Research article

New insights on thioredoxins (Trxs) and glutaredoxins (Grxs) by *in silico* amino acid sequence, phylogenetic and comparative structural analyses in organisms of three domains of life

Soumila Mondal, Shailendra P. Singh [*]

Centre of Advanced Study in Botany, Institute of Science, Banaras Hindu University, Varanasi-221005, UP, India

ABSTRACT

Thioredoxins (Trxs) and Glutaredoxins (Grxs) regulate several cellular processes by controlling the redox state of their target proteins. Trxs and Grxs belong to thioredoxin superfamily and possess characteristic Trx/Grx fold. Several phylogenetic, biochemical and structural studies have contributed to our overall understanding of Trxs and Grxs. However, comparative study of closely related Trxs and Grxs in organisms of all domains of life was missing. Here, we conducted *in silico* comparative structural analysis combined with amino acid sequence and phylogenetic analyses of 65 Trxs and 88 Grxs from 12 organisms of three domains of life to get insights into evolutionary and structural relationship of two proteins. Outcomes suggested that despite diversity in their amino acids composition in distantly related organisms, both Trxs and Grxs strictly conserved functionally and structurally important residues. Also, position of these residues was highly conserved in all studied Trxs and Grxs. Notably, if any substitution occurred during evolution, preference was given to amino acids having similar chemical properties. Trxs and Grxs were found more different in eukaryotes than prokaryotes due to altered helical conformation. The surface of Trxs was negatively charged, while Grxs surface was positively charged, however, the active site was constituted by uncharged amino acids in both proteins. Also, phylogenetic analysis of Trxs and Grxs in three domains of life supported endosymbiotic origins of chloroplast and mitochondria, and suggested their usefulness in molecular systematics. We also report previously unknown catalytic motifs of two proteins, and discuss in detail about effect of abovementioned parameters on overall structural and functional diversity of Trxs and Grxs.

1. Introduction

Thioredoxins (Trxs) and glutaredoxins (Grxs) are heat stable, small (~9–16 kDa) redox-controlling thiol-disulphide oxidoreductases that share di-cysteine active site motif (CXXC) and a common Trx/Grx fold [1, 2]. Trxs and Grxs are found in all domains of life where two proteins are responsible for maintaining cellular redox homeostasis [1, 2, 3, 4]. Trx/Grx fold is characterized by the presence of four β strands and three flanking α helices. The β strands are oriented in a 4312 fashion where 3rd strand is antiparallel to the rest of the β strands [2, 5]. Although Trxs and Grxs belong to the same superfamily and share Trx/Grx fold, the two proteins differ in their source of reducing power. Trxs are reduced by thioredoxin reductase (TR) in an NADPH-dependent reaction while reduced glutathione (GSH) acts as a source of reducing equivalents for Grxs [1, 2, 3]. After catalyzing reduction of their substrates, oxidized Grxs are reduced by GSH which results in the generation of oxidized glutathione (GSSG). GSSG is reduced by NADPH-dependent glutathione reductase (GR) to give GSH. Together, Grxs, GSH, GSSG, and NADPH-dependent GR constitute the glutaredoxin system [4].

Trx was first discovered in *Escherichia coli* (*E. coli*) as an electron donor for the reduction of ribonucleotide reductase (RNR) enzyme [6, 7]. Later, Grx was identified as a backup system of Trx in *E. coli* for the reduction of the RNR enzyme [8]. However, subsequent studies in different organisms established importance of Trxs and Grxs in development, protection of proteins from oxidative damage, signal transduction, protein folding, photosynthesis, abiotic stress resulting reactive oxygen species (ROS), programmed cell death (PCD), cardiac, neurodegenerative and cancerous diseases [1, 2, 9, 10, 11,12]. Thus, Trxs and Grxs regulate diverse range of cellular functions and affect the overall fitness and development of different organisms by controlling the redox state of their target proteins.

Different eukaryotic organisms possess Trxs and Grxs that are targeted to different cellular compartments. However, all Trxs essentially retain catalytic CGPC motif and Trx/Grx fold despite their different intracellular locations and specificity for substrates [5, 13]. In contrast, Grxs are broadly categorized into two groups, i.e., monothiol and dithiol Grxs, based on the number of cysteine residues present in their catalytic site [1, 14, 15]. Grxs can also be divided in six different classes having motif sequence CXX [C/S] (Class I), CGFS (Class II), CC-type (CCXX, CXXC, CCXS; Class III), CXX [C/S] with DER or DUF 547 domain (Class IV), CPWG with extended C-terminal (Class V) and CPW [C/S] with one additional DUF 236 domain at N-terminal (Class VI) [14, 15]. The CC-type class III Grx is only found in higher plants while class V and VI are only present in a few marine cyanobacteria [14, 15]. Grxs catalyze forward reaction of glutathionylation *via* a dithiol mechanism similar to Trxs; however, they can also act using the monothiol mechanism which is required for deglutathionylation of proteins [1, 4]. Besides controlling the redox state of proteins, Grxs, specifically monothiol Grxs, play a significant role in iron homeostasis by participating in biosynthesis and targeting of iron-sulfur clusters [16, 17].

Earlier studies focused on biochemical and structural characterization of Trxs and Grxs. In addition to their catalytic motif based classification, computational studies established evolutionary relationship of Grxs or Trxs from different organisms [1, 5, 14, 15]. Here, we conducted *in silico* comparative structural analysis combined with sequence and phylogenetic analyses of Trxs and Grxs in 12 organisms of three domains of life to get better insights into their evolutionary and structural relationship. Results obtained suggested that substitutions with amino acids having similar chemical properties helped Trxs and Grxs to conserve their Trx/Grx fold and function during evolution. Trxs and Grxs are structurally more similar in prokaryotes than eukaryotes though two proteins have opposite electrostatic surface potential. However, catalytic motifs are constituted by uncharged amino acids in both proteins. Results of phylogenetic analysis suggested the usefulness of Trxs and Grxs sequences in establishing an evolutionary lineage.

2. Materials and methods

2.1. Experimental organisms and sequence retrieval from biological databases

Total 12 organisms such as *Archaeoglobus veneficus*, *Escherichia coli* K12, *Synechococcus elongatus* PCC 7942, *Saccharomyces cerevisiae* S288c, *Arabidopsis thaliana, Caenorhabditis elegans*, *Drosophila melanogaster*, *Danio rerio*, *Xenopus laevie*, *Gekko japonicus*, *Gallus* and *Homo sapiens* were selected as a representative of archaea, bacteria, cyanobacteria, fungi, plant, nematode, arthropod, fish, amphibian, reptile, bird and mammal, respectively. These organisms are commonly used as a model biological system to study various biological processes including but not limited to signal transduction, gene regulation, metabolism, the function of a particular protein or a gene, redox and iron homeostasis, various diseases and developmental process. Trxs and Grxs amino acid sequences from the abovementioned organisms were manually retrieved from NCBI Genome Database (https://www.ncbi.nlm.nih.gov/genome/?term) and UniProt Database (https://www.uniprot.org/) [18]. The duplicate, truncated and missannotated sequences were eliminated manually. Total 153 sequences of Trxs and Grxs were used for further analysis. The retrieved sequences were sorted into respective classes based on their active site motif and their location within a cell [19]. Sub-cellular localization of Trxs and Grxs in different organisms was predicted using CELLO (http://cello.life.nctu.edu.tw/) [20] and WoLF PSORT servers (https://www.genscript.com/wolf-psort.html) [21].

2.2. Primary sequence analysis

The physiochemical properties like theoretical isoelectric point (pI), molecular weight (MW), extinction coefficient (EC) and peptide length were analyzed using Expasy Protparam server (http://web.expasy.org/protparam/) [22]. The catalytic site residues and protein domains were identified by NCBI Conserved Domain Database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Protein sequences were subjected to multiple sequence alignment (MSA) using Clustal W software with default parameters setting and Gonnet as protein weight matrix [23]. The WebLogo (http://weblogo.berkeley.edu/logo.cgi) diagram for Trxs and Grxs were built using MSA files [24]. The percentage amino acid composition of Trxs and Grxs were computed using Molecular Evolutionary Genetics Analysis software version X (MEGA-X) with the help of MSA files [25].

2.3. Phylogenetic analysis

Phylogenetic analysis was done using MEGA-X software [25]. The evolutionary tree was built using the maximum likelihood method [26] and evolutionary distance was computed using the JTT method [27]. All positions containing gaps and missing data in MSA were eliminated during the construction of the phylogenetic tree. Bootstrap analysis was done assigning 500 replication cycles [28]. The tree was drawn to scale with branch lengths measured in the number of substitutions per site. The branch length in the phylogenetic tree is directly proportional to the rate of amino acid substitution and evolutionary distance.

2.4. Comparative structural analysis

The high-resolution 3D protein structures of selected organisms available in the Protein Data Bank were retrieved for structural analysis [29]. The solved tertiary structures of Trxs and Grxs of *E. coli*, *S. cerevisiae*, *Synechocystis* sp. PCC 6803, *D. melanogaster*, *A. thaliana* and *H. sapiens* were retrieved from the PDB database. The Trxs structures having PDB id 1SRX, 1THX, 3F3R, 1XWC, 1 ER T and 1XFL, while Grxs structures having PDB id 2WCI, 4MJE, 5J3R, 2WUL and 3IPZ were used in this study for structural analysis. The unsolved structure was modelled by the Swiss Model server using a suitable template and validated through the Ramachandran plot [30]. Structures were analyzed using UCSF Chimera 1.14 software based on electrostatic surface potential and hydrophobicity index [31]. Structural heterogeneity was computed on the basis of root mean square deviation (RMSD) and percentage similarity. The topology diagrams of proteins were generated using Pro-origami online server (http://munk.csse.unimelb.edu.au/pro-origami/) [32].

3. Results

3.1. Distribution and characteristics of Trxs and Grxs: grxs are more versatile than trxs

We included *E. coli* K12, *A. veneficus*, *S. elongatus* PCC 7942, *S. cerevisiae* S288c, *A. thaliana, C. elegans*, *D. melanogaster*, *D. rerio*, *X. laevie*, *G. japonicus*, *G. gallus* and *H. sapiens* in this study. These organisms were selected as a representative of three domains of life to study the characteristics and relationship of Trxs and Grxs (Table1). Total 153 protein sequences were manually retrieved from biological databases using the names of the abovementioned organisms. Out of 153 proteins, 88 protein sequences were of Grxs while 65 sequences were of Trxs. Active site motif analysis revealed that Trxs generally possesses a conserved active site motif CGPC in all studied organisms except *A. thaliana*. We report active site motifs such as CGPC, CGGC, CPPC, CVPC, CASC, CRKC and CGSC in Trxs of *A. thaliana* (Table 2).

A. thaliana possessed maximum number, i.e., 27, and types, i.e., f, h, m, o, x and y, of Trxs in comparison to other studied organisms (Table 2). The mitochondrion of *A. thaliana* have only 'o' type Trx while chloroplast contains m, y, x and f Trxs. The absence of these Trxs in the cytoplasm stated their specificity towards cellular organelles. Notably, organelle-specific Trxs typically had CGPC active site motif sequence (Table 2). Due to the absence of signal peptide, Trx 'h' was considered as a

Table 1. The number of thioredoxins and glutaredoxins found in studied organisms.

Class	Organism	Thioredoxin	Glutaredoxin
Archaea	*Archaeoglobus veneficsus*	5	3
Cyanobacteria	*Synechococcus elongatus* PCC 7942	3	2
Bacteria	*Escherichia coli* K-12	4	4
Fungi	*Saccharomyces cerevisiae* S288C	3	8
Plant	*Arabidopsis thaliana*	27	35
Nematode	*Caenorhabditis elegans*	7	6
Arthropod	*Drosophila melanogaster*	3	3
Fish	*Danio rerio*	3	9
Amphibian	*Xenopus laevis*	4	2
Reptile	*Gekko japonicus*	1	5
Bird	*Gallus*	2	4
Mammal	*Homo sapiens*	3	7
Total		**65**	**88**

cytoplasmic Trx. This proposal is supported by the fact that 'h' Trx is found in the phloem sap of rice and other plants where it can translocate through plasmodesmata due to its small size [33]. *A. veneficus* Trxs were found to have CGPC, CPSC, and CPYC active site motifs. Trxs analyzed in this study were approximately 140 amino acids long, had a single Trx family domain, and showed a wide range of theoretical pI, i.e., 4.2 to 9.5 (Table 2).

