Heliyon 5 (2019) e02943

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Substrate specificity of thioredoxins and glutaredoxins – towards a functional classification

Manuela Gellert [a,1], Md Faruq Hossain [a,1], Felix Jacob Ferdinand Berens [a,b], Lukas Willy Bruhn [a,b], Claudia Urbainsky [a], Volkmar Liebscher [b], Christopher Horst Lillig [a,*]

[a] Institute for Medical Biochemistry and Molecular Biology, University Medicine Greifswald, Germany
[b] Institute for Mathematics and Informatics, University of Greifswald, Germany

ARTICLE INFO

Keywords:
Mathematical biosciences
Biocomputational method
Biomolecules
Molecular docking
Gromov-Wasserstein distance
Protein-protein interaction
Redox signaling
Electrostatics
Thioredoxin
Glutaredoxin

ABSTRACT

The spatio-temporal reduction and oxidation of protein thiols is an essential mechanism in signal transduction in all kingdoms of life. Thioredoxin (Trx) family proteins efficiently catalyze thiol-disulfide exchange reactions and the proteins are widely recognized for their importance in the operation of thiol switches. Trx family proteins have a broad and at the same time very distinct substrate specificity – a prerequisite for redox switching. Despite of multiple efforts, the true nature for this specificity is still under debate. Here, we comprehensively compare the classification/clustering of various redoxins from all domains of life based on their similarity in amino acid sequence, tertiary structure, and their electrostatic properties. We correlate these similarities to the existence of common interaction partners, identified in various previous studies and suggested by proteomic screenings. These analyses confirm that primary and tertiary structure similarity, and thereby all common classification systems, do not correlate to the target specificity of the proteins as thiol-disulfide oxidoreductases. Instead, a number of examples clearly demonstrate the importance of electrostatic similarity for their target specificity, independent of their belonging to the Trx or glutaredoxin subfamilies.

1. Introduction

Redox modifications of cysteinyl and also methionyl side chains are a vital part of numerous signal transduction pathways as well as the reaction cycle of essential metabolic enzymes [1, 2, 3, 4]. Many of these redox reactions are directly or indirectly catalyzed by members of the Trx family of proteins. This group of proteins share a common basic structural motif – the Trx fold. Their active sites, in most cases consisting of two cysteinyl residues separated by two amino acids (Cys-X-X-Cys), are the basis of their redox activity [5]. Proteins of this family are known to catalyze the reduction of disulfides in target proteins, the de- or glutathionylation of proteins, they catalyze the oxidative folding of proteins and are able to reduce redox modifications like sulfenic acids [4]. Trx family proteins are encoded in essentially all genomes and are localized in all compartments of eukaryotic cells, *e.g.* the cytosol, ER, mitochondria, nucleus, and plastids – often in multiple isoforms. Most members of the family have a broad, but distinct substrate specificity. The nature of this specificity is the focus of this work.

The Trx family divides into subfamilies, two of the major groups are the Trxs themselves and the glutaredoxins (Grxs). Trxs catalyze thiol-disulfide exchange reactions and the trans-nitrosylation of thiol groups [4, 6]. The disulfide formed in their consensus Cys-Gly-Pro-Cys active site during their reaction cycle is reduced by specific reductases named thioredoxin reductases (TrxRs) [7, 8]. The members of one of the Grx subfamilies (dithiol Grxs, with a consensus active site of Cys-Pro-Tyr-Cys) catalyze thiol-disulfide oxidoreductions as well, however, when oxidized these proteins are reduced by the tripeptide glutathione (GSH). The mechanisms of these reactions have been discussed in great detail before, see for instance [4, 9, 10, 11]. The members of a second subclass of the Grxs (monothiol Grxs, with a consensus Cys-Gly-Phe-Ser active site) do not catalyze thiol-disulfide exchange reactions at significant rates. Instead, they function in the regulation of iron metabolism or in the transfer of iron-sulfur centers [12, 13, 14].

Traditionally, Grxs and Trxs were named in each organism in order of their discovery, for instance in mammals, the firstly discovered cytosolic Trx1 [15] and the later discovered mitochondrial Trx2 [11]. An other

* Corresponding author.
E-mail address: horst@lillig.de (C.H. Lillig).
[1] These authors contributed equally.

https://doi.org/10.1016/j.heliyon.2019.e02943
Received 22 May 2019; Received in revised form 15 November 2019; Accepted 25 November 2019

example are the yeast Grxs 1–8, summarized in [16]. This historical naming, however, does not include any information on structural or functional differences with the Trx family of proteins. A more advanced classification and naming system based on the active site sequences was presented for plants Grxs and became widely accepted [17]. This nomenclature defines three classes, *i.e.* class I, that contains the largely redox-active dithiol Grxs, class II, including all monothiol Grxs, and class III, including the land plant-specific CC-type Grxs, also known as ROXYs [18].

In most species, including bacteria, fungi, mammals, and plants, multiple Trx family proteins are present in the same compartment, prompting questions on overlapping functions. Various proteomic studies, also summarized in this work, indicate a rather high degree of substrate specificity for each member of the family, with only some overlapping substrate/target proteins. Various attempts have been made to understand the substrate specificity and reactivity of the Trx family proteins. Previous suggestions primarily addressed the thermodynamics of the reaction, including the nucleophilicity of the more N-terminal active site thiol, the differences in redox potential, and entropic changes during the reaction [19, 20, 21]. Recently, based on the analysis of *E. coli* phosphoadenylyl sulfate (PAPS) reductase that can react with many, but not all Trxs and Grxs [22, 23], we proposed that different electrostatic properties of the redoxins govern their target specificity and reactivity [24].

In this work, we provide a detailed comparison on the coherence between the similarity of redoxins in (1^{st}) primary structure, (2^{nd}) tertiary structure, and (3^{rd}) electrostatic properties. Where possible, we correlate these different methods of clustering/classification to known functions of the redoxins. We have focused on the redoxins encoded in the human genome and all redoxins from various species with experimentally determined structures deposited in the protein data base. Our results provide further evidence for the importance of the electrostatic properties of the proteins for their distinct target specificity. This clustering may allow a new functional classification of the redoxins and may enable the prediction of common functions and interactions partners.

