4 is peculiar to a subset of dimeric Prxs that have a C-terminal extension extending into the active site of the other monomer in the dimer (*5*). Often, but not always, this C-terminal extension contains a resolving Cys.

DISCUSSION

An Evolutionary Relationship among Trxs, CMPs, and Prxs Is Supported by Several Lines of Evidence. An evolutionary relationship among Trxs, CMPs, and Prxs is supported by several lines of evidence. First, as noted by many previous studies, they share a common structural fold. Second, the identification of CMPs as relatives of both Trxs and Prxs provides a sequence link between the Trxs and Prxs. Third, CMPs and Prxs contain insertions at comparable positions in the canonical Trx fold (Figure 6). Fourth, despite the enormous sequence divergence among the proteins used in the analysis, Trxs and CMPs share a strongly conserved


a)
4A
4B
4C
4D
CMP.TlpABc
CMP.TlpABj
CMP.cytcmatprotBh
CMP.ccbOi
CMP.cytcbioDr
CMP.CcmGBj
Prx1.bfcmprotAp
Prx1.PrxTv
Prx1.bfcmprotBh
Prx1.bfcmprotCt
Prx1.PrxSs
Prx2.PrxEc
Prx2.simTPxBs
Prx2.hypMm
Prx2.putTPxBt
Prx2.simPrxSa
Prx2.PrxTpxSp
Prx3.PrxAt
Prx3.Prx5Hs
Prx3.PrxXc
Prx3.PrxSc
Prx3.prxfamprotNo
Prx4.PrxEc
Prx4.rbcPrxHs
Prx4.PrxTv
Prx4.ahpcSt
Prx4.SeCysPrxEa
Prx4.Horf6Hs
Trx fold
N-terminal extension
β1
α1
4E
4F
4G
4H
CMP.TlpABc
CMP.TlpABj
CMP.cytcmatprotBh
CMP.ccbOi
CMP.cytcbioDr
CMP.CcmG
Prx1.bfcmprotAp
Prx1.PrxTv
Prx1.bfcmprotBh
Prx1.bfcmprotCt
Prx1.PrxSs
Prx2.PrxEc
Prx2.simTPxBs
Prx2.hypMm
Prx2.putTPxBt
Prx2.simPrxSa
Prx2.PrxTpxSp
Prx3.PrxAt
Prx3.Prx5Hs
Prx3.PrxSc
Prx3.PrxXc
Prx3.peroxprotNo
Prx4.PrxEc
Prx4.rbcPrxHs
Prx4.PrxTv
Prx4.ahpcSt
Prx4.SeCysPrxEa
Prx4.Horf6Hs
Trx fold
β2
central insertion
α2
β3
β4
α3
Prx4.PrxEc
Prx4.rbcPrxHs
Prx4.PrxTv
Prx4.ahpcSt
Prx4.SeCysPrxEa
Prx4.Horf6Hs
Trx fold
C-terminal extension