The frequency of occurrence of different amino acids at different positions of the peptide chain dictates diversity in the protein sequence. Trxs had very high percentage of non-polar amino acids such as alanine, valine, leucine, serine, and charged amino acids such as aspartic acid and lysine (Figure 1). However, valine was the most abundant amino acid while tryptophan, tyrosine, histidine, arginine, methionine and glutamine residues were the least abundant amino acids in Trxs (Figure 1). The frequency of phenylalanine as compared to other aromatic amino acids was considerably high in all Trxs. Trxs had ~1.5% cysteine while some amino acids were completely absent in Trxs. For example, arginine and histidine were absent in Yeast, histidine was absent in *Drosophila*, and arginine was absent in the Trxs of fish, birds and mammals (Figure 1; Supplementary Table 1). The cysteine residues present in the active sites of Trxs were highly conserved (Supplementary Fig. 1). Also, a *cis*-proline residue located five residues after the catalytic site, and its neighbor threonine residue, two glycine, one phenylalanine, one valine and two aspartic acid residues were highly conserved in all studied Trxs (Supplementary Fig. 1).

Contrary to Trxs, Grxs showed diversity in their active site sequence. In addition to commonly known CPFC, CPYC and CGFS active site motifs, we report CSYC, CGYC, CPYS, CSYS, and CFYC active sites in Grxs of studied organisms. Importantly, the diversity in the active site was more common in eukaryotes (Table 3). *A. thaliana* possessed 17 monothiol and 14 dithiol Grxs. Unlike Trxs, Grxs were either monomeric or multimeric proteins similar to previous reports [14, 15, 33]. The multidomain Grxs had either PICOT domain or multiple Grx domains. Also, multidomain

Table 2. Sub-cellular localization and primary sequence analysis of thioredoxins (Trxs) based on active site, attached protein domain, theoretical pI, amino acids length, molecular weight (Dalton) and extinction coefficient ($M^{-1} cm^{-1}$). Asterisks indicate non-CGPC active sites identified in this study. sv; splice variants of same gene id 25828.

Organism	Location	Accession No	Name of protein	Active site	Domain attached	pI	Length	Molecular weight	Extinction coefficient
Homo sapiens	Mitochondria	XP_005261565.1	*Trx isoform* X1 (sv)	CGPC	Trx	8.85	197	21728.05	15470
	Mitochondria	XP_006724289.1	*Trx isoform* X2 (sv)	CGPC	Trx	8.46	166	18383.30	13980
	Cytosol	NP_003320.2	*Trx isoform* 1	CGPC	Trx	4.82	105	11737.50	6990
Danio rerio	Mitochondria	NP_991204.1	Trx	CGPC	Trx	8.50	166	18458.31	8480
	Cytosol	Q7ZUI4	Trx1	CGPC	Trx	4.69	108	12014.59	9970
	Cytosol	NP_001002461.1	Trx2	CGPC	Trx	5.30	107	11874.56	8480
Drosophila melanogaster	Cytosol	NP_572212.1	TrxT	CGPC	Trx	4.23	157	17488.54	12950
	Cytosol	NP_523526.1	Trx2	CGPC	Trx	4.23	106	17488.54	12950
	Cytosol	P47938	Trx1	CGPC	Trx	4.73	107	11736.66	8480
Caenorhabditis elegans	Cytosol	NP_001256207.1	Trx7	CGPC	Trx	9.45	119	12384.48	11460
	Cytosol	NP_503440.2	Trx6	CGPC	Trx	4.73	142	11736.66	8480
	Cytosol	NP_500961.2	Trx5	*CGHC	PDIa Family	4.42	136	12938.52	6990
	Cytosol	NP_500578.2	Trx4	CGPC	Trx	4.88	158	16172.37	21430
	Cytosol	NP_001021885.1	Trx1	CGPC	Trx	5.74	115	15672.72	38960
	Cytosol	NP_001021886.1	Trx2	CGPC	Trx	7.56	114	18338.15	28420
	Cytosol	NP_491142.1	Trx3	CGPC	Trx	5.59	107	13323.48	11460
Escherichia coli K12	Cytosol	NP_418228.2	Trx1	CGPC	Trx	4.69	109	12967.04	9970
	Cytosol	NP_417077.1	Trx2	CGPC	Trx	4.66	139	12103.94	6990
	Cytosol	WP_074455222.1	TrxA	CGPC	Trx	4.67	109	11806.62	13980
	Cytosol	SQD02739.1	TrxC	CGPC	Trx	5.00	139	15554.77	16500
Saccharomyces cerevisiae S288C	Cytosol	NP_011725.3	Trx2	CGPC	Trx	4.79	104	11203.88	11460
	Cytosol	NP_013144.1	Trx1	CGPC	Trx	4.79	103	11234.98	9970
	Mitochondria	5YKW_A	Trx3	CGPC	Trx	9.08	127	14432.10	11460
Archaeoglobus veneficus	Cytosol	WP_013682912.1	Trx1	CGPC	Trx	5.37	106	12081.09	6990
	Cytosol	WP_083809303.1	Trx2	*CPYC	Trx	6.13	124	14289.68	7450
	Cytosol	WP_013683699.1	Trx3	CGPC	Trx	7.70	134	15242.81	19480
	Cytosol	WP_013683913.1	Trx4	*CPYC	Trx	4.93	81	9034.64	2980
	Cytosol	WP_013684269.1	Trx5	*CPSC	Trx	5.65	196	21997.26	22920

(*continued on next page*)

Table 2 (*continued*)

Organism	Location	Accession No	Name of protein	Active site	Domain attached	pI	Length	Molecular weight	Extinction coefficient
Synechococcus elongatus PCC 7942	Cytosol	WP_011244574.1	Trx1	CGPC	Trx	4.90	107	11648.40	13980
	Cytosol	WP_208672674.1	Trx2	CGPC	Trx	7.73	111	12680.62	26470
	Cytosol	WP_011378192.1	Trx3	CGPC	Trx	4.49	107	11924.69	20970
Arabidopsis thaliana	Chloroplast	NP_849585.1	TrxM1	CGPC	Trx	9.14	179	19664.58	22460
	Cytosol	NP_001117249.1	Trx4 CYS HIS rich	*CGGC	Trx	6.88	177	19372.15	5960
	Mitochondria	NP_564371.1	TrxO2	CGPC	Trx	9.01	159	17623.18	18450
	Chloroplast	NP_175021.2	TrxY2	CGPC	Trx	8.45	167	18592.32	16960
	Cytosol	NP_175128.1	TrxH5	*CPPC	Trx	5.19	118	13122.32	11000
	Cytosol	OAP18792.1	TrxH4	*CPPC	Trx	9.06	182	19647.73	12950
	Cytosol	NP_564566.1	TrxX	CGPC	Trx	7.80	129	14531.75	13980
	Cytosol	NP_176182.1	TrxH7	CGPC	Trx	7.80	119	14531.75	13980
	Cytosol	NP_177146.1	TrxH8	CGPC	Trx	8.97	148	17250.19	26470
	Cytosol	NP_001325846.1	TrxH10	*CVPC	Trx	8.75	154	17370.92	24980
	Chloroplast	NP_177802.2	TrxY1	CGPC	Trx	9.03	172	19250.19	11460
	Mitochondria	NP_001078006.1	TrxO1	CGPC	Trx	9.45	194	21191.29	16960
	Cytosol	NP_186922.1	TrxF1	CGPC	Trx	9.12	178	19325.43	18450
	Cytosol	NP_187329.1	TrxZ	CGPC	Trx	5.65	183	20670.08	11460
	Cytosol	NP_001325992.1	TrxH9	CGPC	Trx	5.12	140	15334.37	16500
	Chloroplast	Q9SEU8	TrxM2	CGPC	Trx	9.35	186	20312.43	20970
	Chloroplast	Q9SEU7	TrxM3	CGPC	Trx	8.65	173	19500.30	18450
	Chloroplast	NP_188155.1	TrxM4	CGPC	Trx	9.62	193	21172.28	19480
	Cytosol	NP_190672.1	TrxH1	CGPC	Trx	5.64	114	12672.73	16500
	Cytosol	NP_001330106.1	Trx5 CYS HIS rich	*CGGC	Trx	8.17	186	20427.38	15470
	Cytosol	OAO89812.1	TrxH3	*CPPC	Trx	5.06	118	13109.28	11000
	Cytosol	Q38879	TrxH2	CGPC	Trx	5.74	133	14675.85	11000
	Cytosol	NP_198811.1	Trx2	CGPC	Trx	5.74	134	14732.90	11000
	Cytosol	NP_194346.1	Trx1 CYS HIS rich	*CGSC	Trx	8.72	221	24352.51	23950
	Cytosol	NP_567831.1	Trx2 CYS HIS rich	*CASC	Trx	9.06	235	25843.64	20970
	Cytosol	AED90720.1	Trx2 WCRKC	*CRKC	Trx	8.32	192	21836.13	25440
	Chloroplast	OAO91968.1	TrxF2	CGPC	Trx	9.06	185	19999.20	18450
Xenopus laevis	Cytosol	NP_001080066.1	Trx2L homoeolog	CGPC	Trx	7.71	170	18584.59	5500
	Cytosol	NP_001088487.1	TrxL homoeolog	CGPC	Trx	5.34	105	11755.55	9970
	Cytosol	A2VDE6	Trx1	CGPC	Trx	4.96	105	11864.72	6990
	Cytosol	NP_001085522.1	Trx2	CGPC	Trx	5.34	105	11755.55	9970
Gekko japonicus	Mitochondria	XP_015265683.1	Trx	CGPC	Trx	9.27	174	19212.44	6990
Gallus gallus	Cytosol	NP_990784.1	Trx	CGPC	Trx	5.10	105	11700.52	6990
	Mitochondria	NP_001026581.1	Trx	CGPC	Trx	9.34	140	15170.64	12490

Grxs were commonly found in eukaryotic systems. Grxs were approximately 100–150 amino acids long and had theoretical pI ranging from 4.5 to 9.5 (Table 3). Important to note that Trxs and Grxs had an almost similar range of theoretical pI values despite their difference in amino acids composition. However, the vast range of theoretical pI was the result of the diversity of amino acids present in the two proteins. Grxs had a comparatively higher percentage of a cysteine residue, i.e., 2%, than Trxs. Grxs possessed high percentage of leucine, glutamic acid, glycine, alanine, serine and valine amino acids (Figure 2). The presence of tryptophan and histidine was less than 1%, while phenylalanine and tyrosine were higher than 2.5% (Figure 2).

However, other amino acids were moderately present in all Grxs (Supplementary Table 2). Similar to Trxs, several amino acids were absent in Grxs. For example, histidine was absent in some Grxs of archaea while tryptophan was absent in Grxs of fungus, nematode, reptile and mammal. Grxs of fish lacked aspartate and tryptophan, and histidine and tryptophan amino acids were absent in the bird's Grxs (Figure 2; Supplementary Table 2). Although Grxs showed diversity in their amino

acids composition, some of the residues were highly conserved. The N-terminal cysteine residue of active site motif, the *cis*-proline and the two consecutive glycine residues were highly conserved in all Grxs (Supplementary Fig. 2). Notably, amino acid residues of glutathione binding sites were not conserved and substitution by a similar group of amino acid residues was observed at glutathione binding sites in Grxs (Supplementary Fig. 2).