2. Methods and procedures

2.1. Structures and molecular modeling

Structures, when available, were obtained from the protein data bank (https://www.rcsb.org); the PDB entries used are listed in the supplementary table. Molecular modeling was performed using the Swiss Model web server [25, 26, 27, 28, 29]. The final model was chosen from structures modeled with different templates that displayed the highest sequence identity with the target protein based on the quality assessment provided, *i.e.* the lowest QMEAN with no major outlier in the global or local quality estimates of Cβ, all atom, solvation, or torsion. The individual template structures used and the QMEAN values are summarized in the supplementary table (structures).

2.2. Sequence and structure comparison

Sequences of the proteins were obtained from the uniprot resources. In case of multi-domain proteins, the sequences encoding the redoxin domains only were mostly extracted as annotated in the respective uniprot entries, i.e. based on PROSITE-ProRule [30]. In some critical cases, for instance human nucleoredoxin, multiple sequence alignments were performed including sequences from various species. These typically share a higher degree of homology within the functional domains and a lower in the joining peptides. Primary structure alignments and the generation of the corresponding distance trees were performed with the CLC sequence viewer (Qiagen bioinformatics, Hilden, Germany) and Clustal omega [31]. The three dimensional structures were aligned using UCSF chimera [32] (MatchMaker) including structure-based multiple sequence alignments. From these primary structure and 3-D structure alignments, the corresponding distance trees were generated with the CLC sequence viewer applying the neighbor joining method and the Jukes-Cantor protein distance measure.

2.3. Electrostatic calculations

The structures in the PDB files were aligned in the desired orientation using UCSF chimera. The electrostatic properties of the proteins were computed from the pdb files as follows: the reconstruction of any missing atoms, the addition of hydrogens, the assignment of atomic charges and radii was performed using pdb2pqr with the amber force field [33]. The electrostatic parameters were calculated using the Adaptive Poisson-Boltzmann Solver (APBS) [34] within the vmd (visual molecular dynamics) software package [35]. The following parameters were used: 150 mM mobile ions, solvent dielectric constant: 78.54, temperature: 298.15 K. Images were rendered depicting the secondary structures of the proteins (with the N-terminal active site cysteinyl residues facing towards the camera perspective), the electrostatic potential mapped to the surface of the proteins (from -4 in red to 4 K T·e^{-1} in blue), and the isosurfaces of the electrostatic potential at -1 in red and 1 K T·e^{-1} in blue. These pictures were used to generate a summary picture using ImgageMagick. All steps following the 3D alignment of the structures were automatized with the help of scripts and a graphical interface. These can be obtained from: https://github.com/WillyBruhn/MutComp.

2.4. Electrostatic distances and clustering

Here, we compared the 3-dimensional isosurfaces of both the negative and positive electrostatic potential using the Gromov-Wasserstein-distance [36, 37]. Solving this problem, is NP-hard as the objective function is not convex. However, three lower bounds for the Gromov-Wasserstein-distance can be calculated in polynomial time. Our empirical tests demonstrated that not all points of the isosurface shall be calculated, instead we limited the sample to n points randomly distributed on the isosurfaces and calculate the lower bound for them. This was repeated m times, the obtained values were summarized in form of a histogram. This comparison was performed pairwise for all proteins. To get a measure of similarity between the histograms, the earth--mover's-distance was used [38]. For the hierarchical clustering, the unweighted pair group method with arithmetic mean (UPGMA) was used. This method yields the mean distance between all points from the new cluster to all points of another cluster. The result of this clustering were displayed in form of a dendrogram. Further details of this mathematical approach will be presented elsewhere. All code of the software produced here (C++, R, scripts) can be obtained free of charge and open source here: https://github.com/BerensF/ComparingProteins.

2.5. Interactome data and comparison

Interactome data were retrieved from the IntAct database [39] (as of Sept. 2018) as well as the BioGRID resources [40] (Vers. 3.4, as of Sept. 2018). ID mapping was performed using the uniprot resources (https://www.uniprot.org). All entries are listed with their unique UniprotKB ID. When available, additional resources for interactions – not yet listed in the upper mentioned databases – were included, for instance from some dedicated publications. All entries identified are listed in the supplementary tables (interactome). The matrix of common interactions as well as the Venn diagrams were computed using R (https://www.r-project.org) from RStudio (https://www.rstudio.com).

3. Results and discussion

All presently established classifications of Trx family proteins are based on the comparison of, and the clustering according to primary structures. We know, however, that these systems do not, or at best partially, reflect the various functions of the proteins. As examples, some

Table 1. Redoxins and redoxin domains encoded in the human genome. loc: localization, c: cytosol, e: endoplamic reticulum, g: golgi apparatus, l: lipid membrane, m: mitochondria, n: nucleus, o: outside of the cell, secreted; '.d' marks individual domains. n.a.: not analyzed (as no information on the individual domains were available).