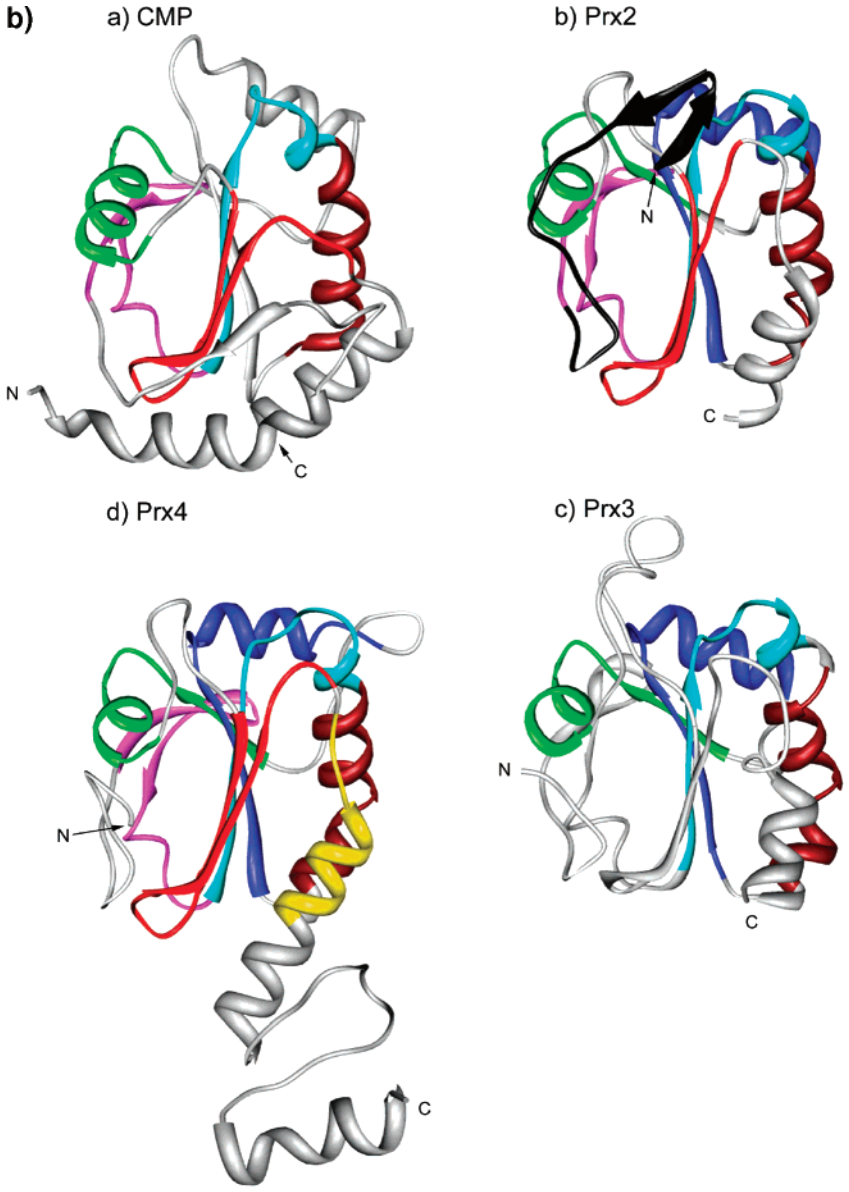


FIGURE 4: (a) Motifs found in a set of 67 CMPs, Prxs, and homologous proteins of unknown function by MEME. Only a subset of the sequences is shown. The secondary structural elements of the canonical Trx fold and the positions of the N-terminal extension and central insertion are shown below the motifs. Residues in the motifs are highlighted in colors corresponding to those used in Figure 5. The cysteines highlighted in motifs 4D and 4E correspond to resolving cysteines described in Table 2 and are not represented in Figure 5. The residue designated as X in the cyan motif (4A) in Prx4.SeCysPrxEa is selenocysteine. On the basis of motif patterns, 22 proteins were assigned as CMPs, 15 as Prx1s, nine as Prx2s, eight as Prx3s, and 13 as Prx4s. The *E* values and number of occurrences for each motif are as follows: black motif (4A), 8.3e-014, 8 Prx2s; magenta motif (4B), 3.1e-163, 19 CMPs, 12 Prx1s, 8 Prx2s, and 6 Prx4s; cyan motif (4C), 3.3e-359, 20 CMPs, 15 Prx1s, 9Prx2s, 8 Prx3s, and 13 Prx4s; brown motif (4D), 9.5e-025, 20 CMPs, 11 Prx1s, 6 Prx3s, and 13 Prx4s; blue motif (4E), 1.2e-202, 2 CMPs, 15 Prx1s, 9 Prx2s, 8 Prx3s, and 13 Prx4s; green motif (4F), 3.1e-068, 17 CMPs, 12 Prx1s, 7 Prx2s, and 12 Prx4s; red motif (4G), 1.7e-170, 20 CMPs, 14 Prx1s, 9 Prx2s, 2 Prx3s, and 13 Prx4s; and yellow motif (4H), 1.7e-048, 13 Prx4s. CMP.tlpABc, TlpA, *Ba. cereus*, gi30021713. CMP.tlpABj, TlpA, *B. japonicum*, gi15988313, PDB entry 1JFU. CMP.cytcmatprotBh, cytochrome *c* maturation protein, *Bacillus halodurans*, gi15614140. CMP.ccbOi, cytochrome *c* biogenesis protein, *Oceanobacillus iheyensis*, gi23098307. CMP.cytcbioDr, cytochrome *c* biogenesis protein, *D. radiodurans*, gi15805374. CMP.CcmGBj, CcmG, *B. japonicum*, gi27375582. Prx1.bfcmprotAp, bacterioferritin comigratory protein, *Aeropyrum pernix*, gi14600438. Prx1.PrxTv, Prx, *Thermoplasma volcanium*, gi13541117. Prx1.bfcmprotBh, bacterioferritin comigratory protein, *Ba. halodurans*, gi15613511. Prx1.bfcmprotCt, bacterioferritin comigratory protein, *Chlorobium tepidum*, gi21674812. Prx1.PrxSs, Prx, *Sulfolobus solfataricus*, gi15899339. Prx2.PrxEc, thiol peroxidase, *E. coli* O157:H7 EDL933, gi15801846, PDB entry 1QXH. Prx2.simTPxBs, similar to thiol peroxidase from *Bacillus subtilis*, gi16080001. Prx2.hypMm, hypothetical protein Magn029381, *Magnetospirillum magnetotacticum*, gi23014942. Prx2.putTPxBt, putative thiol peroxidase, *Bacteroides thetaiotaomicron* VPI-5482, gi29346739. Prx2.simPrxSa, similar to thioredoxin peroxidase, *Staphylococcus aureus*, gi21283385. Prx2.TpxSp, thiol peroxidase, *Streptococcus pneumoniae*, gi33358146, PDB entry 1PSQ. Prx3.PrxAt Prx TPx2, *A. thaliana*, gi4704732. Prx3.Prx5Hs, Prx 5, *Homo sapiens*, gi15826629, PDB entry 1HD2. Prx3.PrxXc, Prx, *Xanthomonas campestris*, gi21112072. Prx3.PrxSc, Ahp1p, *S. cerevisiae*, gi6323138. Prx3.prxfamprotNo, peroxiredoxin 2 family protein, *Nostoc* sp. PCC 7120, gi17229033. Prx4.PrxEc, thioredoxin peroxidase, *Encephalitozoon cuniculi*, gi19173077. Prx4.rbcPrxHs, thioredoxin peroxidase B, red blood cells, *H. sapiens*, gi9955016, PDB entry 1QMV. Prx4.PrxTv, Prx, *Thermoplasma volcanium*, gi13541053. Prx4.ahpSt, Ahpc, *Salmonella typhimurium*, gi20151112, PDB entry 1KYG. Prx4.SecysPrxEa, selenocysteine-containing peroxiredoxin, PrxU, *Eubacterium acidaminophilum*, gi6850955. Prx4Horf6.Hs, human Orf6, *H. sapiens*, gi3318841, PDB 1PRX. (b) Motifs shown in Figure 4a mapped onto the backbone structures of (a) a CMP, *B. japonicum* TlpA (PDF entry 1JFU), (b) a class 2 Prx, *S. pneumoniae* thiol peroxidase (PDG entry 1PSQ), (c) a class 3 Prx, human Prx5 (PDB entry 1HD2), and (d) a class 4 Prx, human red blood cell thioredoxin peroxidase (PDB entry 1QMV).

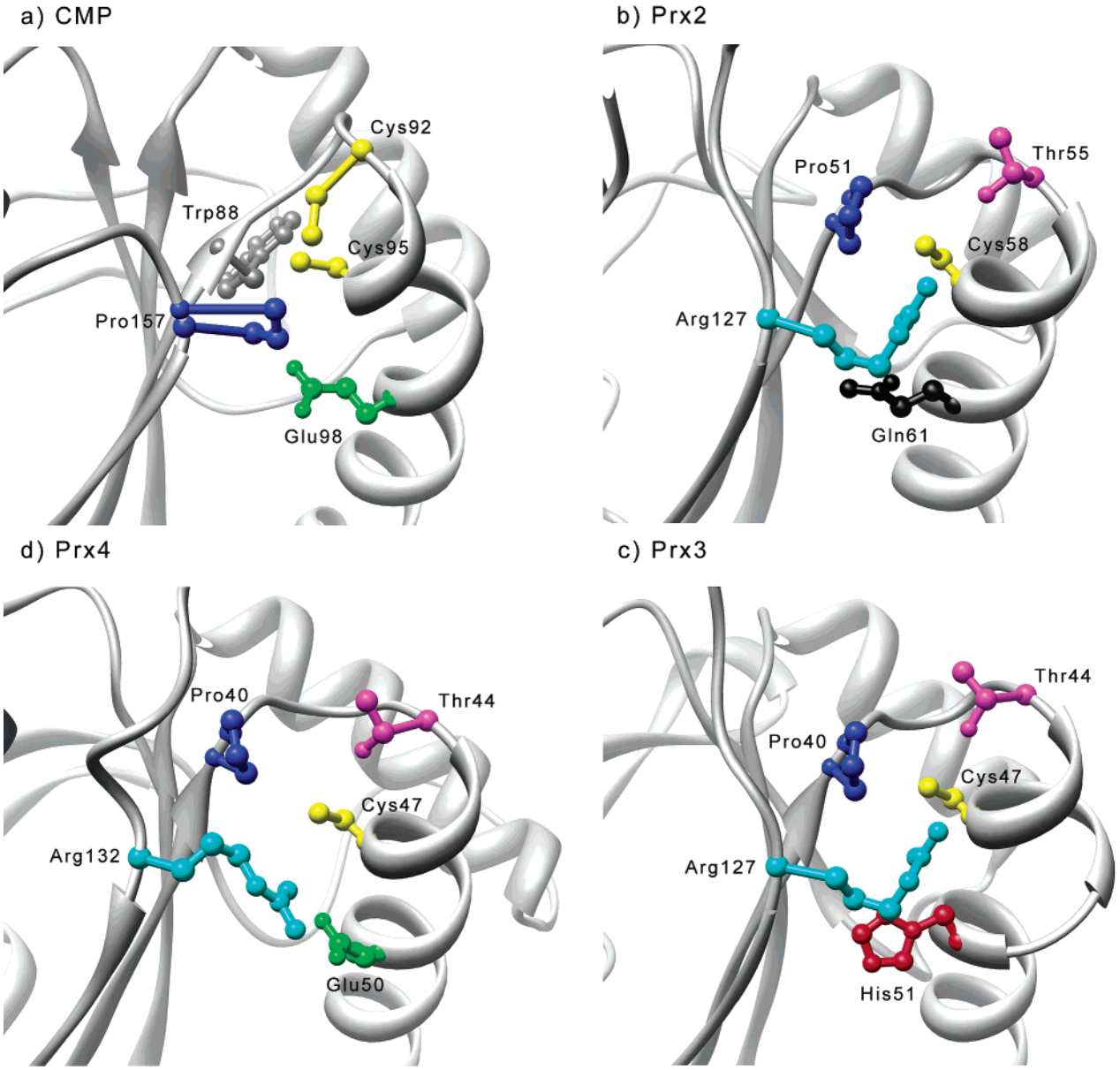


FIGURE 5: Close-up view of active sites of (a) a CMP, *B. japonicum* CcmG (PDB entry 1KNG), (b) a class 2 Prx, *S. pneumoniae* thiol peroxidase (PDB entry 1PSQ), (c) a class 3 Prx, human Prx5 (PDB entry 1HD2), and (d) a class 4 Prx, human Horf6 peroxidase (PDB entry 1PRX).