3.2. Trxs and Grxs phylogeny support endosymbiotic theory for origin of chloroplast and mitochondria

Phylogenetic tree consisting of 65 Trxs and 88 Grxs protein sequences from 12 different organisms was constructed using maximum likelihood method to decipher the evolutionary relationship between two groups of proteins (Figure 3). Trxs and Grxs got separated from a common ancestor at the very beginning of the phylogenetic tree and resulted in two individual groups of proteins (Figure 3). Trxs group was differentiated into different clusters based on phylogenetic analysis. The mitochondrial Trxs
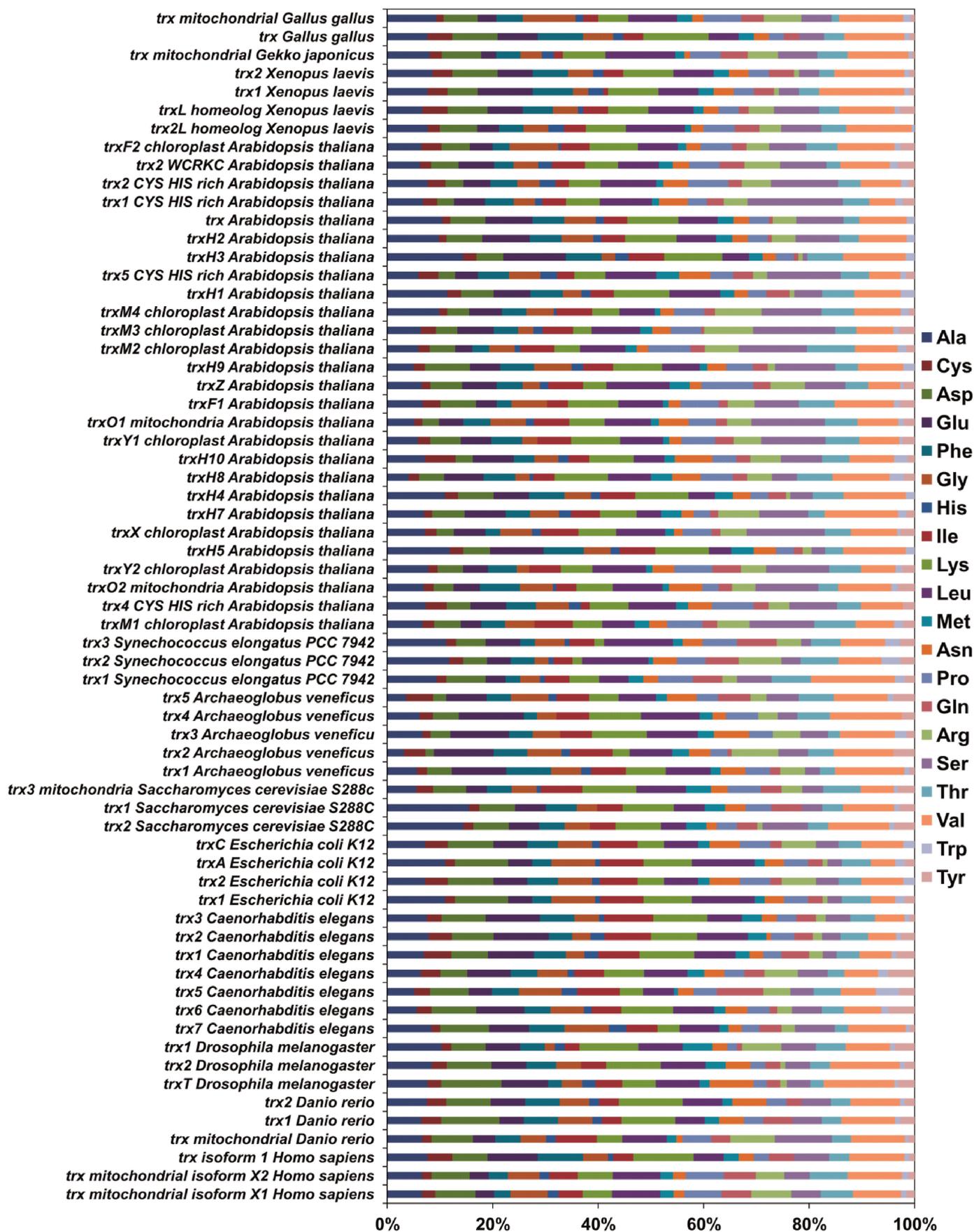
Figure 1. Frequency of occurrence of different amino acids in 65 thioredoxins (Trxs) of different organisms of three domains of life. The frequencies of different amino acids in Trxs of different organisms are shown on the x-axis. The twenty different amino acids are marked with different color codes in the diagram.

Table 3. Subcellular localization and primary sequence analysis of glutaredoxins (Grxs) based on active site, attached protein domain, theoretical pI, amino acids length, molecular weight (Dalton) and extinction coefficient (M^{-1} cm^{-1}).

Organism	Location	Accession No	Name of protein	Active site	Domain attached	pI	Length	Molecular weight	Extinction coefficient
Synechococcus elongatus PCC 7942	Cytosol	WP_011244089.1	Grx3	CPFC	Grx	5.50	87	9474.79	8480
	Cytosol	WP_208672557.1	Grx4	CGFS	Grx	4.31	108	12064.98	9970
Homo sapiens	Cytosol	NP_001230587.1	Grx1	CPYC	Grx	8.33	106	11775.74	2980
	Mitochondria	NP_057150.2	Grx2 isoform 1	CSYC	Grx	9.54	165	18723.46	14440
	Mitochondria	NP_001306220.1	Grx2 isoform 3	CSYC	Grx	8.34	124	14127.21	7450
	Mitochondria	XP_016856886.1	Grx2 isoform X1	CSYC	Grx	8.34	124	14127.21	7450
	Cytosol	NP_001186797.1	Grx3 isoform 1	CGFS	2 Grx_PICOT 1 Trx_PICOT (N-terminal)	5.31	335	37432.03	33920
	Cytosol	NP_001308909.1	Grx3 isoform 2	CGFS	Grx	4.96	189	21497.71	21430
	Cytosol	XP_016870963.1	Grx3 isoform X1	CGFS	Grx	4.96	189	21497.71	21430
Xenopus laevis	Cytosol	XP_018080656.1	Grx3	CGFS	2 Grx_PICOT 1 Trx_PICOT (N-terminal)	5.38	326	36608.99	32430
	Mitochondria	AAH70695.1	Grx5	CGFS	Grx	7.91	150	16642.00	16960
Drosophila melanogaster	Cytosol	Q9W2D1	Grx1	CPYC	Grx	8.82	116	13024.94	7450
	Cytosol	NP_609641.1	Grx4	CGFS	1 Grx_PICOT 1 Trx_PICOT (N-terminal)	4.80	216	23630.97	19940
	Mitochondria	Q8SXQ5	Grx5	CGFS	Grx	7.73	159	17668.34	9970
Danio rerio	Cytosol	NP_001005942.1	Grx	CPYC	Grx	7.63	105	11348.26	4470
	Cytosol	AAH96952.1	Grx	CSYC	Grx	6.54	105	11379.24	4470
	Cytosol	NP_001005942.1	Grx1	CSYC	Grx	6.54	105	11337.16	4470
	Cytosol	XP_005161791.1	Grx2 isoform X1	CPYC	Grx	8.14	157	17185.56	8480
	Cytosol	XP_005161792.1	Grx2 isoform X2	CPYC	Grx	8.61	152	16632.90	5960
	Cytosol	XP_005161790.1	Grx2 isoform X3	CPYC	Grx	7.50	170	18420.04	4470
	Cytosol	XP_009294644.1	Grx2 isoform X4	CPYC	Grx	6.79	134	14408.31	2980
	Cytosol	NP_001005950.1	Grx3	CGFS	1 Grx_PICOT 1 Trx_PICOT (N-terminal)	5.18	326	36335.44	28420
	Mitochondria	NP_998186.1	Grx5	CGFS	Grx	8.71	155	17376.94	15470
Caenorhabditis elegans	Cytosol	NP_490812.1	Grx1	CPYC	Grx	8.46	105	11297.93	9970
	Cytosol	NP_001040891.1	Grx2	CTFC	GrxC	5.76	119	13130.91	7450
	Cytosol	NP_001040892.1	Grx3	CTFC	GrxC	5.71	96	10693.07	7450
	Cytosol	NP_499610.1	Grx4	CGFS	Grx	6.73	142	15767.99	9970
	Cytosol	NP_001033391.1	Grx5	CGYC	Grx	7.65	131	14619.74	10430
	Mitochondria	NP_001023757.1	Grx6	CGFS	1 Grx_PICOT 1 Trx_PICOT (N-terminal) 1 Grx domain (C-terminal)	5.19	342	37939.14	23950
Escherichia coli K12	Cytosol	MBI0727830.1	GrxA	CPYC	Grx	4.81	85	9684.85	11460
	Cytosol	STF72101.1	GrxB	CPYC	Grx	7.72	215	24350.21	22920
	Cytosol	MBA1844136.1	GrxC	CPYC	Grx	6.71	83	9137.49	4470
	Cytosol	2WCIA	GrxD	CGFS	Grx	4.69	115	12878.76	18450
Saccharomyces cerevisiae S288C	Cytosol	NP_009895.1	Grx1	CPYC	Grx	4.98	110	12380.19	5960
	Cytosol	NP_010801.1	Grx2	CPYC	Grx	6.73	143	15861.47	4470
	Cytosol	NP_010383.4	Grx3	CGFS	1 Grx_PICOT 1 Trx_PICOT (N-terminal)	4.37	250	28261.34	18450
	Cytosol	NP_011101.3	Grx4	CGFS	1 Grx_PICOT 1 Trx_PICOT (N-terminal)	4.57	244	27492.90	16960
	Mitochondria	NP_015266.1	Grx5	CGFS	GrxD	4.85	150	16931.45	11460
	Cytosol	NP_010274.1	Grx6	CSYS	Grx	6.01	231	25783.28	15930
	Cytosol	P38068	Grx7	CPYS	Grx	5.64	203	22565.20	14440
	Cytosol	Q05926	Grx8	CPDC	GrxC	7.78	109	12519.40	24980
Archaeoglobus veneficus	Cytosol	WP_013684320.1	Grx1	CPHC	2 HEAT repeat (C-terminal)	4.71	315	10459.27	12950
	Cytosol	F2KMX5	Grx2	CPYC	Grx	4.93	81	9034.64	2980
	Cytosol	WP_013683032.1	Grx3	CPKC	Grx	5.17	95	34962.43	1490

(continued on next page)

Table 3 (continued)

Organism	Location	Accession No	Name of protein	Active site	Domain attached	pI	Length	Molecular weight	Extinction coefficient
Gekko japonicus	Cytosol	AAW51391.1	Grx	CGFS	1 Grx_PICOT 1 Grx domain (C-terminal)	9.24	162	18454.83	16960
	Cytosol	XP_015283682.1	Grx1	CPYC	Grx	6.81	103	11533.46	5960
	Cytosol	XP_015281914.1	Grx2 isoform X1	CSYC	Grx	9.66	145	16353.52	9970
	Cytosol	XP_015281914.1	Grx2 isoform X2	CSYC	Grx	8.86	128	14280.24	4470
	Cytosol	XP_015262026.1	Grx3	CGFS	2 Grx_PICOT 1 Trx_PICOT (N-terminal)	5.55	335	37312.74	39420
Gallus gallus	Cytosol	NP_990491.1	Grx1	CPYC	Grx	8.40	101	11397.45	2980
	Cytosol	XP_004943187.1	Grx2	CFYC	Grx	7.67	123	13390.37	2980
	Cytosol	XP_422200.3	Grx2 X1	CFYC	Grx	9.42	137	15164.57	2980
	Cytosol	NP_001264313.1	Grx3	CGFS	2 Grx_PICOT 1 Trx_PICOT (N-terminal)	5.47	328	36573.94	33920
Arabidopsis thaliana	Cytosol	NP_566522.1	Grx4	CGFS	Grx	5.91	169	18734.21	9970
	Cytosol	NP_568962.1	GrxC1	CGYC	Grx	5.39	125	13610.61	9970
	Cytosol	Q29PZ1	GrxC10	CCMC	Grx	5.23	148	15709.22	6990
	Cytosol	NP_191854.1	GrxC11	CCMC	Grx	8.86	103	11312.19	8480
	Cytosol	NP_001318444.1	GrxC12	CCMC	Grx	7.74	103	11261.06	8480
	Cytosol	O82255	GrxC13	CCLC	Grx	7.63	102	11275.32	4470
	Cytosol	NP_191855.1	GrxC14	CCLC	Grx	7.65	102	11283.22	2980
	Cytosol	Q9FNE2	GrxC2	CPYC	Grx	6.71	111	11756.42	8480
	Cytosol	KAG7590455.1	GrxC3	CPYC	Grx	5.78	130	14247.38	4470
	Cytosol	Q8LFQ6	GrxC4	CPYC	Grx	5.68	135	14827.11	11460
	Chloroplast	Q8GWS0	GrxC5	CSYC	Grx	9.12	174	18813.59	12950
	Chloroplast	Q8L9S3	GrxC6	CCMC	Grx	4.95	144	15566.01	6990
	Cytosol	Q96305	GrxC7	CCMC	Grx	6.88	136	14201.60	15470
	Cytosol	Q8LF89	GrxC8	CCMC	Grx	8.78	140	14969.60	9970
	Cytosol	Q9SGP6	GrxC9	CCMC	Grx	5.71	137	14758.19	8480
	Cytosol	NP_171801.1	GrxS1	CCMS	Grx	5.75	102	11067.88	8480
	Cytosol	Q9LIF1	GrxS10	CCMS	Grx	8.54	102	11056.05	13980
	Cytosol	Q9M9Y9	GrxS11	CCLS	Grx	8.30	99	10889.98	2980
	Chloroplast	Q8LBS4	GrxS12	CSYS	Grx	8.37	179	19124.90	20970
	Cytosol	Q84TF4	GrxS13	CCLG	Grx	8.95	150	16365.85	20970
	Chloroplast	Q84Y95	GrxS14	CGFS	Grx	8.67	173	19309.40	9970
	Mitochondria	NP_001030704.1	GrxS15	CGFS	Grx	5.91	169	18734.21	9970
	Chloroplast	KAG7569621.1	GrxS16	CGFS	1 Grx_PICOT 1 GIY-YIG Domain (N-terminal)	7.72	293	32205.72	26930
	Cytosol	Q9ZPH2	GrxS17	CGFS	3 Grx_PICOT 1 Trx_PICOT (N-terminal)	5.00	488	53115.26	39420
	Cytosol	NP_197361.1	GrxS2	CCMS	Grx	6.06	102	11039.09	8480
	Cytosol	O23421	GrxS3	CCMS	Grx	7.76	102	11168.20	6990
	Cytosol	O23419	GrxS4	CCMS	Grx	8.54	102	11161.25	6990
	Cytosol	O23420	GrxS5	CCMS	Grx	6.71	102	11205.21	6990
	Cytosol	NP_191852.1	GrxS6	CCMS	Grx	8.55	102	11111.03	6990
	Cytosol	Q6NLU2	GrxS7	CCMS	Grx	8.54	102	11244.34	6990
	Cytosol	NP_193301.1	GrxS8	CCMS	Grx	7.73	102	11311.38	8480
	Cytosol	NP_180612.1	GrxS9	CCMS	Grx	6.80	102	11080.08	1490
	Cytosol	NP_186849.1	ROXY1	CCMC	Grx	6.88	136	14201.60	15470
	Cytosol	NP_174170.1	ROXY19	CCMC	Grx	5.71	137	14758.19	8480
	Cytosol	KAG7550582.1	ROXY2	CCMC	Grx	8.78	140	14969.60	9970