Protein name	ID	Active site	pdb	Functions	loc.	Interactome (number)
Homo sapiens glutaredoxins						
Grx1	P35754	CPYC	1b4q, ...	(de)-glutathionylation	c/n	24
Grx2a/c	Q9NS18	CSYC	2fls, ...	Fe/S, redox sensor [52]	c/m/n	64 [53]
Grx3.d2	O76003	CGFS	3zyw	Fe/S, iron metabolism [54, 55]	c	98
Grx3.d3		CGFS	2yan			
Grx5	Q86SX6	CGFS	2wul, ...	Fe/S biogenesis [56, 57]	m	22
TrxR1v3.d1	Q16881-1	CTRC			c	n.a.
TrxR3.d1	Q86VQ6	CPHS	2h8q		c	n.a.
GrxCR1.d2	A8MXD5	FERC			c	9
PTGES2.d1	Q9H7Z7	CPFC	2pbj	prostaglandin synthesis [58]	c/l	31
SH3BGRL3	Q9H299	KSQQ	1sj6	cancer progression [59]	c/n	10
Homo sapiens thioredoxins						
Trx1	P10599	CGPC	1ert, ...	electron donor, redox signaling, ... [4]	c/n/o	247 [60]
Trx2	Q99757	CGPC	1wh4, ...		m	102
Grx3.d1	O76003	APQC	2diy, ...	Fe/S, iron metabolism [54, 55]	c	see above
Nrx.d2	Q6DKJ4	CPPC		redox signaling	c/n	629 [61]
Nxnl1	Q96CM4	CPQC			n/l	4
Nxnl2	Q5VZ03	CAPS			c	2
Txnl1	O43396	CGPC	1gh2, ...		c/n	62
Txnl4A	P83876	DPTC	1qgv, ...		n	39
Txnl4B	Q9NX01	DPVC	3gix, ...		n	21
Tmx1	Q9H3N1	CPAC	1x5e		e/l/o	79
Tmx2	Q9Y320	SNDC	2dj0		l/o	103
Tmx4	Q9H1E5	CPSC			l/o	25
Txndc2	Q86VQ3	CGPC		spermatogenesis [62, 63]	c	4
Txndc3	Q8N427	CGPC		spermatogenesis [64]	c	8
Txndc6	Q86XW9	CGPC		microtuble dynamics, cancer progression [65, 66]	c	-
Txndc8	Q6A555	CGPC		spermatogenesis [67]	c/g	-
Txndc9	O14530	TFRC		protein complex assembly [68]	c/n	91
Txndc11.d1	Q6PKC3	CGQS			e/l	83
Txndc11.d2		CGFC				
Txndc12	O95881	CGAC	1sen, ...	PDI	e	30
Txndc15	Q96J42	CRFS		ciliogenesis [69]	l/o	44
Txndc16	Q9P2K2	QAVS		meningioma-associated antigen [70]	e/s	29
Txndc17	Q9BRA2	CPDC	1wou	disulfide and cystine reduction, denitrosylation [71, 72]	c	39
Qsox1	O00391	CGHC	3q6q, ...	disulfide formation [73]	g/o	17
Qsox2	Q6ZRP7	CGHC		disulfide formation [73]	c/l/n/o	28

only distantly related Trxs and Grxs share overlapping functions, while some closely related Grxs do not. The disulfide formed in *E. coli* PAPS reductase during its catalytic cycle can be reduced by the distantly related Trx1 and Grx1 (primary sequence identity: 19.6%), but not by Grx3 (sequence identity to Grx1: 41.7%) [23, 41]. The primary aim of this study is to provide a thorough comparison between the clustering and classification according to primary structure, 3-D structure, and electrostatic characteristics. Further more, we aim to evaluate the practicality of the comparison of electrostatic properties for the functional classification of Trx family proteins and the prediction of functions.

In the first part, we focus on the Trx family proteins from human (see Table 1). Functions of the proteins were deduced from the literature, identified interacting proteins were collected from the literature, our own data sets, and some major interaction databases, *i.e.* IntAct and BioGRID (see methods section). These sources summarize studies dedicated to single interactions as well as proteomic approaches. If available, 3-D structures were obtained from the protein data base (pdb). Missing structures were obtained by homology modeling. The electrostatic calculations were performed using pdb2pqr and APBS implemented in vmd. For the comparison and clustering of the electrostatic similarities, we have adopted new strategies based on Gromov-Wasserstein distances of lower boundaries and Earth-Movers distances. These computational strategies are outlined in the methods section.

3.1. Human Grxs and Trxs

The human genome encodes about 9 Grxs or Grx domain- containing proteins and 24 Trxs or Trx domain-containing proteins (see Table 1). From the in total 35 Trx-fold domains in these 33 proteins, the 3-D structure of 20 had already been determined experimentally. The remaining 15 were predicted by homology modeling, see supplementary Table, sheet 1 – modeling. Similar to the comparison of the plant redoxins, the human redoxins divide into similar groups in the trees generated from primary and tertiary structure (Figure 1A and B). The subgroup of the Grxs both show the separate monothiol and dithiol Grx groups. The Trx group is quite diverse, however, the Nrxs and the Trxs with the consensus active site motif Cys-Gly-Pro-Cys (minus mitochondrial Trx2) clearly separate from the others in both analyses. The electrostatic characteristics (Figure 2) define two major groups (Figure 1C). Group 'I' contains Trxs and Trx domain-containing proteins exclusively. Group 'II' contains all of the Grxs and Grxs domains next to some Trx-related proteins and domains.

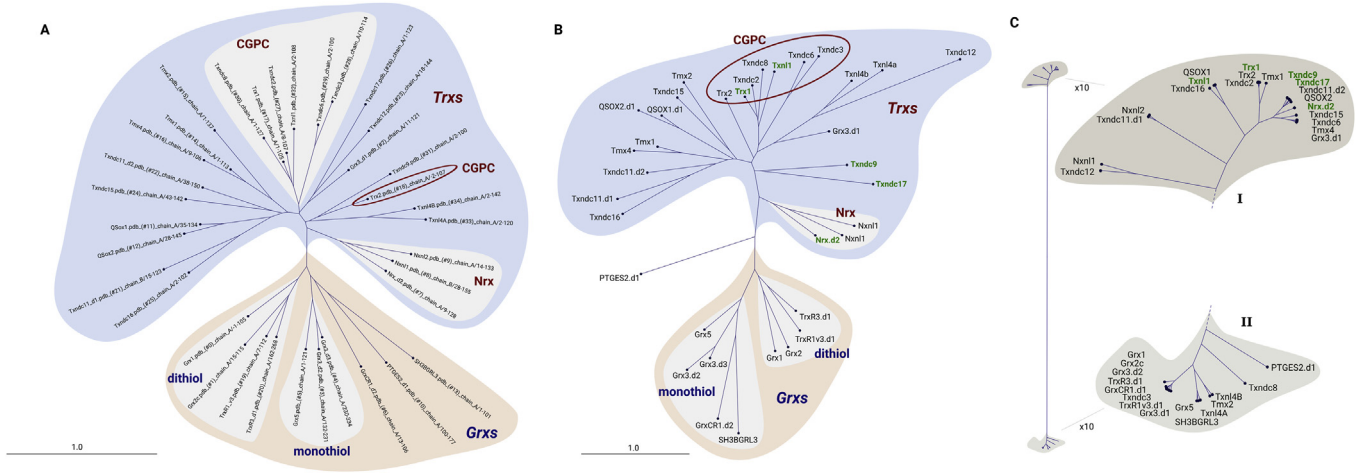

Figure 1. Clustering of human redoxins. (**A**) Phylogram based on primary structure comparison, computed by Clustal Omega and CLC sequence viewer. (**B**) Similarity tree based on the similarity of the 3D structures extracted from the pdb and generated by homology modeling; the tree was computed using UCSF Chimera and the CLC sequence viewer. (**C**) The electrostatic similarity of the whole proteins was computed as outlined in the methods section; the tree was generated using 'R'. The protein abbreviations highlighted in green are referred to in the main text. The Trx proteins with a Cys-Gly-Pro-Cys active site motif were highlighted with a red circle in B.