motif whose catalytically important residues cluster in the respective active sites, along with a structurally superimposable *cis*-Pro in the active site (Figure 3), and CMPs and Prxs share five strongly conserved motifs in the active sites and in the insertions in the canonical Trx fold (Figure 4). Thus, as described further below, the accumulated evidence provides strong support for an evolutionary relationship among Trxs, CMPs, and Prxs, and, most importantly, a record of sequence changes in the CMPs and Prxs that can be correlated with the appearance of new functional properties through successive modifications of the ancestral architecture. We note that when a motif analysis was carried out using a set of Trxs and Prxs alone, no motifs were found to be common to the Trx and Prx families.[4] Thus, understanding the relationship between Trxs and Prxs depended critically upon the identification of CMPs whose sequences link those of Trxs and Prxs, which have diverged beyond the point where sequence methods can relate them directly.

Motif Analysis Supports a Reclassification of Prxs. Our motif analysis and structural comparisons suggest a new system for classification of Prxs. Prxs have previously been classified according to quaternary structure, sequence characteristics around the active site, and/or the presence or absence of a resolving Cys. The most widely used classification system includes 1-Cys Prxs, typical 2-Cys Prxs, and atypical 2-Cys Prxs (*5*, *34*). (The resolving Cys in the typical 2-Cys Prxs is found in the C-terminal region of the other monomer in the dimeric protein, while in the atypical 2-Cys Prxs, the resolving Cys is located within the same monomer.) This scheme has recently been refined to include four classes of atypical 2-Cys Prxs, based upon the location of the resolving Cys (*35*).

[4] Since the divergence in the active sites of Prx2s prevents recognition of motif 4D in Figure 4a, we repeated the analysis using a set of Trxs and Prxs in which the Prx2s had been removed, and still found no common motifs.

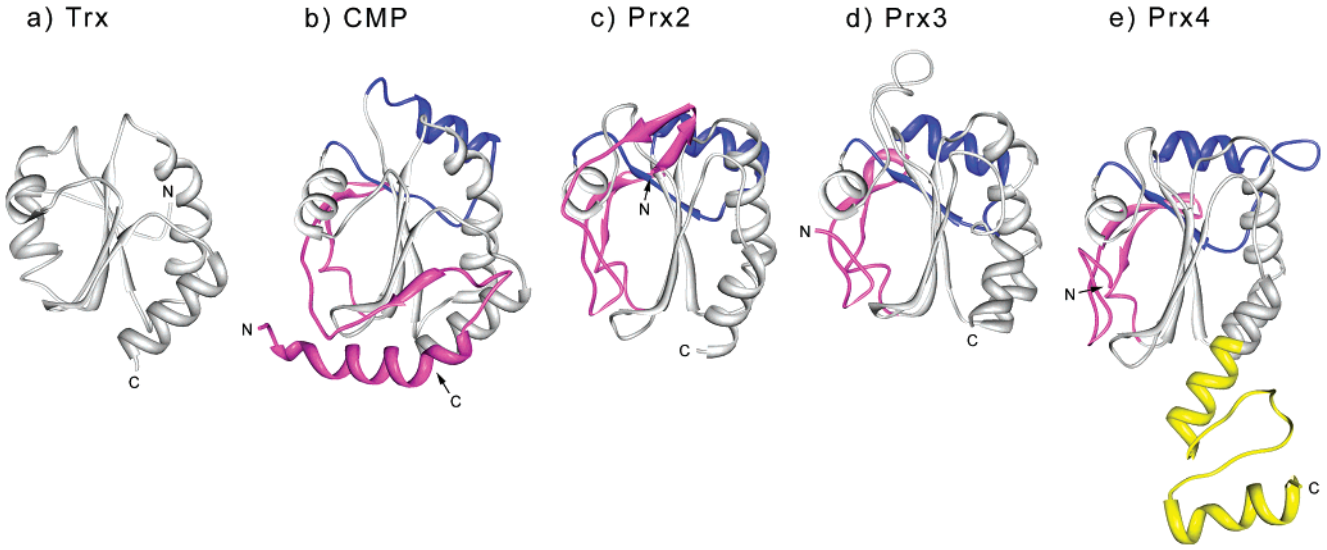


FIGURE 6: Structures of (a) a Trx (1XOA); (b) a CMP (*B. japonicum* TlpA, PDB entry 1JFU); (c) a class 2 Prx, *S. pneumoniae* thiol peroxidase (PDB entry 1PSQ); (d) a class 3 Prx, human Prx5 (PDB entry 1HD2); and (e) a class 4 Prx, human red blood cell thioredoxin peroxidase (PDB entry 1QMV), showing the canonical Trx fold (gray) and the positions of the N-terminal extension (magenta), the central insertion (blue), and the C-terminal extension found only in Prx4s (yellow). (Only monomers are shown for the three dimeric Prxs.)

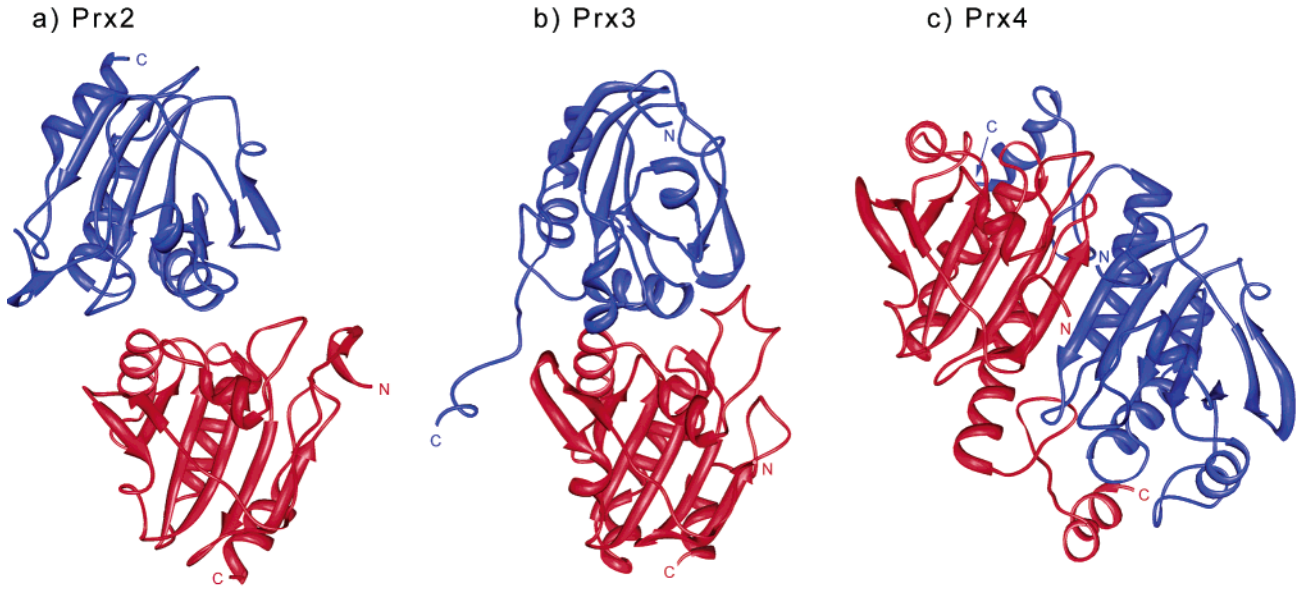


FIGURE 7: Quaternary structures of three Prxs: (a) a class 2 Prx, *S. pneumoniae* thiol peroxidase (PDB entry 1PSQ), (b) a class 3 Prx, human Prx5 (PDB entry 1OC3), and (c) a class 4 Prx, human red blood cell thioredoxin peroxidase (PDB entry 1QMV).