clustered together and formed a separate group with Trxs of bacteria and archaea (Figure 3). The chloroplast specific Trxs clustered with the Trxs of cyanobacteria and nematode. However, cytoplasmic Trxs of *A. thaliana* formed distinct clusters and shared ancestral relationships with nematodes and amphibians (Figure 3). Trxs of higher organisms shared common clusters in the phylogenetic tree (Figure 3). In contrast, Grxs were divided into three major subgroups. One of the subgroups had monothiol CGFS type Grxs while another subgroup had CC-type Grxs from *A. thaliana* forming a separate cluster that got separated at the beginning of the tree (Figure 3). The third subgroup possessed
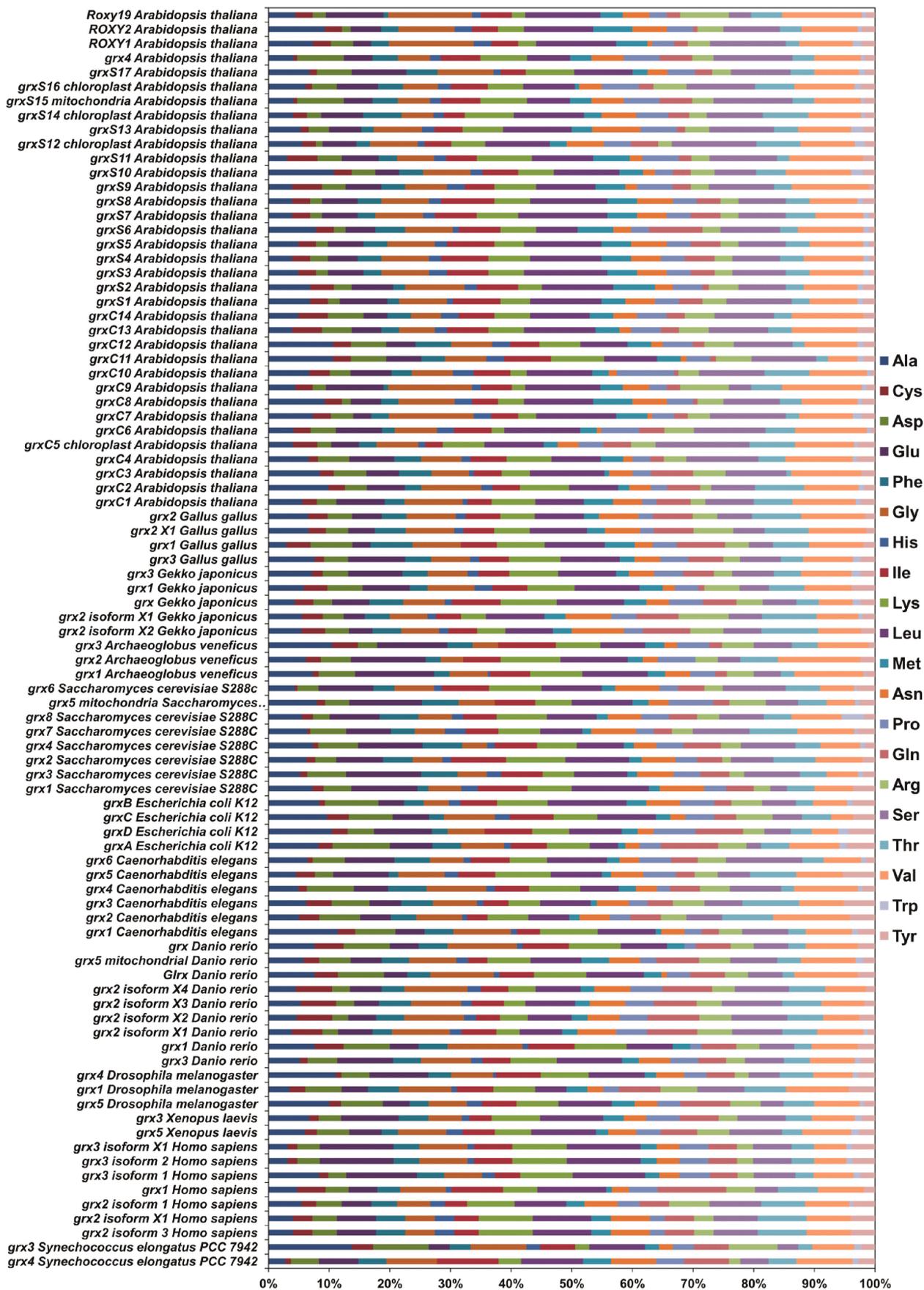
Figure 2. Frequency of occurrence of different amino acids in 88 glutaredoxins (Grxs) of different organisms of three domains of life. The frequencies of different amino acids in Grxs of different organisms are shown on the x-axis. The twenty different amino acids are marked with different color codes in the diagram.
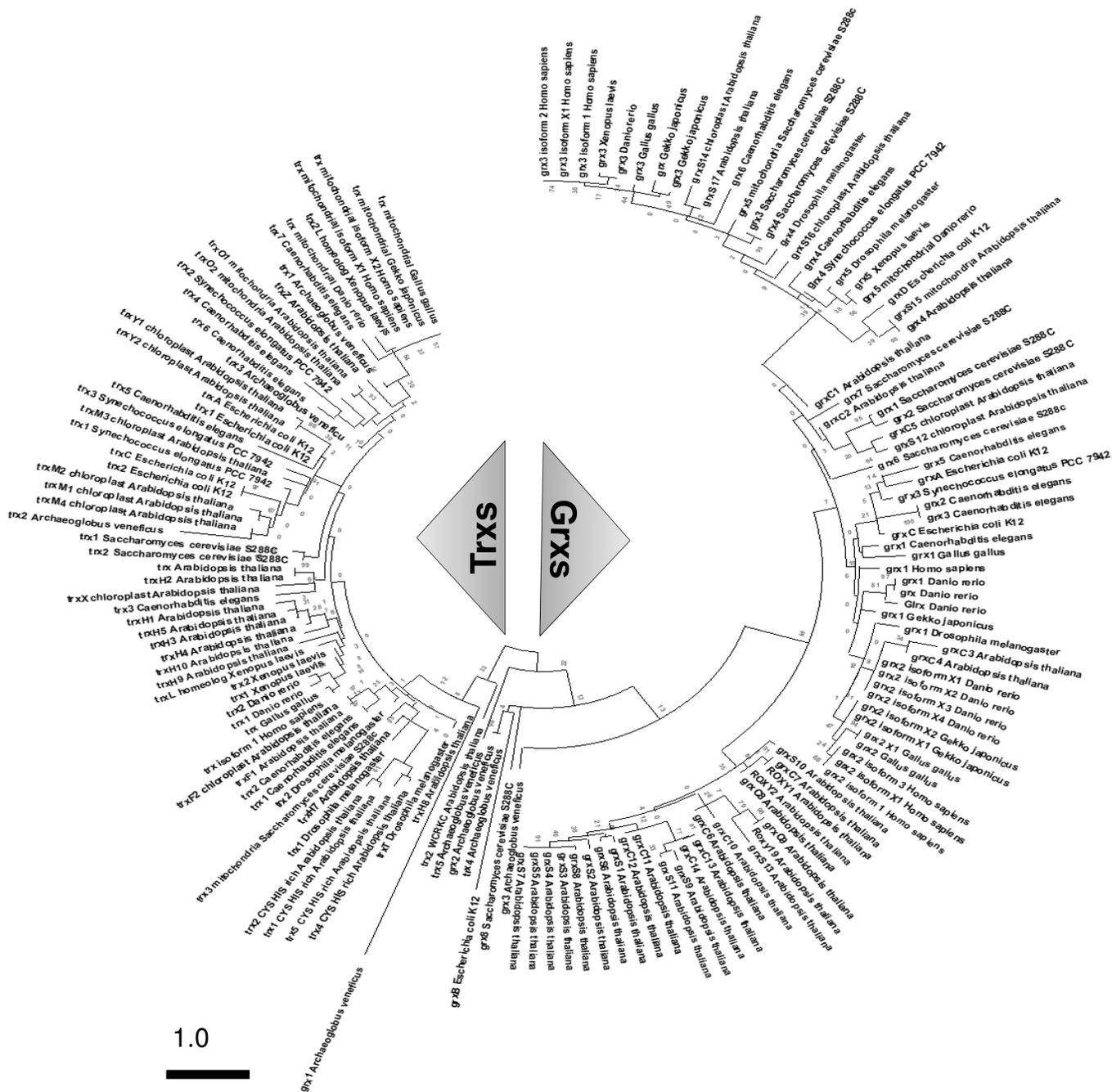
Figure 3. Evolutionary relationship of thioredoxins (Trxs) and glutaredoxins (Grxs). The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. The tree with the highest log likelihood (-3169.91) is shown. The tree was drawn to scale and branch length in the tree is directly proportional to the rate of amino acid substitution and evolutionary distance. The analysis involved 65 Trxs and 88 Grxs of 12 different organisms representing three domains of life.

particularly dithiol Grxs from all organisms. It is worth mentioning here that similar to Trxs, Grxs located in the chloroplast and mitochondria shared common clusters with Grxs of cyanobacteria and bacteria, respectively (Figure 3).

3.3. Evolutionary relationship of Trxs in three domains of life

We further explored the evolutionary relationship among Trxs of 12 distantly related organisms (Supplementary Fig. 3). The phylogenetic tree of Trxs was divided into four major subgroups. Subgroup 1 was comprised of Trxs of bacteria, archaea, cyanobacteria together

with mitochondrial and chloroplastic Trxs, while subgroup 2 had cytoplasmic Trxs of eukaryotes. However, in subgroup 2, *A. thaliana* Trxs formed an independent cluster which shared common ancestry with cytoplasmic Trxs of amphibian, mammal, fish and bird (Supplementary Fig. 3). Notably, Trxs f1 and f2 of *A. thaliana* shared close ancestry with Trx4 and Trx5 of *A. veneficus* which suggested that plant Trx f has an archaebacterial origin. Trxs of bird, mammal, amphibian, fish, and nematode formed a separate cluster. Two cytosolic Trxs of *A. thaliana* (Table 2; accession number NP_198811.1 and NP_190672.1) shared an ancestral relationship with 3 Trxs of *S. cerevisiae* and formed a small cluster in subgroup 2. The cysteine-

histidine-rich Trxs of *A. thaliana* formed a small subgroup 3. Two Trxs of *A. veneficus* and three Trxs (two Trxs f and one WCRKC Trx) of *A. thaliana* build subgroup 4. Importantly, two Trxs of *A. veneficus* got separated at the very beginning from an inner node of the tree (Supplementary Fig. 3).

3.4. Evolutionary relationship of Grxs in three domains of life

Phylogenetic tree of 88 Grxs showed three major subgroups (Supplementary Fig. 4). Subgroup 1 had Grxs of eukaryotic organisms which further divided to form two individual groups. The first group generally possessed Grxs of plants where monothiol Grxs formed a separate cluster (S1 to S8) from dithiol Grxs. The other part of subgroup 1 was largely dominated by the Grxs of bird, fish, mammal and fungus but Grxs of nematode, arthropod and some Grxs of the plant were also clustered in this group (Supplementary Fig. 4). Notably, Grxs of fungus and plant were present in close proximity and got separated at the beginning of the inner node while Grxs of other organisms were found to be descendent of fungus and plant. The presence of Grxs of fungus and plants in close proximity suggested their common ancestral relationship. Subgroup 2 had Grxs of both prokaryotes and eukaryotes whereas Grxs of archaea got separated at the very beginning from other organisms (Supplementary Fig. 4). The mitochondrial Grxs shared ancestral relationships with the Grxs of bacteria, while Grx3, Grx4 and Grx5 of different organisms formed small out-groups.