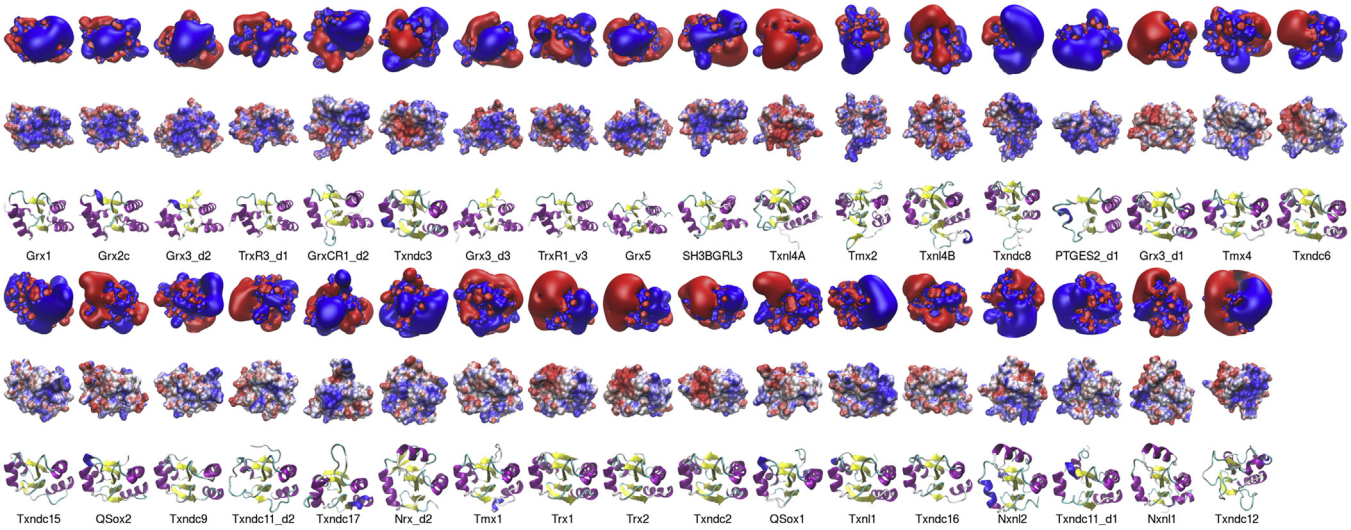

Figure 2. Electrostatic features of the active site contact areas of the human redoxins. The first rows depict the electrostatic potential isosurfaces at +/- 1 K T·e^{-1}. The second row depicts the electrostatic potential at +/- 4 K T·e^{-1} mapped to the water-accessible surface of the proteins. Blue: positive, red negative potential. The third row depicts the proteins in cartoon models, helices are colored in purple, sheets in yellow. The proteins were arranged with the N-terminal active site thiol in the middle of the models. The electrostatic similarity of the whole proteins was computed as outlined in the methods section.

Using various sources, see Table 1, we collected almost 2000 potential interactions partners of these 33 proteins (all entries are available in the supplementary tables, sheet H.s. interactome). The pair-wise comparison of common targets in these sets is summarized in Figure 3A. The largest degree of overlapping interacting proteins was found between the thioredoxins Trx1, Txnl1, Txndc9, Txndc17, and Nrx, also summarized in the Venn diagram depicted in Figure 3B. This suggests some overlapping functions of various pairs of these five proteins. Triple and quadruple overlaps, on the other hand, are rare and no common potential interaction partner of all five of these redoxins was identified nor suggested so far. In the trees generated from the primary structure and tertiary structure comparisons, these five proteins of the Trx subfamily are localized in three different branches, separated by large distances (Figure 1A and B, green labels), only two of them contain the consensus Cys-Gly-Pro-Cys active site motif (Trx1 and Txnl1, see Table 1). In the tree based on their similarity in electrostatic properties, these five proteins are all localized close to each other in the terminal branches of cluster 'I' (Figure 1C, green labels). The glutaredoxins Grx1, Grx2, Grx3, and Grx5 share only three potential interaction partners, one between Grx1 and Grx2, two between Grx3 and Grx5, out of a total of 210 suggested target proteins (see Figure 3C). The primary reasons for this may be their different subcellular localization, dithiol Grx1 and monothiol Grx3 are cytosolic, dithiol Grx2 and monothiol Grx5 primarily mitochondrial.

This study primarily focused on redox-interactions of the redoxins with target proteins. The main reason for this was the availability of data. However, it has been suggested early on that the Trx-fold domains might act as platform for protein-protein interactions, e.g. as processivity factor of T7 DNA polymerase [42] or as the basis for their redox-independent chaperon activity, summarized in [43]. Our study implies that redox-inactive redoxins (mostly domains) will show a similar target specificity as electrostatically similar redox-active redoxins. The N-terminal Trx domain of human Grx3 (Grx3_d1), for instance, contains the 'active' site motif Ala-Pro-Gln-Cys, hence it cannot catalyze thiol-disulfide exchange reactions. Electrostatically, this domain is most similar to the redox-active Grx1 and Grx2, as well as the Grx-domains of