We recognize four distinct classes of Prxs (Table 1 and Figures 4−7). Class 1 includes Prxs that contain a highly conserved TPVCTKE motif at the active site. Many of these proteins were originally classified as bacterioferritin comigratory proteins, the name stemming from observations of electrophoretic mobility before their function was identified (*36*). Studies of these proteins in solution suggest that they are monomers (*36*). Class 2 includes a group of closely related bacterial proteins (with pairwise sequence identities of >50%) that have a Ser or Gln in place of the active site Glu of Prx1s and that form dimers by what we will term a "face-to-edge" interaction (Figure 7a). Class 3 includes Prxs that lack the conserved Glu but contain a conserved His downstream of the active site Cys in a characteristic TPTCSxxH motif. These proteins form dimers by a face-to-edge interaction between the two monomers [Figure 7b (*37*, *38*)] that is distinct from that typical of the class 2 Prxs, in terms of both structure and the nature of the residues at the interface. The active site regions of class 4 Prxs are similar to those of class 1 Prxs, but the class 4 Prxs are dimers in which the monomers interact via an "edge-to-edge" interaction between the β-sheets at the core of the molecule that results in an extended 12- or 14-stranded β-sheet spanning both monomers (Figure 7c). Some class 4 Prxs appear to form decamers when the protein is oxidized (*31*, *32*). Class 4 Prxs are also distinguished by a large C-terminal extension that extends into the active site of the opposite monomer. Within three of the four classes of Prxs we have defined, only some proteins contain a resolving Cys (Table 2). This finding suggests that acquisition of a resolving Cys is a refinement that has occurred in some lineages but not others. Furthermore, in each of the four classes, the resolving Cys is in a different location in the tertiary structure, suggesting that nature has invented four different solutions to the problem of regenerating the active site Cys. This is not surprising given the different quaternary structures of

Table 2: Resolving Cysteine Residues (Yellow) Are Found in Different Places in the Four Classes, and in Only Some Representatives of Classes 1, 3, and 4[a]

Prx1s– resolving Cys near active site Cys (in green)

Prx1.PrxSs	(32)	VVLAFYPAAFTSVCTKEMCTFRDSM
Prx1.PrxTv	(32)	VVLAFFPGAFTSVCTKEMCTFRDSM
Prx1.bfcmprotPa	(34)	VVLLFFPGAFTSVCTKELCTFRDKM
Prx1.bfcmprotBh	(30)	IVLYFYPKDMTPGCTTEACDFRDRV
Prx1.bfcmprotMt	(36)	VIVYFYPAASTPGCTKQACDFRDNL
Prx1.bfcmprotCt	(31)	VLLIFYPGDDTPVCTAQLCDYRNNV
Prx1.PrxhomSc	(94)	VVFFVYPRASTPGCTRQACGFRDNY
Prx1.bfcmprotAp	(31)	VVLYFYPRAMTPGCTREGVRFNELL
Prx1.bfcmprotCc	(56)	VVLYFFPAAYTAGCTAEAREFAEAT
Prx1.antioxidprotMm	(34)	VLLSFHPLAWTQVCAQQMKSLEENY
Prx1.antigenEh	(16)	VVLLFYPLDWTFVCPTEMIGYSEVA
Prx1.putantioxidprotDr	(32)	VVLVFYPLDFSPVCSMQLPEYSGSQ
Prx1.hyrAHalo	(67)	VVLFFYPFDFSPVCATELCAIQNAR
Prx1.bfcmprotTm	(26)	TILFFFPKAGTSGCTREAVEFSREN
Prx1.ahpEMt	(32)	VLLVFFPLAFTGICQGELDQLRDHL
Prx1.PrxLl	(45)	VLISVFPDINTRVCSLQTKHFNLEA

Prx2s– resolving Cys in central insertion

Prx2.PrxEc	(80)	VLCISADLPFAQSRFCGAEGLNNVITLSTFRN
Prx2.PrxAa	(80)	VTVVSMDLPFAQKRFCESFNIQNVTVASDFRY
Prx2.simPrxSa	(78)	VLTISADLPFAQKRWCASAGLDNVITLSDHRD
Prx2.TpxSp	(77)	VLTVSXDLPFAQKRWCGAEGLDNAIXLSDYFD
Prx2.hypMm	(79)	VLCVSADLPFAHKRFCGAEGLERVQSLSDMRD
Prx2.tagDVc	(78)	LLCVSADLPFAMSRFCTEHAVANVTNASFFRE

Prx3s – resolving Cys in loop between last beta strand and alpha helix of the Trx fold

Prx3.PrxXc	(147)	KVSAADYVLQHL
Prx3.peroxisomalprotPc	(156)	ELSRADHVLKQL
Prx3.PrxfamilyNostoc	(152)	EVSDAQTMLKHI
Prx3.PrxSc	(165)	TVSSVESVLAHL
Prx3.PrxAt	(151)	TVSSAEDILKAL
Prx3.Prx5Hs	(150)	TCSLAPNIISQL
Prx3.hypPrxSm	(150)	TASGAAAMLELL
Prx3.peroxisomalprotLk	(155)	AGSGVDAVLAAL

Prx4s – resolving Cys in C-terminal extension of other monomer

Prx4.PrxLm	(166)	GGLCPINWQPGEKTI
Prx4.rbcPrxHs	(168)	GEVCPAGWKPGSDTI
Prx4.ahpcSt	(162)	GEVCPAKWKEGEATL
Prx4.PrxTv	(158)	GGLCPVNWHEGEPTL
Prx4.PrxStc	(161)	DELCPCNWSKGDETL
Prx4.OrfOs	(169)	GYACPVNWNFGDQVF
Prx4.probPrxSe	(168)	GDVCPAQWEKGKEGM
Prx4.PrxPf	(171)	PVATPVNWKEGDKCC
Prx4.SecysPrxEa	(167)	AEVAPSGWKPGKKTL
Prx4.Horf6Hs	(174)	RVATPVDWKDGDSVM
Prx4.PrxEc	(163)	GDICPAEWRSENKDN
Prx4.probPrxMt	(161)	GVAAPANWPDNQLIG
Prx4.PrxCc	(159)	NKLCPYSWNNDSKSI

[a] The role of the putative resolving Cys residues has not been experimentally verified in all cases.

the four classes.