Subgroup 3 had a few Grxs of both prokaryotes and eukaryotes forming the smallest subgroup which further divided to form two distinct out-groups. One group possessed Grxs of archaea, bacteria and fungus while the other group primarily had Grxs of eukaryotes along with bacterial and cyanobacterial Grxs. Importantly, subgroup 3 did not contain Grxs of plant and arthropod.

3.5. Trxs and Grxs have opposite electrostatic surface potential

We conducted a comparative analysis of Trxs and Grxs of distantly related organisms to better understand their evolutionary and structural relationship. To achieve this, Trxs and Grxs of *E. coli*, *S. cerevisiae*, *Synechocystis* sp. PCC 6803, *D. melanogaster*, *A. thaliana*, and *H. sapiens* were selected as their solved structures were available in the database except for Grx of *D. melanogaster*. Therefore, we modelled Grx of *D. melanogaster* using a suitable template by the Swiss Model server [30]. We also modelled the Trx structure of *A. thaliana* as its 2D overlapped structure solved by NMR was available in the PDB database but it is difficult to conduct structural analysis with such structures. The modelled structures were validated by Ramachandran plot analysis where 99% of residues fall under the allowed region (data not shown).

Similar to previous findings [15, 34], the core of Trxs and Grxs was made up of β strands while the catalytic motif was positioned on the surface of two proteins (Figures 4 and 5; Supplementary Fig. 5). The anchor residues, which are crucial for the thermodynamic and redox properties of proteins
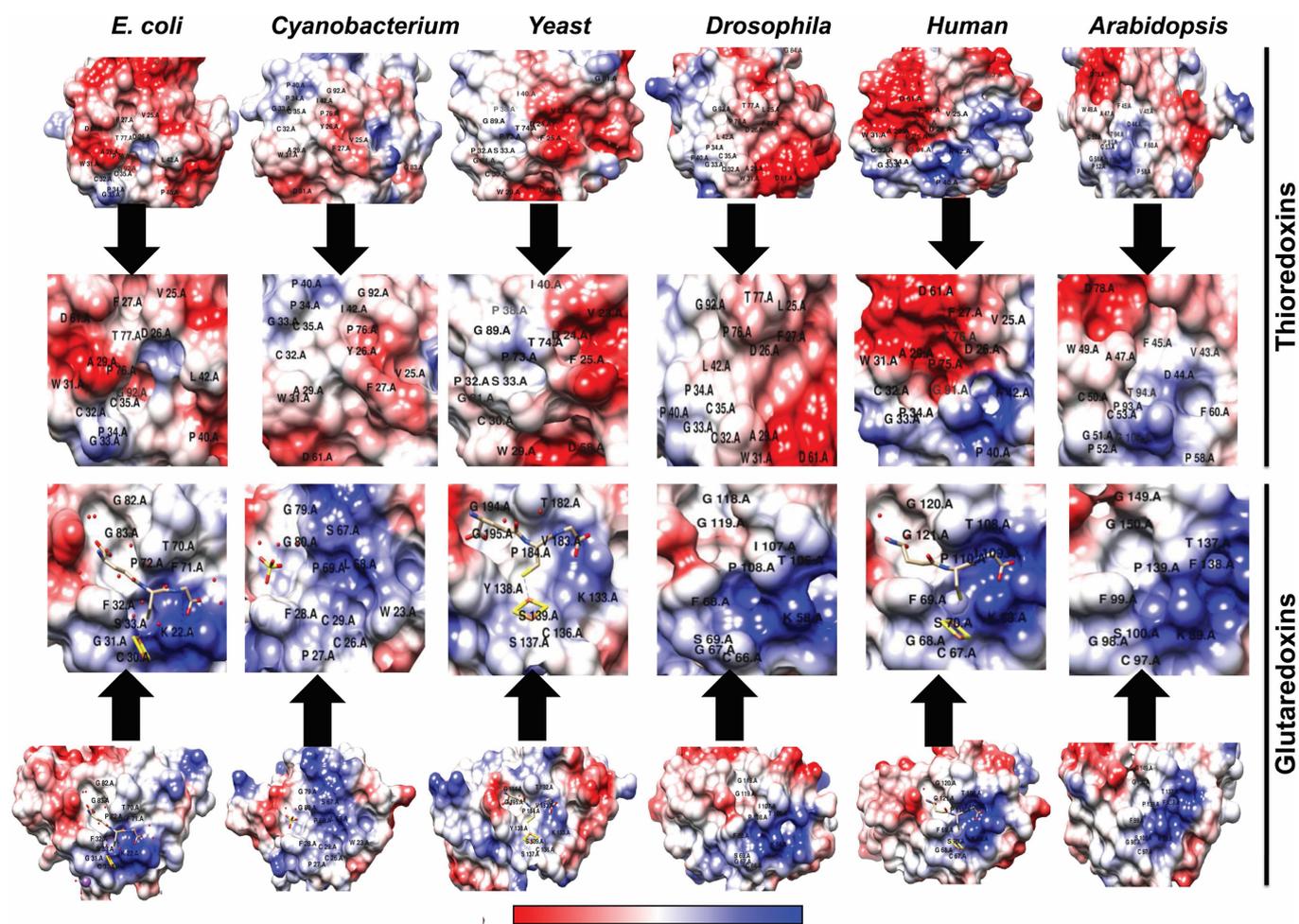

Figure 4. Electrostatic surface potential views of thioredoxins (Trxs) and glutaredoxins (Grxs) structures of different organisms representing three domains of life. The positive surface potential is denoted by blue color and negative surface potential is shown in red color. The protein surface possessing a neutral charge is shown in white color. Anchor residues are shown on the surface using a single letter code.

and their catalytic activity [35], were present on the surface in close proximity to the active site (Figure 4). The surface of Trxs was dominated by negatively charged and neutral amino acids, while Grxs had neutral and positively charged amino acids on their surface. However, in both cases, the active site was constituted by neutral (uncharged) amino acids (Figure 4). The catalytic site of Trxs was surrounded by negatively charged amino acids, while positively charged amino acids surrounded the catalytic site in Grxs (Figure 4). The surface of both Trxs and Grxs was dominated by hydrophilic amino acids together with a few hydrophobic residues (Figure 5). However, the catalytic site in all Trxs and Grxs was comprised of both hydrophilic and hydrophobic amino acid residues (Figure 5). The presence of hydrophilic residues on the surface of all studied Trxs and Grxs suggested their soluble nature. Thus, Trxs and Grxs have similar protein surface properties except for their electrostatic potential.

3.6. Trxs maintain domain architecture and topology which is not conserved in Grxs

Trxs and Grxs showed similar topology and had a conserved Trx/Grx fold which was primarily composed of 4β strands and at least 3α helices (Supplementary Fig. 5). For better understanding, we divided Trx/Grx fold into two domains, i.e., the N-terminal domain shown in orange color and the green-colored C-terminal domain (Supplementary Fig. 5). The N-terminal domain of both Trxs and Grxs was composed of 2 parallel β strands and one α helix while the C-terminal domain had 2 anti-parallel β strands and one α helix. Important to mention that β_1 and α_1 are extra

secondary structures found in all Trxs that were not part of the conserved fold (Supplementary Fig. 5a). The N-terminal and C-terminal domains of the two proteins were connected by an α helix; specifically α_3 helix in Trxs and α_2 helix in Grxs. The α_2 and α_4 helices of Trxs were located on one side of the central β-sheet while the α_3-helix was located on the opposite side. Similarly, α_1 and α_3 helices of Grxs were located on one side of the central β-sheet while α_2-helix was located on the opposite side (Supplementary Fig. 5b). The α_3-helix of Trxs was oriented perpendicularly to α_2 and α_4 helices, and α_2-helix of Grxs was oriented perpendicularly to α_1 and α_3 helices (Supplementary Fig. 5). All studied Trxs structures were typically containing 5β strands and 4α helices, and similarly, all studied Grxs structures possessed additional α helix other than 3α helices which was not part of the Trx/Grx fold (Supplementary Fig. 6). For example, Grxs of *E. coli*, *D. melanogaster*, *H. sapiens* and *A. thaliana* had two additional α helices; one at N-terminal and another at the C-terminal end. Similarly, cyanobacterial Grxs had one additional α helix at their C-terminal end while Yeast Grxs had two additional α helices; one at N-terminal and one at C-terminal, and two antiparallel β strands (Supplementary Fig. 6). In summary, topology analysis suggested that all Trxs strictly maintained domain architecture and topology which was not conserved in Grxs.

3.7. Trxs and Grxs are structurally more different in eukaryotes than prokaryotes due to altered helical conformation

We calculated structural similarity and computed root mean square deviation (RMSD) between equivalent atom positions after optimal
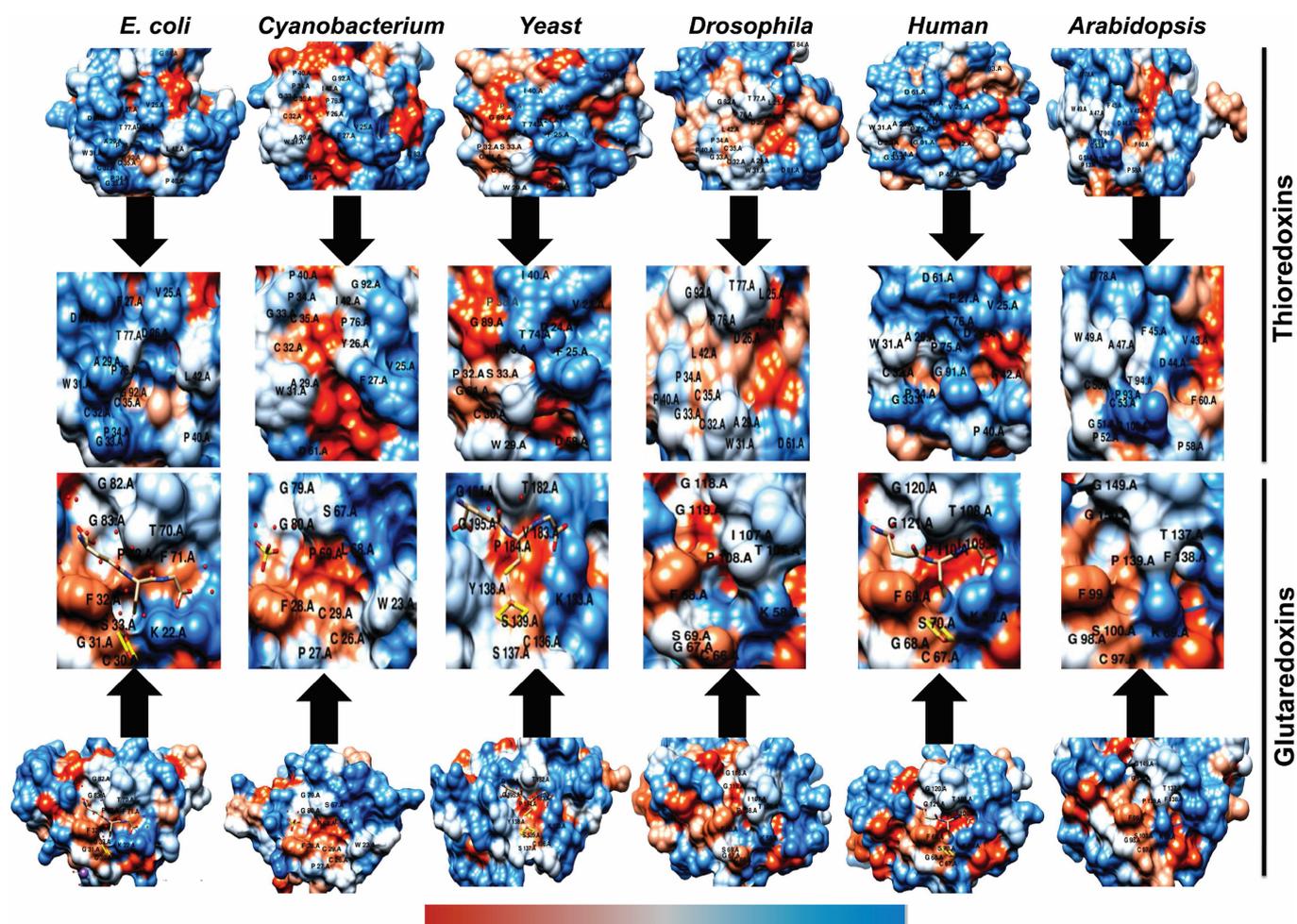