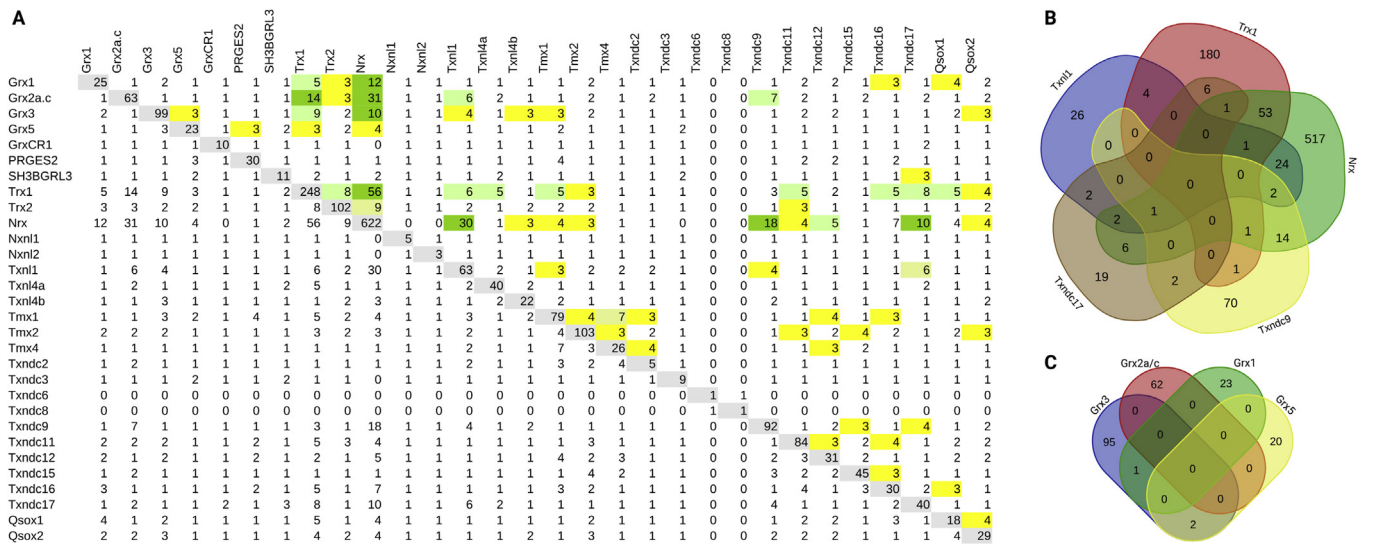

Figure 3. Common interaction partners between the human redoxins. (**A**) Pair-wise comparison between all human redoxins. The total numbers of potential interactions partners collected from various data sources is depicted with gray background in the diagonal; yellow background: 3–4 common interaction partners; light green background: 5–9 common interaction partners; green background: ≥ 10 common interaction partners. The full list of interaction partners can be found in the supplementary tables. (**B–C**) Venn diagrams of the overlapping potential interactions partners between Trx1, Nrx, Txndc9, Txndc17, and Txnl1 (B) as well as Grx1, Grx2, Grx3, and Grx5 (B).

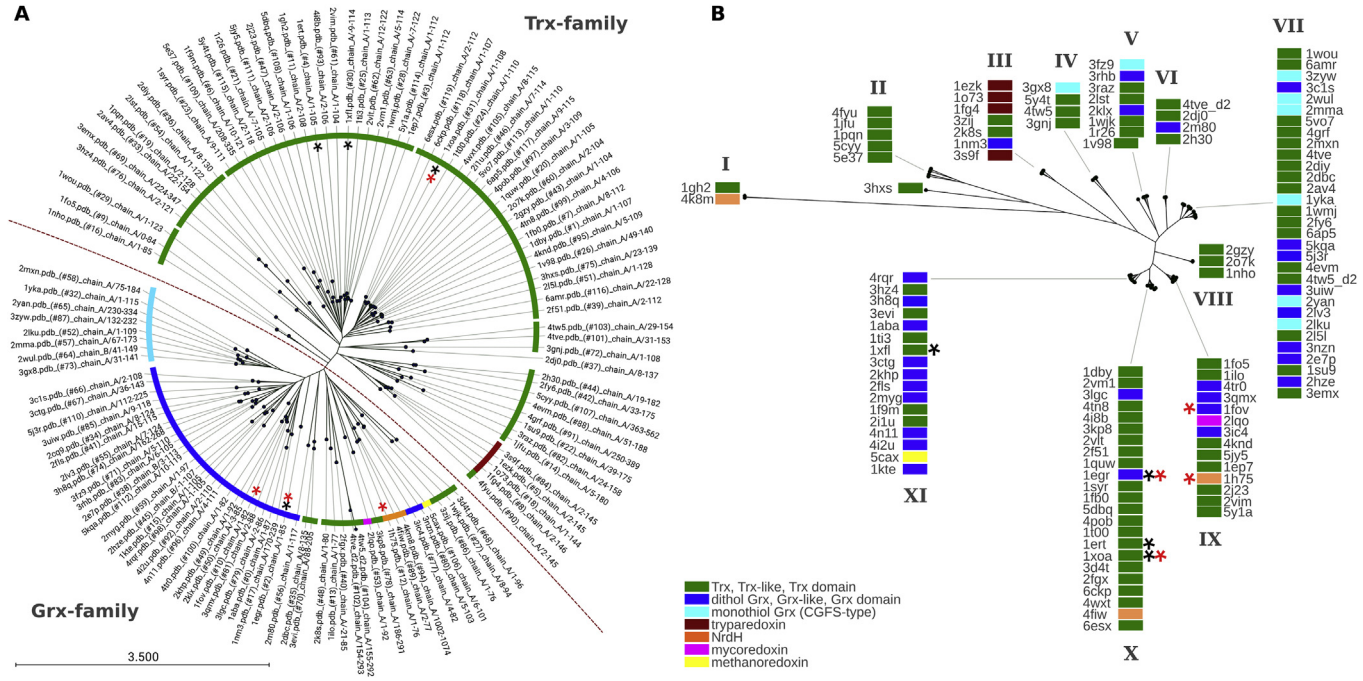

Figure 4. Clustering of all representative redoxins in the pdb. (**A**) Phylogram based on primary structure comparison, computed by Clustal Omega and CLC sequence viewer. The dashed red line separates the Trx and Grx subfamilies. (**B**) The electrostatic similarity of the whole proteins was computed as outlined in the methods section; the tree was generated using 'R'. The red asterisks mark proteins interacting with *E. coli* RNR, the black asterisks proteins interacting with *E. coli* PAPS reductase. The color code is included in the figure. Further information on the protein structures can be obtained from the supplementary Table, sheet 5.