We suggest that this classification scheme, which integrates information from motifs found throughout the protein structures, as well as structural information, is an improvement upon previously proposed schemes. We note that the phylogenetic analyses of many Prx sequences carried out by Verdoucq et al. (*39*) and Hoffman et al. (*12*) suggest similar groupings of Prxs.[5]

Differences in Sequence Correlate with Differences in Function between Trxs and CMPs. Trxs and CMPs carry out comparable thiol oxidoreductase reactions, but differ substantially in substrate specificity. These classes are distinguished by the presence of two major insertions in CMPs (Table 1 and Figure 6), as well as variations in sequence in the region of the active site cysteines and *cis*-Pro, the likely result of which is a relatively more defined

[5] We note that the robustness of the published phylogenetic trees cannot be assessed because neither study reports bootstrap values, and Hoffman et al. do not specify the method or sequences used to generate the tree.

substrate binding site in the CMPs. The following discussion focuses upon residues that are conserved in both Trxs and CMPs or whose divergence can be rationalized in terms of diversification of function in the two groups of proteins.

The most important motif found in the set of Trxs and CMPs includes the active site Cys residues. This motif has long been recognized in association with the fundamental thiol oxidoreductase reaction carried out by Trxs and CMPs. A notable feature peculiar to the CMPs is a conserved Glu three residues after the CXXC (Figures 2 and 3). This Glu (Glu98 in the CMP in Figure 3), which is partially buried in the CMP structure, has been proposed to be involved in recognition of the specific proteins bound by CcmG (*25*). An additional distinctive feature is the conserved Asp (Asp26 in the Trx in Figure 3) preceding by several residues the CXXC in Trxs that is substituted with Asn (Asn86 in Figure 3) in CMPs. In Trxs, this Asp removes the proton from the C-terminal Cys of the active site CXXC as it attacks the mixed disulfide between the N-terminal Cys and the target protein (*40*). The Asn in CMPs that replaces this Asp is located within hydrogen bonding distance of the conserved Glu found only in CMPs (Glu98 in Figure 3). One explanation for this difference is the need to prevent repulsion between two closely spaced negatively charged residues that would occur in the CMPs if the Asp had not been replaced with Asn. Notably, the D26N mutant of *E. coli* Trx is only modestly impaired; the rate of cleavage of the mixed disulfide is decreased by only 5−10-fold (*40*). Thus, the change from Asp to Asn is compatible with retention of the ability to carry out thiol−disulfide exchange reactions.

The active sites of both Trxs and CMPs contain a completely conserved *cis*-Pro residue; these residues superimpose nearly exactly, as shown in Figure 3. The *cis* conformation of the Pro exposes the main chain oxygen of the preceding residue. Structures of Trxs with covalently attached polypeptide substrates show that this oxygen provides an important interaction site for the extended polypeptide chain of the substrate (*29*, *30*). In addition, the side chain of the Pro residue is in contact with the disulfide bridge formed between the active site and the target polypeptide, and may help position the disulfide for subsequent attack by the C-terminal Cys residue (*29*, *30*). While this interaction is also likely to be important in CMPs, there is enough sequence divergence in this region overall that MEME recognizes distinct motifs for the two families. As discussed above, this sequence divergence may be related to both the enhanced substrate specificity displayed by the CMPs and the likely ancient divergence of the Trxs and CMPs.

The insertions found in CMPs also appear to provide surface features that allow recognition of particular binding partners required for their specialized functions. The blue motif (2C) shown in Figure 2 encompasses the entire central insertion located between β2 and α2 of the canonical fold. In the structurally characterized CMPs (CcmG and TlpA), this insertion provides a surface-exposed α-helix and short β-sheet (*25*, *26*); the α-helix flanks the active site and is required for cytochrome *c* formation in the case of CcmG and thus is believed to be required for binding of the specific protein target of CcmG (*25*). The magenta motif (2A) is found near the end of the N-terminal extension in most CMPs. This region is situated on the side of the protein opposite from the active site, so is not likely to be involved in catalysis. It has been proposed that the several proteins involved in maturation of cytochrome *c* may form a "maturase" complex (*41*); it is possible that the region containing the magenta motif is conserved because it is involved in interactions with other proteins in such a complex. Curiously, the N-terminal insertion adopts different secondary structures in the two structurally characterized CMPs (not shown). The magenta motif (2A) is found in TlpA and not in CcmG, so the structure found in CcmG may be a less common variation in a rather plastic region of the structure.

Differences in Sequence Correlate with Differences in Function between CMPs and Prxs. Examination of the motifs found by MEME to be globally conserved in the set of CMPs and Prxs reveals class-specific residues characteristic of either CMPs or Prxs that may be involved in the particular functions carried out by each of these classes of proteins. The following discussion focuses on those differences in sequence that are likely to contribute to the functional differences between CMPs and Prxs.

The most strongly conserved motif (4C, cyan in Figure 4) contains the active site cysteine residue (in Prxs) or residues (in CMPs). A particularly notable feature is the change of the N-terminal Cys of the CXXC in CMPs to a Thr in the Prxs. This change has the effect of switching the position at which the protein attacks its substrate from the N-terminal Cys in CMPs to the single Cys that aligns with the C-terminal Cys of CMPs. Structural evidence suggests a hydrogen bonding interaction between the active site Cys and the hydroxyl group of this Thr (Figure 5) that may enhance the nucleophilicity of the Cys and/or help position the Cys for nucleophilic attack on the peroxide substrate (*5*, *42*). Indeed, mutation of this Thr to Ser or Val significantly decreases catalytic activity (*42*, *43*). As expected, the CXXC/TXXC sequences align in the structures as well (Figure 5).

An additional striking change is the replacement of a Trp preceding the CXXC in CMPs with a Pro in Prxs (Figures 4a and 5). Note that in Figure 5 these residues occupy identical positions in the active site. In Prxs, this Pro is believed to shield the active site Cys and help prevent overoxidation of the sulfenic acid intermediate formed during the catalytic cycle to the sulfinic or sulfonic acid forms, which cannot be regenerated by the usual pathway shown in Figure 1 (*5*). Although protecting the sulfur atom of the sulfenic acid intermediate is important, ultimately the sulfur atom must be exposed to allow attack by a thiol to form the disulfide in the next step of the reaction. This process requires unwinding of the end of the helix containing the active site Cys to expose the sulfur atom that is buried by the hydroxyl group in the sulfenic acid intermediate (*5*). It is notable that most CMPs in the set have a Pro immediately before the C-terminal Cys of the CXXC. Pro is never found in this position in Prxs, and this change may be important in allowing the flexibility around the Cys residue that is required to allow attack of the resolving Cys on the sulfenic acid intermediate.

The Glu residue following the CXXC in CMPs is also conserved in class 1 and class 4 Prxs (Figures 4a and 5; Glu98 in CMP, Glu50 in Prx4). This conservation is especially interesting because this residue likely plays different roles in the two protein superfamilies. As discussed