Figure 5. Hydrophobicity surface views of thioredoxins (Trxs) and glutaredoxins (Grxs) structures of different organisms representing three domains of life. The hydrophilic regions present on the surface of the proteins are shown in cyan color while a hydrophobic portion of the protein surface is shown in orange color. Anchor residues are shown on the surface using a single letter code.
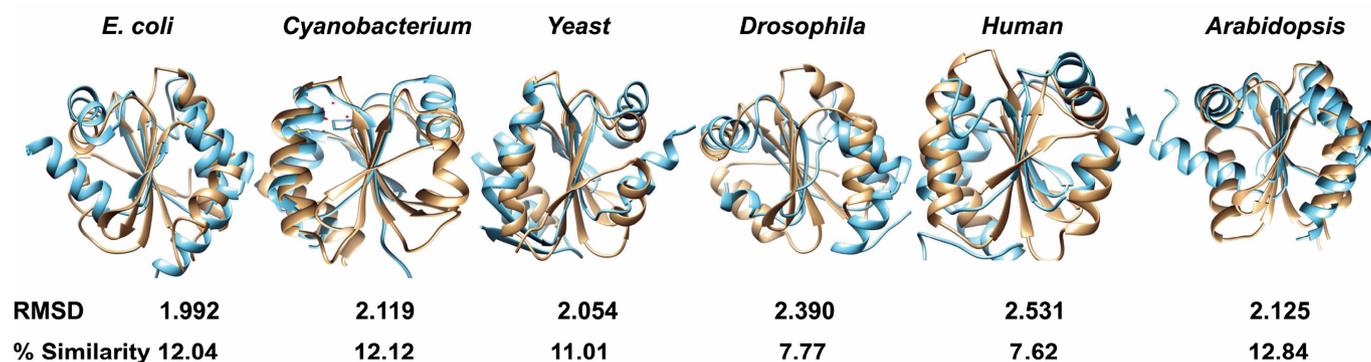
Figure 6. Comparative structure analysis of thioredoxins (Trxs) and glutaredoxins (Grxs). Secondary structure alignment and comparison between Trxs and Grxs in six different organisms representing three domains of life were made using superimposed structure. The structures of Trx and Grx from the same organism were superimposed and structural similarities were computed based on RMSD and percentage identity. Trx structures are shown in brown color while Grx structures are shown in cyan color.

superimposition of the different structures to decipher structural differences between Trxs and Grxs. Global RMSD value of superimposed Trx and Grx structures of the same organism was higher in eukaryotes than prokaryotes (Figure 6). This finding suggested that structures of Trxs and Grxs are more similar in prokaryotes than eukaryotes. The high RMSD values of superimposed Trx and Grx structures of eukaryotes suggested their structural differences and their involvement in distinct cellular processes (Figure 6). Importantly, Trxs and Grxs shared common conformation of core β strands with minimum local RMSD values (data not shown). However, two proteins largely differed in conformation of α helices which resulted high RMSD score (Figure 6). For example, conformation of α_1 helix was different in Trx and Grx of *Arabidopsis*, *Drosophila*, Human, and Yeast, while two proteins showed different conformation of α_2 and α_3 helices in *E. coli* and cyanobacterium. Also, eukaryotic Trxs and Grxs showed a higher conformational change in the helices than prokaryotic equivalents (Figure 6). We also calculated percent similarity which corresponds to the number of residues or percentage of total residues matched in the aligned structures. The percent similarity value depends on both sequence and structural conservation of the proteins, and therefore, proteins having similar sequence and secondary structure possess higher percent identity. The percentage identities were higher for prokaryotic Trx and Grx than eukaryotes which suggested that two proteins are more conserved in terms of their sequence and structure in prokaryotes than eukaryotes (Figure 6).

3.8. Trxs differ in distantly related organisms but conserved conformation of anchor residues

To assess the structural and functional diversity of Trxs, we conducted a comparative structural analysis of Trxs of different organisms based on percent similarity and RMSD value. All studied Trxs of prokaryotes and eukaryotes had conserved conformation of core β strands, helices and loops (Figure 7). However, RMSD values of superimposed Trxs structures were higher for distantly related organisms while lower RMSD values were obtained for closely related organisms (Figure 7). The *E. coli* Trx scored minimum RMSD value with cyanobacterial Trx and showed maximum percent similarity. Similarly, cyanobacterial Trx showed higher similarity to *E. coli* Trx than any other studied structures (Figure 7). Yeast Trx showed maximum similarity to human Trx; *Drosophila* Trx was similar to Human Trx and *vice versa*, and *A. thaliana* shared maximum similarity to Human Trx (Figure 7).

In addition to the conserved fold, the Trxs structures have amino acid residues known as anchor residues which include catalytic residues C32, G33, P34 and C35 along with D26, V25, F27, A29, W31, P40, F42, D61, P76, T77, and G92 [35]. We report that some of the anchor residues in studied Trxs were replaced by other residues having similar chemical properties (Figure 1; Supplementary Fig. 7). The 42nd position in *E. coli* Trx

was occupied by leucine. In cyanobacteria, tyrosine, isoleucine and alanine were present at the 26th, 42nd and 77th position while leucine replaced V25 and F42 in *D. melanogaster*. Similarly, F42 was replaced by I42 in the Trx of Yeast. We analyzed the conformation of these anchor residues by the superimposition of the structures (Supplementary Fig. 7). A low local RMSD score for anchor residues (data not shown) suggested that Trxs strictly conserved conformations of structurally and functionally important residues even in distantly related organisms despite substitutions and other structural differences.

3.9. Grxs possess flexible active site but strictly conserved conformation of cis-proline and GG-kink

The structural conformation of Grxs varied from organism to organism; however, it had same structural fold in all studied organisms (Figure 7). All Grxs shared a common conformation of core β strands and possessed more loop regions than Trxs (Figure 7; Supplementary Fig. 6). The presence of a higher number of loop regions in Grxs rendered higher flexible structure than Trxs. However, the conformation of helices and loops in Grxs was not conserved as compared to Trxs. The conformation of eukaryotic Grxs was more conserved in comparison to prokaryotic Grxs (Figure 7). Interestingly, *E. coli* Grx showed higher similarity with *Drosophila*, Human and *Arabidopsis* Grxs. Similar to Trxs, Grxs possessed structurally and functionally important anchor residues which included catalytic residues, glutathione binding residues, *cis*-proline, and GG-kink [15]. We analyzed the 3D conformation of catalytic residues, *cis*-proline, and GG-kink by the superimposition of studied Grxs structures, and did not include the GSH binding site due to variability in the amino acid residues (Supplementary Fig. 8). However, the position of amino acids in the GSH binding site was strictly conserved in all Grxs. In prokaryotes, GSH binding site residues were located at two positions, i.e., one upstream and the other downstream of the catalytic site, while eukaryotes had GSH binding site residues at three positions, i.e., one upstream and two downstream of the catalytic site. Results of the 3D conformational analysis of superimposed structures indicated that Grxs have a more flexible active site in comparison to Trxs while *cis*-proline and GG-kink residues conserved their 3D conformation during evolution (Supplementary Figs. 7 and 8). Also, the active site conformation of *E. coli* Grx was analogous to human and *Arabidopsis* Grxs while Grxs of human and *Arabidopsis* had similar conformation of catalytic residues (Supplementary Fig. 8).

Overall, structural analyses suggested that the conformation of active sites of Trxs is highly conserved in different organisms despite variation in their overall structural conformation. In contrast, the active sites of Grxs were more flexible in terms of their structural conformation. This suggested that Trxs are more specific for their substrates than Grxs. Thus, Grxs could target different substrates and show functional plasticity. This

Figure 7. Matrix-based comparative structure analysis of thioredoxins (Trxs) and glutaredoxin (Grxs) structures of different organisms. Secondary structure alignment and comparison of Trxs and Grxs in six different organisms representing three domains of life were made using superimposed structure. The structures of Trxs and Grxs found in the different organisms were superimposed based on the matrix and structural similarities were computed based on RMSD and percentage identity. Trxs and Grxs structures of different organisms are shown in different colors as given below the image.

could be also a possible explanation for Grxs to act as a backup system for Trxs depending on their electrostatic surface potential, especially, in prokaryotes where Trxs and Grxs are structurally more similar [5, 8].

4. Discussion

We conducted *in silico* amino acids sequence, phylogenetic, and comparative structural analyses to decipher the evolutionary and structural relationship between Trxs and Grxs of 12 organisms from three domains of life. Total 153 protein sequences, including 88 Grxs and 65 Trxs of distantly related organisms, were analyzed and outcomes suggested that Trxs possess more rigid and conserved catalytic site than Grxs (Tables 2 and 3; Supplementary Figs. 7 and 8). This observation supported that Trxs are more specific for their substrates than Grxs that have flexible active sites due to variation in their amino acids composition [2, 5, 16]. However, N-terminal cysteine residue of the active site was conserved in all Grxs which supported its proposed essentiality for initiating the reaction [15, 36]. In addition to active site residues, several other amino acid residues in Trxs and Grxs were found conserved during MSA analysis (Supplementary Fig. 1). Notably, these residues were abundant in Trxs than Grxs and could be required for maintaining 3D structure and redox properties of two proteins; however, this proposition needs to be experimentally validated.

Generally, Trxs have conserved CGPC active site motif, however, we report Trxs with different catalytic motifs [37]. The maximum variation in the active sites of Trxs was found in *A. thaliana* which had 2 CGGC (Trx4 and Trx5), 3 CPPC (Trx h3, h4 and h5), 1 CVPC (Trx h10), 1 CGSC (Trx1), 1 CASC (Trx2) and 1 CRKC (Trx2) Trxs in the cytosol (Table 2). Similarly, Trx2 and Trx5 of *A. veneficus* had CPYC and CPSC active sites, respectively, and *C. elegans* had CGHC active site motif in Trx5 (Table 2). We report presence of disulfide isomerase (PDI) family domain in Trx5 of *C. elegans*. PDI is a redox-active protein commonly found in eukaryotes and catalyze oxidative protein folding in the endoplasmic reticulum (ER) [38, 39]. PDI can also reduce disulfide bonds and prevent protein aggregation, and facilitate the folding of newly synthesized proteins by acting as chaperones [40]. These proteins usually contain redox-active multiple Trx domains containing CXXC active site motif and could also possess one or more redox inactive Trx-like domains [38]. Based on domain information, it is predicted that Trx5 of *C. elegans* could participate in oxidative protein folding, however, further investigation is required for the proposed role of Trx5 in this organism.

Notably, multimodular Grxs were observed in higher organisms while prokaryotes such as *E. coli*, *A. veneficus* and *S. elongatus* PCC 7942 had single modular Grxs (Table 3). However, it is important to mention that cyanobacterial class V and VI Grxs are multimodular proteins that were absent in freshwater *S. elongatus* PCC 7942. *S. elongatus* PCC 7942 possess

class I and II Grxs while class V and VI Grxs are exclusively found in a few marine cyanobacteria [15]. We report several multimodular Grxs where monothiol Grxs of eukaryotes contain one N-terminal Trx domain and one or more Grx domains similar to PICOT proteins (Table 3). PICOT from the plant contains three repeats of Grx-like domain, metazoan other than insect has two repeats while fungus contains only one domain of Grx (Table 3). PICOT proteins are glutaredoxin-3 or Protein Kinase C (PKC) interacting proteins that show homology to Trxs and Grx-PICOT-like proteins [41, 42]. We identified two HEAT repeats in Grx1 of *A. veneficus* and a GIY-YIG domain in Grx S16 of *A. thaliana* (Table 3). The HEAT repeat is a tandem repeat of 37–47 amino acids long module which was found in several cytoplasmic proteins [43]. HEAT repeat-containing proteins are involved in intracellular transport processes where HEAT repeat domain facilitates protein-protein interaction [44]. The GIY-YIG domain-containing proteins are involved in several cellular processes such as DNA repair and recombination, transfer of mobile genetic elements, genomic stability and restriction of foreign DNA [45, 46, 47]. Thus, the presence of HEAT and GIY-YIG domains in Grxs suggest their role in intracellular transport and maintenance of genomic DNA, respectively. However, further experimental evidence is required for this proposal. Together, these observations suggested that the prevalence of additional domains and flexibility of catalytic sites renders higher functional diversity to Grxs.

The analysis of amino acids composition gives an idea about the change in the frequency of occurrence of different amino acids in a family of proteins during evolution [48]. Also, the amino acids composition of a protein plays an important role in determining its structure, biological function and cellular localization. The flexibility in the frequency of occurrence of different amino acids was observed in both Trxs and Grxs, however, two proteins preserved approximately the same amount of non-polar amino acids in their structures (Figures 1 and 2; Supplementary Tables 1 and 2). The straight-chain non-polar amino acids are required for helix formation [49], and therefore, it is proposed that three helices of the Trx/Grx fold were maintained during evolution by preserving non-polar amino acids in both Trxs and Grxs of different organisms. Similarly, Trxs and Grxs had least amount of aromatic amino acids, particularly tryptophan (Figures 1 and 2; Supplementary Table 1 and 2). Phenylalanine, tyrosine and tryptophan are typically hydrophobic but compared to common hydrophobic residues such as leucine and valine, aromatic amino acids play important role in structural conformation [50]. For example, tyrosine and tryptophan contribute to hydrogen bonds formation while tryptophan is involved in the cation-π interaction. It is a strong non-covalent binding interaction that contributes to the secondary structure of proteins and protein-ligand interactions [50]. Thus, the low frequency of occurrence of aromatic amino acids is the characteristic feature of Trxs and Grxs proteins.