TrxR1 transcript variant 3 [44] and TrxR3 [45], see Figure 1C (cluster II). It remains to be established whether these proteins and domains share a similar pattern of interacting proteins or other substrates as implied by our study.

3.2. All representative redoxin structures from the pdb

For a more comprehensive clustering of Trx-family proteins, we have selected all Trx- and Grx-fold structures from the pdb for analysis. For the electrostatic analysis, we selected only non-mutated proteins and, if required, extracted the Trx-/Grx-domains. Other molecules included in some of the structures, *i.e.* cofactors or water, were excluded. For the present analysis, we did not include other Trx-family members, such as DsbA/B/C proteins, protein disulfide isomerases, GSH peroxidases, or arsenate reductases. Our final collection included 119 structures (see supplementary Table, sheet 5 – pdb structures) from all domains of life, including some phage/virus-encoded proteins. For these structures, we generated both primary structure and electrostatic similarity trees, see

(caption on next page)

Figure 5. Electrostatic features of all representative redoxins in the pdb. The first rows depict the electrostatic potential isosurfaces at +/- 1 K T·e^{-1}. The second row depicts the electrostatic potential at +/- 4 K T·e^{-1} mapped to the water-accessible surface of the proteins. Blue: positive, red negative potential. The third row depicts the proteins in cartoon models, helices are colored in purple, sheets in yellow. The pdb entry code of the structures is indicated in the fourth row. The proteins were arranged with the N-terminal active site thiol in the middle of the models. The electrostatic similarity of the whole proteins was computed as outlined in the methods section.

Figure 4. The sequence-based tree clearly separates the structures into the Grx and Trx subfamilies (Figure 4A), independent of their sometimes confusing annotation in the pdb and sequence databases. Within the Grx branch, the genuine monothiol Grxs (Cys-Gly-Phe-Ser) form a well defined side branch, as well as the dithiol Grxs. The remaining structures include redoxins such as mycoredoxin, methanoredoxin, and NrdHs proteins. The Trx subfamily contains six distinct groups, two eukaryotic, three bacterial, and the tryparedoxin branches (Figure 4A). In contrast to sequence similarity, the electrostatic similarity tree separates into eleven branches, marked as I-XI in Figure 4B, see also Figure 5, of which most include structures from both the Trx and Grx subfamilies. None of these groups overlap with a branch of the sequence-based tree.

So what can be said about the correlation between the two clustering methods and the functional interaction of the different redoxins with target proteins? The two first discovered functions of Trxs and Grxs offer some insights. Trx1 (pdb: 1xoa) from *E. coli* was first discovered as electron donor for ribonucleotide reductase (RNR) [46]. Later studies demonstrated that this function could also be fulfilled by Grx1 (1egr) [47, 48], Grx3 (1fov) [49], and NrdH (1h75) [50] from *E. coli*. In the sequence based tree, these four proteins can be found in three distant branches (Figure 4A, red asterisks). In the electrostatic similarity tree, however, these four proteins can be found in the very close neighbouring branches IX and X (Figures 4B and 5, red asterisks). The requirement of Trxs for sulfate reduction was first described in yeast [51]. Subsequently, Trx1 was also demonstrated to be a cosubstrate of PAPS reductase in *E. coli*. Similar to RNR, Grx1 can replace Trx1 *in vivo* and *in vitro* in this function [23, 41], however, Grx3 and NrdH failed to do so [23, 24]. *E. coli* PAPS reductase can be reduced by a number of redoxins from various species, while some others cannot interact with the protein [22, 24]. Positive interactions partners are, for instance human and *Arabidopsis thaliana* Trx1 (1ert) and TrxH1 (1xfl), respectively. Again, these functionalities and non-functionalities cannot be predicted or explained by their similarities in primary sequence (Figure 4A, black asterisks). In the electrostatic similarity tree, three of the four proteins are located in cluster 'X', one (A.t. TrxH1) in cluster 'XI'. It should also be mentioned that cluster 'XI' contains proteins that cannot interact productively with PAPS reductase, *i.e.* human Grx2 (2fls), and T4 Grx (1aba) [24]. It appears that – at least some – proteins included in the electrostatic characteristics cluster 'X' can interact with both PAPS reductase and RNR, proteins of cluster 'IX' also with RNR, but not PAPS reductase. These two examples clearly support our hypothesis that the electrostatic similarity between the redoxins correlates to their profile of interacting proteins or may even guide these specific interactions [24].

For this study, the points analyzed for electrostatic similarity were distributed randomly on the surface of the proteins. One might argue that more weight on the properties of the immediate contact surface surrounding the active site could even further improve the model. To test this, we have developed a strategy to extract the electrostatic isosurfaces of the areas surrounding the N-terminal active site residue, that is usually the cysteinyl residue that forms an intermediate mixed-disulfide with the (redox-) target proteins and analyzed them in the same way as the features of the whole proteins. This strategy and the results obtained when it was applied to the 'representative structures from the pdb' data set were summarized in the supplementary material (suppl. Figure 1 and text). In brief, the hierarchical clustering of the electrostatic similarities in the area close to the active site did not reflect the functions of some redoxins as electron donor of PAPS reductase or RNR anywhere near as good as the clustering of the electrostatic similarities of the whole proteins. The redoxins that were previously shown to donate electrons to PAPS reductase clustered all within the distance of 9.8% of the maximum distance of all redoxins, the ones donating electrons to RNR within 12.7% of the maximum distance (Figure 4B) when the features of the whole proteins were compared. When only the electrostatic features of the active site faces of the proteins were compared, these redoxins did not cluster in close proximity and the relative distances increased to 37.4% and 62.6% of the maximum distance for PAPS reductase and RNR, respectively (suppl. Figure 1, nodes a and b). These results indicate that the global properties of the proteins may play a more important role than previously assumed. Moreover, they favor a model of redoxin-target interaction in which the recognition of the two proteins is controlled by attractive and repulsive electrostatic forces that, presumably, take part in pre-orientation of the two proteins before they can form a productive encounter complex, rather than contact surface complementarity only. Further experimental studies will have to address this hypothesis.