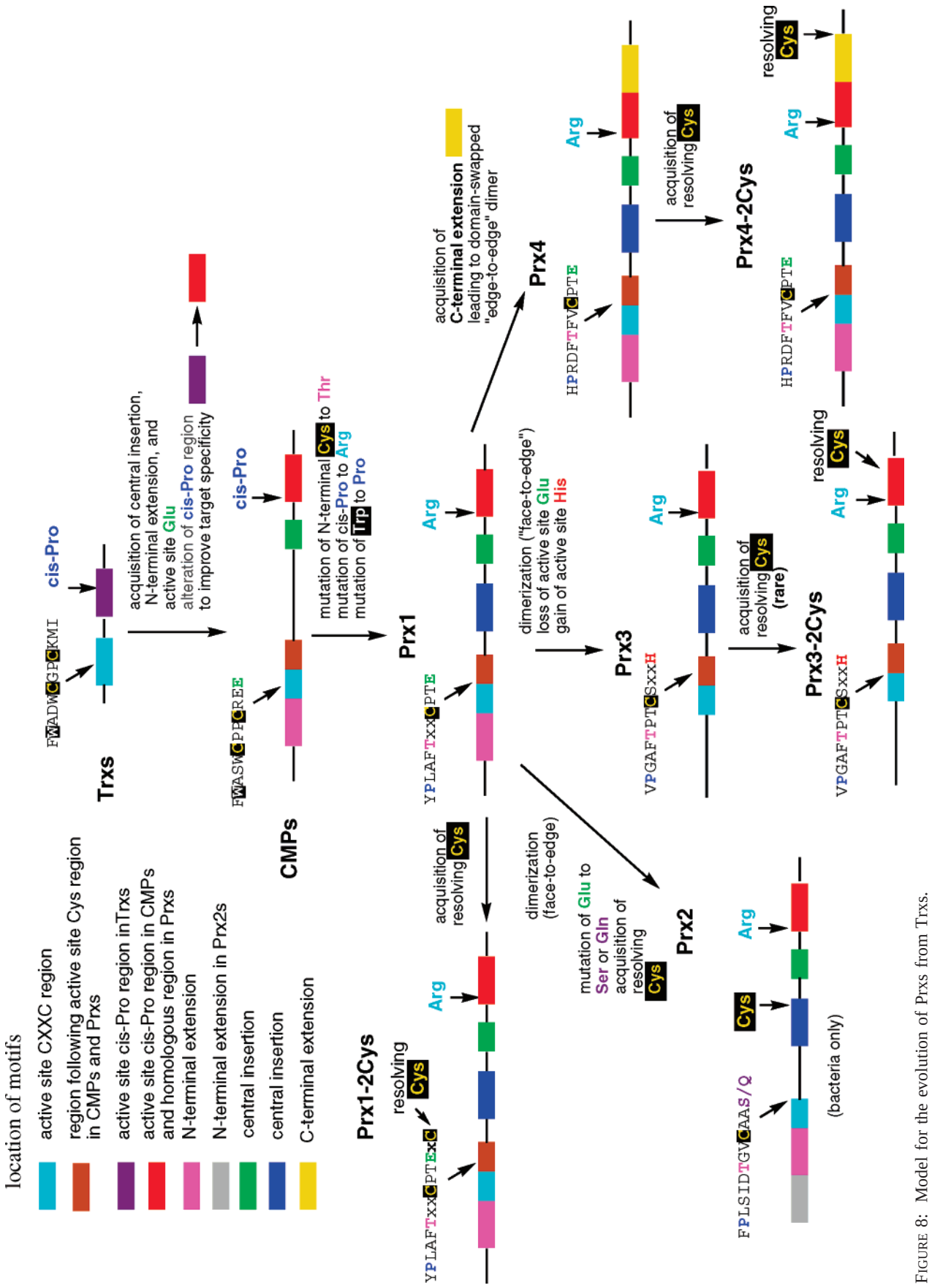


FIGURE 8: Model for the evolution of Prxs from Trxs.

above, it is likely to be involved in substrate recognition in the CMPs. However, in Prxs, this Glu interacts with a conserved Arg (Arg132 in Prx4 in Figure 5) that replaces the conserved *cis*-Pro found in Trxs and CMPs [see the red motif (4G) in Figure 4a and Pro157 in the CMP structure in Figure 5]. The *cis*-Pro in Trxs and CMPs helps to form the binding platform for polypeptide substrates. Since the substrates for Prxs are peroxides rather than polypeptides, the *cis*-Pro is no longer needed. The Arg that appears in the homologous position of Prxs is postulated to interact with the active site Cys and influence its nucleophilicity. Indeed, mutation of this Arg abolishes or diminishes (*43*) Prx activity. Thus, the conserved Glu in class 1 and class 4 Prxs appears to interact with and orient the Arg, and thereby help to tune the nucleophilicity of the active site Cys, rather than serving as a recognition element for binding of a specific substrate, as it does in CMPs. Notably, class 2 Prxs have a conserved Ser/Gln and class 3 Prxs a conserved His rather than a Glu in the active site. These differences result in subtle variations in the active site architecture (Figure 5). In Prxs with an active site His or Gln, the conserved Arg appears to be in direct contact with the active site Cys, whereas it is ~1 Å farther away in the Prxs that have a conserved Glu. The position of the Arg may affect the pK_a of the nucleophilic Cys; this may be important in terms of the functions of these proteins, but there are no experimental data as yet to address this issue.

In general, the structural superpositions given in Figures 3 and 5 are consistent with the sequence alignments given in Figures 2 and 4a. This analysis suggests that nature has reused specific sites in the active site architecture common to all of these proteins by changing the specific nature of certain residues to accommodate new mechanisms and functions. Such conservation of position in active site residues has been observed in other superfamilies (*44*−*46*).

A Model for the Evolution of Prxs from Trxs. By providing explicit correlations between conserved elements of sequence and structure and conserved aspects of the chemical mechanism in each of the Trx, CMP, and Prx classes, we can propose a model for the evolutionary divergence of Prxs from a Trx-like ancestor (see Figure 8). Such a model is useful in thinking about how nature has engineered new capabilities into an existing structural fold. The model proposes that Trxs are the most ancient among the Trxs, CMPs, and Prxs. Trxs are among the most ubiquitous proteins in nature; they are represented in every lineage in the Clusters of Orthologous Groups (COGs) Database (http://www.ncbi.nlm.nih.gov/COG/) (*47*, *48*). Most proteins that are represented in every lineage in the COGs database are involved in transcription and translation. These ubiquitous proteins are considered likely to have been present in the last universal common ancestor (*47*, *49*). In contrast, cytochromes and cytochrome biogenesis proteins are more sparsely represented in the COGs database, suggesting that they arose later in specific lineages. Thus, we hypothesize that substrate-specific thiol−disulfide oxidoreductases involved in cytochrome maturation arose by specialization of previously existing Trx-like proteins of broader specificity. We also propose that Prxs arose from CMPs after the emergence of oxygenic photosynthesis ~2.3 billion years ago (*50*).

The active site of Trxs is well suited for reduction of disulfide bonds in target proteins because it is exposed at the surface and allows binding of an extended β-sheet in the target protein. Our model postulates that CMPs may have evolved from a remote Trx ancestor by acquiring new structural elements that allowed more specific recognition of target proteins. These changes included acquisition of an N-terminal extension and an insertion between β2 and α2 of the Trx fold that form surface features that contribute to recognition of specific target proteins, as well as point mutations leading to changes in substrate specificity (e.g., the conserved Glu following the CXXC).

The evolution of Prxs apparently required several changes in specific residues that allowed a shift in the chemistry from the relatively easy thiol−disulfide exchange reaction to the more difficult reduction of peroxides. Movement of the site of nucleophilic attack to deeper within the protein by mutation of the N-terminal Cys of the CXXC to Thr and utilization of the C-terminal Cys as the nucleophile could potentially allow better control over substrate orientation and more extensive provision of catalytic residues to assist in the more difficult cleavage of the O−O bond. The Thr that replaces the N-terminal Cys and the Arg that replaces the *cis*-Pro of CMPs both interact with the active site Cys and are critical for Prx activity. The replacement of the Trp preceding the active site Cys with Pro is also important for protecting the reactive sulfenic acid intermediate from overoxidation before it is able to react with a resolving Cys in the same molecule or a different molecule.