The specific requirement of cysteine for catalytic reaction [15, 36] could be responsible for their conserved frequency in both Trxs and Grxs (Figures 1 and 2; Supplementary Table 1 and 2). All studied Trxs and Grxs had almost similar percentage of different amino acids in their sequence, however, the presence or absence of one or more charged amino acids in their sequence resulted in a wide range of theoretical pI (Figures 1 and 2; Tables 2 and 3; Supplementary Tables 1 and 2). The MSA analysis confirmed a similar percentage of different amino acids in Trxs and Grxs but their position in the peptide chain was different (Figures 1 and 2; Supplementary Tables 1 and 2; Supplementary Figs. 1 and 2). This observation is important as Trxs and Grxs are found in subcellular compartments [1, 3, 5]. The change in the frequency of the amino acids and their position leads to a change in the theoretical pI value which could help proteins to adjust with the environment of subcellular compartments. However, a further experimental investigation by point mutation studies is required to support the proposition that the position of certain amino acids in a peptide chain can help proteins in adjusting to environment of their intracellular location.

The phylogenetic analysis of 153 Trxs and Grxs from 12 different organisms suggested that two proteins originated from a common ancestor and diverged later during evolution to form two groups of proteins (Figure 3). Also, monothiol and dithiol Grxs appeared as two separate clades which indicated their ancestral relationship, and based on our analysis, we propose that monothiol Grxs originated from dithiol Grxs (Figure 3). This proposal is supported by the previous study where the phylogenetic tree of Grxs was clearly divided into two distinct groups of dithiol and monothiol Grxs [51]. However, it is important to mention that in the previous study two groups of Grxs got separated from a common ancestor at the very beginning of the phylogenetic tree in contrast with the present study where Trxs were also part of analysis (Figure 3). Further detailed but separate phylogenetic analyses of Trxs and Grxs (Supplementary Figs. 3 and 4) supported the endosymbiotic theory that suggests that mitochondria and chloroplast of a eukaryotic cell originated from endosymbiosis of alpha-proteobacteria and cyanobacteria, respectively [52]. Also, the results of the phylogenetic study suggested that Trxs and Grxs sequences can be used to establish an evolutionary lineage in molecular systematics.

The plant possesses substrate-specific Trxs such as f, m, x, y, h, and o types that carry out diverse cellular functions in different cellular compartments [53]. Trx f, Trx m, Trx x and Trx y are plastidial Trxs, Trx o is mitochondrial while Trx h is a cytoplasmic Trx. Trx f activates fructose 1, 6–biposphatase (FBPase), Trx m activates malate dehydrogenase (MDH), and Trxs x and y catalyses 2-cys peroxiredoxins (Prx) and PrxQ. Trx o regulates the activity of enzymes involved in the TCA cycle while specific substrate(s) of Trx h is still unknown but due to the absence of any signal sequence, it is considered to control redox level in the cytoplasm [53]. Trxs f1 and f2 of *A. thaliana* were found close to the Trx4 and Trx5 of *A. veneficous* in phylogenetic analysis of Trxs, and therefore, it appears that plant Trx f has an archaebacterial origin (Supplementary Fig. 3). In contrast, other Trxs of *A. thaliana*, except Trx y and Trx m, appeared in the same clad with eukaryotic Trxs which suggested their eukaryotic origin (Supplementary Fig. 3). Trx y and Trx m were found in chloroplast and showed a closed relationship with cyanobacterial Trxs which supported the endosymbiotic origin of chloroplast from cyanobacteria. However, it is noticeable that some of the mitochondrial Trxs of *S. cerevisiae* and *A. thaliana* shared eukaryotic origin (Supplementary Fig. 3). This observation suggested that these mitochondrial Trxs were evolved later during evolution and their coding sequence were incorporated into the mitochondrial genome.

It was proposed earlier that both f- and h-type Trxs are of archaebacterial origin [54], however, here, we report that h-type Trx of *A. thaliana* showed eukaryotic ancestry (Supplementary Fig. 3). Similarly, six divergent sequences of Trxs of archaebacteria were reported to be originated from animal and eubacterial ancestors [54]. However, we report that not all but a few archaebacterial Trxs such as Trx1, Trx2 and Trx3 share an ancestral relationship with animals and eubacteria (Supplementary Fig. 3). The phylogenetic analysis of Grxs suggested that CC-type Grxs of *A. thaliana* are of eukaryotic origin while Grx3 of cyanobacteria shared a close relationship with Grx1 of archaebacteria as well as eukaryotic Grxs. Owing to the presence of two prokaryotic sequences in this clad, it is proposed that these Grxs share a common prokaryotic origin (Supplementary Fig. 4). The presence of chloroplastidial and mitochondrial Grxs in the same clade supported the endosymbiotic theory and their prokaryotic origin [51]. Similar to Trxs, some Grxs of *A. veneficous* shared ancestral relationship with eukaryotes and eubacteria which suggested their common origin.

The structural conformation of proteins impacts their function, and therefore, proteins of the same family conserved their overall structural conformation despite the difference in their amino acid sequence. All protein structures studied were found to have two regular states, i.e., α-helix and β-strand. The remaining unassigned regions known as an irregular state (coil) corresponded to a large number of different conformations (Supplementary Figs. 5 and 6). Notably, all studied Trxs and Grxs had a common structural Trx/Grx fold despite difference in their amino acids sequences (Figures 1 and 2). The robust structure of proteins despite difference in their amino acids suggested an error in gene

replication event which permitted variation in proteins during evolution [55, 56, 57, 58]. However, the two cysteine residues C32 and C35 of active site and their structural conformation were highly conserved in all Trxs (Figures 4 and 5; Supplementary Fig. 7). Similarly, three conserved proline were present in all studied Trxs (Figures 4 and 5; Supplementary Fig. 7). First proline was situated in the catalytic CGPC motif (P34) which is the key residue for reducing power as its substitution by a serine or a threonine affected the redox and stability properties of Trxs [59, 60, 61].

We report substitution at the 34th position of Trxs by residues such as tyrosine, serine, glycine, or lysine (Table 2). Notably, the substitution of P34 by amino acids having similar properties to proline did not alter the overall structure of the protein; however, the effect of these substitutions on the function of Trxs needs to be investigated. The second conserved proline, i.e., P40, introduces a kink in the α_2 helix that separates the active site CGPC motif from the rest of the helix (Figures 4 and 5; Supplementary Fig. 7). Thus, P40 helps in the proper positioning of the catalytic site in the α_2 helix. However, substitution of P40 destabilizes the structure of the protein without affecting the redox properties [62, 63]. The third proline, i.e., P76, was positioned on the opposite side of the CGPC active site motif and it was always found in cis-conformation (Figures 4 and 5; Supplementary Fig. 7). P76 is essential for maintaining the conformation of the active site and redox potential of the protein, and its substitution by alanine resulted in decreased catalytic efficiency of Trx [64].

The conserved threonine (T77) situated next to the P76 is involved in structuring the area opposite to the CGPC active site motif [65]. The G33 of active site motif CGPC helps in maintaining the conformation of the active site and also influence the redox potential of Trxs [65]. The G33 and P34 collectively provide a flat surface around the active site due to the absence of protruding side chains in these amino acids. The G84 and G92 determine the length of the β_5 strand while F12 present at N-terminal α_1 helix is required for correct positioning of the α_1 helix (Figures 4 and 5; Supplementary Fig. 7). Also, F12 together with F27 found at the C-terminal of β_2 strand and isoleucine and valine residues of the central β sheet create a hydrophobic site which was proposed to act as a site for interaction with other proteins [65]. The W31 residue is important for the thermodynamic stability of Trxs which interact with A29 located in a turn through van der Waals interaction (Supplementary Fig. 7) [66]. A29 is known to prevent a shift in the position of the indole side chain of W31 due to the small size of alanine. However, important to mention that substitution of W31 by alanine resulted in swapping of domain dimer that caused loss of biochemical activities of Trxs-fold containing proteins [66]. The conserved D26 and K57 residues are part of a charged region present between the core β sheet and the kinked α_2 helix, and D26 is the key residue that is considered to activate the nucleophilic activity of C35 of the active site motif [61, 68].

All studied Grxs shared a common Trx/Grx fold with Trxs proteins but in this study we report presence of additional secondary structures in Grxs that varied in different organisms (Supplementary Fig. 6). This observation suggested that the topology of Grxs varies from organism to organism but domain architecture is always maintained. Also, dithiol and monothiol Grxs conserved their residues which were important for their structural conformation and function (Figures 4, 5, 6, and 7; Supplementary Fig. 8). These residues included catalytic site motif, glutathione binding site, *cis*-proline and GG-kink which were conserved in all classes of Grxs (Supplementary Fig. 8). Importantly, the residues constituting glutathione binding site may vary among different classes of Grxs, however, their position in a peptide chain remain highly conserved. The conserved *cis*-proline and GG-kink are signature residues present in all Grxs, however, the structural and functional importance of these residues in Grxs of different organisms is still not well studied (Figures 4 and 5; Supplementary Fig. 8). The *cis*-proline is known to play a significant role in protein folding and redox dynamics while two glycines in the GG-kink are required for proper orientation of the $\alpha3$ helix [15, 69]. The substitution of either Gly115 or Gly116 with valine or serine residues resulted in the loss of yeast Grx5 function [69].

Generally, Trxs and Grxs were found to have cysteine residues only in their catalytic site, however, human cytosolic Trx1 contains three additional structural cysteine residues (C62, C69, and C73). These structural cysteine residues play important role in substrate recognition, dimerization, and regulate the activity of thioredoxin reductase [70]. In summary, despite variation in the amino acid composition, anchor residues and their conformation is highly conserved in Trxs and Grxs of distantly related organisms. This observation suggests that specific position of highly conserved anchor residues is important for proper functioning of Trxs and Grxs.

Trxs and Grxs were characterized by the presence of conserved Trx/Grx fold and anchor residues, however, two proteins surprisingly had opposite electrostatic surface potential (Figure 4). All Trxs had negative while Grxs had positive electrostatic surface potential that surrounded the catalytic sites constituted by uncharged amino acids (Figure 4). The electrostatic surface potential of a protein dictates the binding of substrates and/or ligands, and recently, Trxs and Grxs were classified based on their surface electrostatic charges [5]. Interestingly, distantly related Trxs and Grxs clustered together when an automated clustering of electrostatic surface potential properties was done for their functional classification and function prediction [5]. Here, it is proposed that recognition of target proteins could be regulated by attractive and repulsive electrostatic surface potential. However, further studies targeting substrates of Trxs and Grxs is required to validate this proposal. It should be noted that electrostatic surface potential is an emerging global property of Trxs and Grxs [5] which can be used together with structural similarity (Figures 6 and 7) for functional classification and explaining their substrate specificity and/or redundancy.

5. Conclusions

Trxs and Grxs show variation in their amino acid sequence, however, diversity in sequence does not alter their structural fold even in distantly related organisms. The structural conformation of Trxs along with their anchor residues are conserved throughout the evolution whereas the structure and active site of Grxs are more flexible. The dynamic catalytic site and presence of additional module in Grxs could permit them to exhibit versatile substrate catalysis and reaction mechanism. Also, flexibility in catalytic site and overlapping electrostatic surface potential could permit Grxs to act as a backup system for Trxs, especially in prokaryotes where two proteins are more similar than eukaryotes.

Declarations

Author contribution statement

Soumila Mondal: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Shailendra P. Singh: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e10776.

Acknowledgements

References

[1] E.M. Hanschmann, J.R. Godoy, C. Berndt, C. Hudemann, C.H. Lillig, Thioredoxins, glutaredoxins, and peroxiredoxins-molecular mechanisms and health significance: from cofactors to antioxidants to redox signaling, Antioxidants Redox Signal. 19 (2013) 1539–1605.

[2] Y. Meyer, C. Belin, V. Delorme-Hinoux, J.P. Reichheld, C. Riondet, Thioredoxin and glutaredoxin systems in plants: molecular mechanisms, crosstalks, and functional significance, Antioxidants Redox Signal. 17 (2012) 1124–1160.

[3] F.J. Cejudo, A.J. Meyer, J.P. Reichheld, N. Rouhier, J.A. Traverso, Thiol-based redox homeostasis and signaling, Front. Plant Sci. 5 (2014) 266.

[4] F.T. Ogata, V. Branco, F.F. Vale, L. Coppo, Glutaredoxin: discovery, redox defense and much more, Redox Biol. (2021), 101975.