4. Conclusions

Here, we evaluated the practicality of a mathematical model for the automated clustering of the electrostatic properties of proteins for the functional classification of Trx family proteins and the prediction of functions. The analysis of the human, and pdb-wide redoxin structures clearly demonstrate that primary and tertiary structure (backbone) similarity do not correlate to the target specificity of the proteins as thiol-disulfide oxidoreductases and neither does their redox potential, see [24]. Instead, the examples of the human and pdb-wide redoxins clearly demonstrate the importance of the electrostatic properties of the whole protein for target specificity and discrimination. The mathematical model evaluated here is the first step towards an automated analysis and comparison of electrostatic properties of a large number of protein structures.

Declarations

Author contribution statement

Manuela Gellert: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data.

Md Faruq Hossain: Performed the experiments; Analyzed and interpreted the data.

Felix Jacob Ferdinand Berens, Lukas Willy Bruhn: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Claudia Urbainsky: Performed the experiments.

Volkmar Liebscher: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Christopher Horst Lillig: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

This work was supported by Deutsche Forschungsgemeinschaft, German Research Foundation (DFG): Li984/3-2, GRK1947-A1. We

acknowledge support for the Article Processing Charge from the DFG (393148499) and the Open Access Publication Fund of the University of Greifswald.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Data associated with this study has been deposited at GitHub under the URLs: https://github.com/WillyBruhn/MutComp and https://github.com/BerensF/ComparingProteins.

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e02943.