The relatively large number of differences in primary, secondary, tertiary, and quaternary structure that distinguish the different classes of Prxs described in our model reflect a pathway for diversification of this class of proteins more complex than the one we found among the currently available representatives of the Trxs or CMPs. We postulate in Figure 8 that the monomeric class 1 Prxs may have been the first to arise from the CMPs. This hypothesis is based upon the observation that the class 1 Prxs are most similar to the CMPs in terms of quaternary structure, active site residues, patterns of conserved sequence motifs, and lack of a C-terminal extension. Furthermore, the Prxs that are most similar to CMPs according to our Shotgun analysis are class 1 Prxs. Class 2−4 enzymes could then have arisen by acquisition of sequence changes associated with different modes of dimerization, followed by adaptations optimizing catalysis within those new structural contexts. As noted earlier, a resolving Cys apparently arose independently and in different locations in some lineages within each class (Table 2).

CONCLUSIONS

Catalysis of different chemical reactions by evolutionarily related proteins has been observed in several fold classes (*45*, *51*−*55*). In many cases, such enzymes preserve a similar partial reaction or chemical capability; this notion is supported by recent large-scale studies across the enzyme universe (*56*, *57*). For example, enzymes in the enolase superfamily preserve the ability to remove a proton from a carbon α to a carboxylate, although the subsequent chemical steps and thus the overall chemical transformations differ substantially (*44*). Enzymes in the crotonase superfamily preserve the ability to stabilize oxyanion intermediates formed in different reactions of coenzyme A thioesters (*58*), and enzymes in the amidohydrolase superfamily utilize a

mononuclear or binuclear metal center to activate a hydrolytic water molecule for nucleophilic attack (*55*). As illustrated by these examples, partial reactions (or other chemical capabilities) conserved within mechanistically diverse superfamilies can be explicitly mapped to similarities in the active sites, thereby defining a structure−function paradigm that unifies the diverse members of a superfamily (*54*). However, some divergently related enzymes do not share a common mechanistic strategy, even though some active site residues may be conserved. The term "suprafamily" has been used to describe such groups of enzymes (*54*, *59*). Given these definitions, Trxs and Prxs are most properly considered as members of a suprafamily that we will term the Trx fold suprafamily. The reduction of disulfide bonds and the reduction of peroxides are similar in some respects. In both cases, the reaction is initiated by attack of an active site Cys on the substrate, and completed by attack of a second Cys on the resulting adduct, followed by reduction of the resulting disulfide. However, the position of nucleophilic attack and the position of the second Cys differ in Trxs and Prxs, as do the residues involved in binding substrates and protecting the sensitive sulfenic acid intermediate in the case of Prxs. Indeed, only one residue in the active site (the Cys corresponding to the C-terminal Cys of the Trx CXXC) is conserved between these two protein superfamilies, and it serves a different role in each.

The analysis presented here provides important insights into the variations on the fundamental themes conserved in the structures and active sites of Trx fold proteins, and leads to a model for the emergence of the new CMP and Prx structures and catalytic capabilities of CMPs and Prxs from an ancestral Trx-like protein. Work is ongoing to expand the model to include other Trx fold superfamilies.

REFERENCES

1. Babbitt, P. C., and Gerlt, J. A. (1997) Understanding enzyme superfamilies: Chemistry as the fundamental determinant in the evolution of new catalytic activities, *J. Biol. Chem. 272*, 30591−30594.
2. Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M., and Creighton, T. E. (1997) The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules, *Curr. Biol. 7*, 239−245.
3. Martin, J. L. (1995) Thioredoxin: a fold for all reasons, *Structure 3*, 245−250.
4. Wang, S., Trumble, W. R., Liao, H., Wesson, C. R., Dunker, A. K., and Kang, C. H. (1998) Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum, *Nat. Struct. Biol. 5*, 476−483.
5. Wood, Z. A., Schröder, E., Harris, J. R., and Poole, L. B. (2003) Structure, mechanism and regulation of peroxiredoxins, *Trends Biochem. Sci. 28*, 32−40.
6. Follman, H., and Häberlein, I. (1995) Thioredoxins: universal, yet specific thiol-disulfide redox cofactors, *BioFactors 5*, 147−156.
7. Powis, G., and Montfort, W. R. (2001) Properties and biological activities of thioredoxins, *Annu. Rev. Biophys. Biomol. Struct. 30*, 421−455.
8. Nishiyama, A., Masutani, H., Nakamura, H., Nishinaka, Y., and Yodoi, J. (2001) Redox regulation by thioredoxin and thioredoxin-binding proteins, *IUBMB Life 52*, 29−33.
9. Arnér, S. J., and Holmgren, A. (2000) Physiological functions of thioredoxin and thioredoxin reductase, *Eur. J. Biochem. 267*, 6102−6109.
10. Russel, M., and Model, P. (1986) The role of thioredoxin in filamentous phage assembly. Construction, isolation, and characterization of mutant thioredoxins, *J. Biol. Chem. 261*, 14997−15005.
11. Huber, H. E., Russel, M., Model, P., and Richardson, C. C. (1986) Interaction of mutant thioredoxin of *Escherichia coli* with the gene 5 protein of phage T7: The redox capacity of thioredoxin is not required for stimulation of DNA polymerase activity, *J. Biol. Chem. 261*, 15006−15012.
12. Hofmann, B., Hecht, H.-J., and Flohé, L. (2002) Peroxiredoxins, *Biol. Chem. 383*, 347−364.
13. Neumann, C. A., Krause, D. S., Carman, C. V., Das, S., Dubey, D. P., Abraham, J. L., Bronson, R. T., Fujiwara, Y., Orkin, S. H., and Van Etten, R. A. (2003) Essential role for the peroxiredoxin Prdx1 in erythrocyte antioxidant defense and tumour suppression, *Nature 424*, 561−565.
14. Komaki-Yasuda, K., Kawazu, S.-I., and Kano, S. (2003) Disruption of the *Plasmodium falciparum* 2-Cys peroxiredoxin gene renders parasites hypersensitive to reactive oxygen and nitrogen species, *FEBS Lett. 547*, 140−144.
15. Wang, X., Phelan, S. A., Forsman-Semb, K., Taylor, E. F., Petros, C., Brown, A., Lerner, C. P., and Paigen, B. (2003) Mice with targeted mutation of peroxiredoxin 6 develop normally but are susceptible to oxidative stress, *J. Biol. Chem. 276*, 25179−25190.
16. Wood, Z. A., Poole, L. B., and Karplus, P. A. (2003) Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling, *Science 300*, 650−653.
17. Choi, H.-J., Kang, S. W., Yang, C.-H., Rhee, S. G., and Ryu, S.-E. (1998) Crystal structure of a novel human peroxidase enzyme at 2.0 Å resolution, *Nat. Struct. Biol. 5*, 400−406.
18. Schröder, E., and Ponting, C. P. (1998) Evidence that peroxiredoxins are novel members of the thioredoxin fold superfamily, *Protein Sci. 7*, 2465−2468.
19. Gerstein, M. (1998) Measurement of the effectiveness of transitive sequence comparison through a third 'intermediate' sequence, *Bioinformatics 14*, 707−714.
20. Pegg, S. C.-H., and Babbitt, P. C. (1999) Shotgun: Getting more from sequence similarity searches, *Bioinformatics 15*, 729−740.
21. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.
22. Bailey, T. L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers, in *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp 28−36, AAAI Press, Menlo Park, CA.
23. Jewett, A. I., Huang, c. C., and Ferrin, T. E. (2003) MINRMS: An efficient algorithm for determining protein structure similarity using root-mean-squared distances, *Bioinformatics 19*, 625−634.
24. Huang, C. C., Couch, G. S., Pettersen, E. F., and Ferrin, T. E. (1996) Chimera: An extensible molecular modeling application constructed using standard components, *Pac. Symp. Biocomput. '96 1*, 724.
25. Edeling, M. A., Guddat, L. W., Fabianek, R. A., Thöny-Meyer, L., and Martin, J. L. (2002) Structure of CcmG/DsbE at 1.14 Å resolution: high-fidelity reducing activity in an indiscriminately oxidizing environment, *Structure 10*, 973−979.
26. Capitani, G., Rossman, R., Sargent, D. F., Grütter, M. G., Richmond, T. J., and Hennecke, H. (2001) Structure of the soluble domain of a membrane-anchored thioredoxin-like protein from *Bradyrhizobium japonicum* reveals unusual properties, *J. Mol. Biol. 311*, 1037−1048.
27. Fabianek, R. A., Hofer, T., and Thöny-Meyer, L. (1999) Characterization of the *Escherichia coli* CcmH protein reveals new insights into the redox pathway required for cytochrome *c* maturation, *Arch. Microbiol. 171*, 92−100.
28. Loferer, H., Bott, M., and Hennecke, H. (1993) *Bradyrhizobium japonicum* TlpA, a novel membrane-anchored thioredoxin-like protein involved in the biogenesis of cytochrome aa_3 and development of symbiosis, *EMBO J. 12*, 3373−3383.
29. Qin, J., Clore, G. M., Kennedy, W. M. P., Huth, J. R., and Gronenborn, A. M. (1995) Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide form the transcription factor NFκB, *Structure 3*, 289−297.
30. Qin, J., Clore, G. M., Kennedy, W. P., Kuszewski, J., and Gronenborn, A. M. (1996) The solution structure of human thioredoxin complexed with its target from Ref-1 reveals peptide chain reversal, *Structure 4*, 613−620.