[5] M. Gellert, M.F. Hossain, F.J.F. Berens, L.W. Bruhn, C. Urbainsky, V. Liebscher, C.H. Lillig, Substrate specificity of thioredoxins and glutaredoxins-towards a functional classification, Heliyon 5 (2019), e02943.

[6] T.C. Laurent, E.C. Moore, P. Reichard, Enzymatic synthesis of deoxyribonucleotides: IV. Isolation and characterization of thioredoxin, the hydrogen donor from *Escherichia coli* B, J. Biol. Chem. 239 (1964) 3436–3444.

[7] E.C. Moore, P. Reichard, L. Thelander, Enzymatic synthesis of deoxyribonucleotides: V. Purification and properties of thioredoxin reductase from *Escherichia coli* B, J. Biol. Chem. 239 (1964) 3445–3452.

[8] A. Holmgren, Glutathione-dependent synthesis of deoxyribonucleotides. Characterization of the enzymatic mechanism of *Escherichia coli* glutaredoxins, J. Biol. Chem. 254 (1979) 3672–3678.

[9] L. Michelet, M. Zaffagnini, S. Morisse, F. Sparla, M.E. Pérez-Pérez, F. Francia, S.D. Lemaire, Redox regulation of the Calvin–Benson cycle: something old, something new, Front. Plant Sci. 4 (2013) 470.

[10] Q. Wu, J. Yang, N. Cheng, K.D. Hirschi, F.F. White, S. Park, Glutaredoxins in plant development, abiotic stress response, and iron homeostasis: from model organisms to crops, Environ. Exp. Bot. 139 (2017) 91–98.

[11] M. Balsera, B.B. Buchanan, Evolution of the thioredoxin system as a step enabling adaptation to oxidative stress, Free Radic. Biol. Med. 140 (2019) 28–35.

[12] L. Vogelsang, K.J. Dietz, Regulatory thiol oxidation in chloroplast metabolism, oxidative stress response and environmental signaling in plants, Biochem. J. 477 (2020) 1865–1878.

[13] M.I. Berggren, B. Husbeck, B. Samulitis, A.F. Baker, A. Gallegos, G. Powis, Thioredoxin peroxidase-1 (peroxiredoxin-1) is increased in thioredoxin-1 transfected cells and results in enhanced protection against apoptosis caused by hydrogen peroxide but not by other agents including dexamethasone, etoposide, and doxorubicin, Arch. Biochem. Biophys. 392 (2001) 103–109.

[14] J. Couturier, J.P. Jacquot, N. Rouhier, Evolution and diversity of glutaredoxins in photosynthetic organisms, Cell. Mol. Life Sci. 66 (2009) 2539–2557.

[15] S. Mondal, V. Kumar, S.P. Singh, Phylogenetic distribution and structural analyses of cyanobacterial glutaredoxins (Grxs), Comput. Biol. Chem. 84 (2020), 107141.

[16] A.P. Fernandes, A. Holmgren, Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system, Antioxidants Redox Signal. 6 (2004) 63–74.

[17] N. Rouhier, Plant glutaredoxins: pivotal players in redox biology and iron–sulphur centre assembly, New Phytol. 186 (2010) 365–372.

[18] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 35 (2006) D61–D65.

[19] P. Schürmann, B.B. Buchanan, The ferredoxin/thioredoxin system of oxygenic photosynthesis, Antioxidants Redox Signal. 10 (2008) 1235–1274.

[20] C.S. Yu, Y.C. Chen, C.H. Lu, J.K. Hwang, Prediction of protein subcellular localization, Proteins 64 (2006) 643–651.

[21] P. Horton, K.J. Park, T. Obayashi, N. Fujita, H. Harada, C.J. Adams-Collier, K. Nakai, WoLF PSORT: protein localization predictor, Nucleic Acids Res. 35 (2007) W585–W587.

[22] E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, The proteomics protocols, 2005, pp. 571–607.

[23] J.D. Thompson, D.G. Higgins, T.J. Gibson, W. Clustal, Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[24] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.

[25] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, X. Mega, Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.

[26] H. White, Maximum likelihood estimation of misspecified models, Econometrica (1982) 1–25.

[27] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences, Comput. Appl. Biosci. 8 (1992) 275–282.

[28] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evol. Lett. 39 (1985) 783–791.

[29] P.W. Rose, A. Prlić, A. Altunkaya, C. Bi, A.R. Bradley, C.H. Christie, R.K. Green, The RCSB protein data bank: integrative view of protein, gene and 3D structural information, Nucleic Acids Res. 52 (2016) 1000.

[30] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.

[31] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.

[32] A. Stivala, M. Wybrow, A. Wirth, J.C. Whisstock, P.J. Stuckey, Automatic generation of protein structure cartoons with Pro-origami, Bioinformatics 27 (2011) 3315–3316.

[33] A. Holmgren, Thioredoxin and glutaredoxin systems, J. Biol. Chem. 264 (1989) 13963–13966.

[34] L.W. Schultz, P.T. Chivers, R.T. Raines, The CXXC motif: crystal structure of an active-site variant of *Escherichia coli* thioredoxins, Acta Crystallogr. D Biol. Crystallogr. 55 (1999) 1533–1538.

[35] A. Holmgren, Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide, Structure 3 (1995) 239–243.

[36] J.H. Bushweller, F. Aaslund, K. Wuethrich, A. Holmgren, Structural and functional characterization of the mutant *Escherichia coli* glutaredoxin and its mixed disulfide with glutathione, Biochemistry 31 (1992) 9288–9293.

[37] Y. Meyer, J.P. Reichheld, F. Vignols, Thioredoxins in *Arabidopsis* and other plants, Photosynth. Res. 86 (2005) 419–433.

[38] B. Wilkinson, H.F. Gilbert, Protein disulfide isomerase, Biochim. Biophys. Acta, Proteins Proteomics 1699 (2004) 35–44.

[39] A.D. Johnston, P.R. Ebert, The redox system in *C. elegans*, a phylogenetic approach, J. Toxicol. 2012 (2012), 546915.

[40] C.W. Gruber, M. Čemažar, B. Heras, J.L. Martin, D.J. Craik, Protein disulfide isomerase: the structure of oxidative folding, Trends Biochem. Sci. 31 (2006) 455–464.

[41] N. Isakov, S. Witte, A. Altman, H.D. Picot, A highly conserved protein domain that is often associated with thioredoxin and glutaredoxin modules, Trends Biochem. Sci. 25 (2000) 537–539.

[42] S. Witte, M. Villalba, K. Bi, Y. Liu, N. Isakov, A. Altman, Inhibition of the c-Jun N-terminal kinase/AP-1 and NF-kappaB pathways by PICOT, a novel protein kinase C-interacting protein with a thioredoxin homology domain, J. Biol. Chem. 275 (2000) 1902–1909.

[43] M.A. Andrade, P. Bork, HEAT repeats in the Huntington's disease protein, Nat. Genet. 11 (1995) 115–116.

[44] M.R. Groves, N. Hanlon, P. Turowski, B.A. Hemmings, D. Barford, The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs, Cell 96 (1999) 99–110.

[45] J.C. Kowalski, M. Belfort, M.A. Stapleton, M. Holpert, J.T. Dansereau, S. Pietrokovski, S.M. Baxter, V. Derbyshire, Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings, Nucleic Acids Res. 27 (1999) 2115.

[46] S. Dunin-Horkawicz, M. Feder, J.M. Bujnicki, Phylogenomic analysis of the GIY-YIG nuclease superfamily, BMC Genom. 7 (2006) 98.

[47] F. Zannini, A. Moseler, R. Bchini, T. Dhalleine, A.J. Meyer, N. Rouhier, J. Couturier, The thioredoxin-mediated recycling of *Arabidopsis thaliana* GRXS16 relies on a conserved C-terminal cysteine, BBA—Gen. Subj. 1863 (2019) 426–436.

[48] D.J. Brooks, J.R. Fresco, Increased frequency of cysteine, tyrosine and phenylalanine residues since the last universal ancestor, Mol. Cell. Proteomics 1 (2002) 125–131.

[49] S. Padmanabhan, R.L. Baldwin, Straight-chain non-polar amino acids are good helix-formers in water, J. Mol. Biol. 219 (1991) 135–137.

[50] D.A. Dougherty, Cation-π interactions involving aromatic amino acids, J. Nutr. 137 (2007) 1504S–1508S.

[51] J. Sagemark, T.H. Elgán, T.R. Bürglin, C. Johansson, A. Holmgren, K.D. Berndt, Redox properties and evolution of human glutaredoxins, Proteins 68 (2007) 879–892.

[52] V. Zimorski, C. Ku, W.F. Martin, S.B. Gould, Endosymbiotic theory for organelle origins, Curr. Opin. Microbiol. 22 (2014) 38–48.

[53] E. Gelhaye, N. Rouhier, N. Navrot, J.P. Jacquot, The plant thioredoxin system, Cell. Mol. Life Sci. 62 (2005) 24–35.

[54] Y. Meyer, L. Verdoucq, F. Vignols, Plant thioredoxins and glutaredoxins: identity and putative roles, Trends Plant Sci. 4 (1999) 388–394.

[55] A.P. Joseph, G. Agarwal, S. Mahajan, J.C. Gelly, L.S. Swapna, B. Offmann, B. Schneider, A short survey on protein blocks, Biophys. Rev. 2 (2010) 137–145.

[56] F. Katzen, J. Beckwith, Transmembrane electron transfer by the membrane protein DsbD occurs via a disulfide bond cascade, Cell 103 (2000) 769–779.

[57] P.W. Haebel, D. Goldstone, F. Katzen, J. Beckwith, P. Metcalf, The disulfide bond isomerase DsbC is activated by an immunoglobulin-fold thiol oxidoreductase: crystal structure of the DsbC–DsbDα complex, EMBO J. 21 (2002) 4774–4784.

[58] B. Rost, Protein structures sustain evolutionary drift, Folding Des. 2 (1997) S19–S24.

[59] F.K. Gleason, C.J. Lim, M. Gerami–Nejad, J.A. Fuchs, Characterization of *Escherichia coli* thioredoxins with altered active site residues, Biochemistry 29 (1990) 3701–3709.

[60] G. Krause, J. Lundstrom, J.L. Barea, C. Pueyo de la Cuesta, A. Holmgren, Mimicking the active site of protein disulfide-isomerase by substitution of proline 34 in *Escherichia coli* thioredoxins, J. Biol. Chem. 266 (1991) 9494–9500.

[61] P.T. Chivers, R.T. Raines, General acid-base catalysis in the active site of *Escherichia coli* thioredoxins, Biochemistry 36 (1997) 15810–15816.

[62] F. de Lamotte-Guéry, C. Pruvost, P. Minard, M.A. Delsuc, M. Miginiac-Maslow, J.M. Schmitter, P. Decottignies, Structural and functional roles of a conserved proline residue in the alpha 2 helix of *Escherichia coli* thioredoxins, Protein Eng. 10 (1997) 1425–1432.

[63] A. Chakrabarti, S. Srivastava, C.P. Swaminathan, A. Surolia, R. Varadarajan, Thermodynamics of replacing an alpha-helical Pro residue in the P40S mutant of *Escherichia coli* thioredoxins, Protein Sci. 8 (1999) 2455–2459.

[64] F.K. Gleason, Mutation of conserved residues in *Escherichia coli* thioredoxin: effects on stability and function, Protein Sci. 1 (1992) 609–616.

[65] H. Eklund, F.K. Gleason, A. Holmgren, Structural and functional relations among thioredoxins of different species, Proteins 11 (1991) 13–28.

[66] A. Garcia–Pino, S. Martinez–Rodriguez, K. Wahni, L. Wyns, R. Loris, J. Messens, Coupling of domain swapping to kinetic stability in a thioredoxin mutant, J. Mol. Biol. 385 (2009) 1590–1599.

[68] V. Menchise, C. Corbier, C. Didierjean, M. Saviano, E. Benedetti, J.P. Jacquot, A. Aubry, Crystal structure of the wild-type and D30A mutant thioredoxin h of *Chlamydomonas reinhardtii* and implications for the catalytic mechanism, Biochem. J. 359 (2001) 65–75.

[69] G. Bellí, J. Polaina, J. Tamarit, M.A. de la Torre, M.T. Rodríguez-Manzaneque, J. Ros, E. Herrero, Structure-function analysis of yeast Grx5 monothiol glutaredoxin defines essential amino acids for the function of the protein, J. Biol. Chem. 277 (2002) 37590–37596.

[70] W.H. Watson, J. Pohl, W.R. Montfort, O. Stuchlik, M.S. Reed, G. Powis, D.P. Jones, Redox potential of human thioredoxin 1 and identification of a second dithiol/disulfide motif, J. Biol. Chem. 278 (2003) 33408–33415.