31. Wood, Z. A., Poole, L. B., Hantgan, R. R., and Karplus, P. A. (2002) Dimers to doughnuts: redox-sensitive oligomerization of 2-cysteine peroxiredoxins, *Biochemistry 41*, 5493−5504.
32. Schröder, E., Littlechild, J. A., Lebedev, A. A., Errington, N., Vagin, A. A., and Isupov, M. N. (2000) Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution, *Structure 8*, 605−615.
33. Trivelli, X., Krimm, I., Ebel, C., Verdoucq, L., Prouzet-Mauléon, V., Chartier, Y., Tsan, P., Lauquin, G., Meyer, Y., and Lancelin, J.-M. (2003) Characterization of the yeast peroxiredoxin Ahp1 in its reduced active and overoxidized inactive forms using NMR, *Biochemistry 42*, 14139−14149.
34. Rhee, G. S., Kang, S. W., Chang, T.-S., Jeong, W., and Kim, K. (2001) Peroxiredoxin, a novel family of peroxidases, *IUBMB Life 52*, 35−41.
35. Choi, J., Choi, S., Choi, J., Cha, M.-K., Kim, I.-H., and Shin, W. (2003) Crystal structure of *Escherichia coli* thiol peroxidase in the oxidized state: insights into intramolecular disulfide formation and substrate binding in atypical 2-Cys peroxiredoxins, *J. Biol. Chem. 278*, 49478−49486.
36. Jeong, W., Cha, M.-K., and Kim, I.-H. (2000) Thioredoxin-dependent hydroperoxide peroxidase activity of bacterioferritin comigratory protein (BCP) as a new member of the thiol-specific antioxidant protein (TSA)/alkyl hydroperoxide peroxidase C (AhpC) family, *J. Biol. Chem. 27*5, 2924−2930.
37. Kim, S. J., Woo, J. R., Hwang, Y. S., Jeong, D. G., Shin, D. H., Kim, K., and Ryu, S. E. (2003) The tetrameric structure of *Haemophilus influenza* hybrid Prx5 reveals interactions between electron donor and acceptor proteins, *J. Biol. Chem. 278*, 10790−10798.
38. Evrard, C., Capron, A., Marchand, C., Clippe, A., Wattiez, R., Soumillion, P., Knoops, B., and Declerq, J.-P. (2004) Crystal structure of a dimeric oxidized form of human peroxiredoxin 5, *J. Mol. Biol. 337*, 1079−1090.
39. 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.
40. Chivers, P. T., and Raines, R. T. (1997) General acid/base catalysis in the active site of *Escherichia coli* thioredoxin, *Biochemistry 36*, 15810−15816.
41. Thöny-Meyer, L. (1997) Biogenesis of respiratory cyochromes in bacteria, *Microbiol. Mol. Biol. Rev. 61*, 337−376.
42. Flohé, L., Budde, H., Bruns, K., Castro, H., Clos, H., 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.
43. König, J., Lotte, K., Plessow, R., Brockhinke, A., Baier, M., and Dietz, K.-J. (2003) Reaction mechanism of plant 2-Cys peroxiredoxin: Role of the C terminus and the quaternary structure, *J. Biol. Chem. 278*, 24409−24420.
44. Babbitt, P. C., Hasson, M., Wedekind, J. E., Palmer, D. J., Lies, M. A., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gerlt, J. A. (1996) The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids, *Biochemistry 35*, 16489−16501.
45. Holm, L., and Sander, C. (1997) An evolutionary treasure: unification of a broad set of amidohydrolases related to urease, *Proteins: Struct., Funct., Genet. 28*, 72−82.
46. Bartlett, G. J., Borkakoti, N., and Thornton, J. M. (2003) Catalysing new reactions during evolution: economy of residues and mechanism, *J. Mol. Biol. 331*, 829−860.
47. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families, *Science 278*, 631−637.
48. Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D., and Koonin, E. V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res. 29*, 22−28.
49. Koonin, E. (2003) Comparative genomics, minimal gene sets, and the last universal common ancestor, *Nat. Rev. Microbiol. 1*, 127−136.
50. Pavlov, A. A., and Kasting, J. F. (2002) Mass-independent fractionation of sulfur isotopes in Archean sediments: strong evidence for an anoxic Archean atmosphere, *Astrobiology 2*, 27−41.
51. Holden, H., Benning, M. M., Haller, T., and Gerlt, J. A. (2001) The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme A thioesters, *Acc. Chem. Res. 34*, 145−157.
52. Koonin, E., and Tatusov, R. L. (1994) Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database searching, *J. Mol. Biol. 244*, 125−132.
53. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2000) The crystal structure of *Bacillus cereus* phosphonoacetaldehyde hydrolase: Insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily, *Biochemistry 39*, 10385−10396.
54. Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies, *Annu. Rev. Biochem. 70*, 209−246.
55. Gerlt, J. A., and Raushel, F. M. (2003) Evolution of function in $(\beta/\alpha)_8$-barrel enzymes, *Curr. Opin. Chem. Biol. 36*, 539−548.
56. Nagano, N. C., Orengo, C. A., and Thornton, J. M. (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures, and functions, *J. Mol. Biol. 321*, 741−765.
57. Todd, A. E., Orengo, C. A., and Thornton, J. M. (2001) Evolution of function in protein superfamilies, from a structural perspective, *J. Mol. Biol. 307*, 1113−1143.
58. Holden, H. M., Benning, M. M., Haller, T., and Gerlt, J. A. (2001) The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme A thioesters, *Acc. Chem. Res. 34*, 145−157.
59. Wise, E., Yew, W. S., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (2002) Homologous $(\beta/\alpha)_8$-barrel enzymes that catalyze unrelated reactions: orotidine 5′-monophosphate decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase, *Biochemistry 41*, 3861−3869.

BI048947R