References

[1] P. Ghezzi, V. Bonetto, M. Fratelli, Thiol disulfide balance: from the concept of oxidative stress to that of redox regulation, Antioxidants Redox Signal. 7 (2005) 964–972.
[2] Y.M.W. Janssen-Heininger, B.T. Mossman, N.H. Heintz, H.J. Forman, B. Kalyanaraman, T. Finkel, J.S. Stamler, S.G. Rhee, A. van der Vliet, Redox-based regulation of signal transduction: principles, pitfalls, and promises, Free Radic. Biol. Med. 45 (2008) 1–17.
[3] A. Drazic, J. Winter, The physiological role of reversible methionine oxidation, Biochim. Biophys. Acta 1844 (2014) 1367–1382.
[4] 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.
[5] J.L. Martin, Thioredoxin–a fold for all reasons, Struct. Lond. Engl. 1993 3 (1995) 245–250.
[6] R. Sengupta, A. Holmgren, The role of thioredoxin in the regulation of cellular processes by S-nitrosylation, Biochim. Biophys. Acta (2011).
[7] E.S. Arnér, A. Holmgren, Physiological functions of thioredoxin and thioredoxin reductase, Eur. J. Biochem. FEBS 267 (2000) 6102–6109.
[8] E.S.J. Arnér, Focus on mammalian thioredoxin reductases - important selenoproteins with versatile functions, Biochim. Biophys. Acta 1790 (2009) 495–526.
[9] J.J. Mieyal, M.M. Gallogly, S. Qanungo, E.A. Sabens, M.D. Shelton, Molecular mechanisms and clinical implications of reversible protein S-glutathionylation, Antioxidants Redox Signal. 10 (2008) 1941–1988.
[10] C.H. Lillig, C. Berndt, Glutaredoxins in thiol/disulfide exchange, Antioxidants Redox Signal. 18 (2013) 1654–1665.
[11] G. Spyrou, E. Enmark, A. Miranda-Vizuete, J. Gustafsson, Cloning and expression of a novel mammalian thioredoxin, J. Biol. Chem. 272 (1997) 2936–2941.
[12] M.T. Rodríguez-Manzaneque, J. Tamarit, G. Bellí, J. Ros, E. Herrero, Grx5 is a mitochondrial glutaredoxin required for the activity of iron/sulfur enzymes, Mol. Biol. Cell 13 (2002) 1109–1121.
[13] S. Bandyopadhyay, F. Gama, M.M. Molina-Navarro, J.M. Gualberto, R. Claxton, S.G. Naik, B.H. Huynh, E. Herrero, J.P. Jacquot, M.K. Johnson, et al., Chloroplast monothiol glutaredoxins as scaffold proteins for the assembly and delivery of [2Fe-2S] clusters, EMBO J. 27 (2008) 1122–1133.
[14] U. Mühlenhoff, S. Molik, J.R. Godoy, M.A. Uzarska, N. Richter, A. Seubert, Y. Zhang, J. Stubbe, F. Pierrel, E. Herrero, et al., Cytosolic monothiol glutaredoxins function in intracellular iron sensing and trafficking via their bound iron-sulfur cluster, Cell Metabol. 12 (2010) 373–385.
[15] E.C. Moore, A thioredoxin — thioredoxin reductase system from rat tumor, Biochem. Biophys. Res. Commun. 29 (1967) 264–268.
[16] E. Herrero, G. Bellí, C. Casa, Structural and functional diversity of glutaredoxins in yeast, Curr. Protein Pept. Sci. 11 (2010) 659–668.
[17] N. Rouhier, E. Gelhaye, J.-P. Jacquot, Plant glutaredoxins: still mysterious reducing systems, Cell. Mol. Life Sci. CMLS 61 (2004) 1266–1277.
[18] N. Gutsche, C. Thurow, S. Zachgo, C. Gatz, Plant-specific CC-type glutaredoxins: functions in developmental processes and stress responses, Biol. Chem. 396 (2015) 495–509.
[19] G. Roos, N. Foloppe, J. Messens, Understanding the pK(a) of redox cysteines: the key role of hydrogen bonding, Antioxidants Redox Signal. 18 (2013) 94–127.
[20] M.M. Gallogly, D.W. Starke, J.J. Mieyal, Mechanistic and kinetic details of catalysis of thiol-disulfide exchange by glutaredoxins and potential mechanisms of regulation, Antioxidants Redox Signal. 11 (2009) 1059–1081.
[21] P.B. Palde, K.S. Carroll, A universal entropy-driven mechanism for thioredoxin-target recognition, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 7960–7965.
[22] J.D. Schwenn, U. Schriek, PAPS-reductase from Escherichia coli: characterization of the enzyme as probe for thioredoxins, Z. Für Naturforschung C J. Biosci. 42 (1987) 93–102.
[23] C.H. Lillig, A. Prior, J.D. Schwenn, F. Aslund, D. Ritz, A. Vlamis-Gardikas, A. Holmgren, New thioredoxins and glutaredoxins as electron donors of 3'-phosphoadenylylsulfate reductase, J. Biol. Chem. 274 (1999) 7695–7698.
[24] C. Berndt, J.-D. Schwenn, C.H. Lillig, The specificity and efficiency of thioredoxins and glutaredoxins is determined by electrostatic and geometric complementarity, Chem. Sci. (2015).
[25] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, SWISS-MODEL: an automated protein homology-modeling server, Nucleic Acids Res. 31 (2003) 3381–3385.
[26] P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, Bioinforma. Oxf. Engl. 27 (2011) 343–350.
[27] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. Gallo Cassarino, M. Bertoni, L. Bordoli, et al., SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, Nucleic Acids Res. 42 (2014) W252–258.
[28] M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli, T. Schwede, Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology, Sci. Rep. 7 (2017) 10480.
[29] S. Bienert, A. Waterhouse, T.A.P. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, The SWISS-MODEL Repository-new features and functionality, Nucleic Acids Res. 45 (2017) D313–D319.
[30] C.J.A. Sigrist, E. De Castro, P.S. Langendijk-Genevaux, V. Le Saux, A. Bairoch, N. Hulo, ProRule: a new database containing functional and structural information on PROSITE profiles, Bioinforma. Oxf. Engl. 21 (2005) 4060–4066.
[31] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539.
[32] 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.
[33] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, Nucleic Acids Res. 35 (2007) W522–525.
[34] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10037–10041.
[35] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (33–38) (1996) 27–28.
[36] F. Mémoli, Gromov–wasserstein distances and the metric approach to object matching, Found. Comput. Math. 11 (2011) 417–487.
[37] F. Mémoli, The Gromov-Wasserstein distance: a brief overview, Axioms 3 (2014) 335–341.
[38] H. Ling, K. Okada, An efficient Earth Mover's Distance algorithm for robust histogram comparison, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 840–853.
[39] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N.H. Campbell, G. Chavali, C. Chen, N. del-Toro, et al., The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases, Nucleic Acids Res. 42 (2014) D358–D363.
[40] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, BioGRID: a general repository for interaction datasets, Nucleic Acids Res. 34 (2006) D535–539.
[41] M. Russel, P. Model, A. Holmgren, Thioredoxin or glutaredoxin in Escherichia coli is essential for sulfate reduction but not for deoxyribonucleotide synthesis, J. Bacteriol. 172 (1990) 1923–1929.
[42] D.F. Mark, C.C. Richardson, Escherichia coli thioredoxin: a subunit of bacteriophage T7 DNA polymerase, Proc. Natl. Acad. Sci. U. S. A. 73 (1976) 780–784.
[43] C. Berndt, C.H. Lillig, A. Holmgren, Thioredoxins and glutaredoxins as facilitators of protein folding, Biochim. Biophys. Acta BBA - Mol. Cell Res. 1783 (2008) 641–650.
[44] A.-K. Rundlöf, M. Janard, A. Miranda-Vizuete, E.S.J. Arnér, Evidence for intriguingly complex transcription of human thioredoxin reductase 1, Free Radic. Biol. Med. 36 (2004) 641–656.
[45] D. Su, S.V. Novoselov, Q.-A. Sun, M.E. Moustafa, Y. Zhou, R. Oko, D.L. Hatfield, V.N. Gladyshev, Mammalian selenoprotein thioredoxin-glutathione reductase. Roles in disulfide bond formation and sperm maturation, J. Biol. Chem. 280 (2005) 26491–26498.
[46] 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.
[47] A. Holmgren, Hydrogen donor system for Escherichia coli ribonucleoside-diphosphate reductase dependent upon glutathione, Proc. Natl. Acad. Sci. U. S. A. 73 (1976) 2275–2279.
[48] A. Holmgren, Glutathione-dependent synthesis of deoxyribonucleotides. Purification and characterization of glutaredoxin from Escherichia coli, J. Biol. Chem. 254 (1979) 3664–3671.
[49] F. Aslund, B. Ehn, A. Miranda-Vizuete, C. Pueyo, A. Holmgren, Two additional glutaredoxins exist in Escherichia coli: glutaredoxin 3 is a hydrogen donor for ribonucleotide reductase in a thioredoxin/glutaredoxin 1 double mutant, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 9813–9817.
[50] A. Jordan, F. Aslund, E. Pontis, P. Reichard, A. Holmgren, Characterization of Escherichia coli NrdH. A glutaredoxin-like protein with a thioredoxin-like activity profile, J. Biol. Chem. 272 (1997) 18044–18050.
[51] P. Gonzalez Porqué, A. Baldesten, P. Reichard, The involvement of the thioredoxin system in the reduction of methionine sulfoxide and sulfate, J. Biol. Chem. 245 (1970) 2371–2374.
[52] C.H. Lillig, C. Berndt, O. Vergnolle, M.E. Lonn, C. Hudemann, E. Bill, A. Holmgren, Characterization of human glutaredoxin 2 as iron-sulfur protein: a possible role as redox sensor, Proc. Natl. Acad. Sci. 102 (2005) 8168–8173.

[53] L.D. Schütte, S. Baumeister, B. Weis, C. Hudemann, E.-M. Hanschmann, C.H. Lillig, Identification of potential protein dithiol-disulfide substrates of mammalian Grx2, Biochim. Biophys. Acta 1830 (2013) 4999–5005.
[54] P. Haunhorst, C. Berndt, S. Eitner, J.R. Godoy, C.H. Lillig, Characterization of the human monothiol glutaredoxin 3 (PICOT) as iron-sulfur protein, Biochem. Biophys. Res. Commun. 394 (2010) 372–376.
[55] P. Haunhorst, E.-M. Hanschmann, L. Bräutigam, O. Stehling, B. Hoffmann, U. Mühlenhoff, R. Lill, C. Berndt, C.H. Lillig, Crucial function of vertebrate glutaredoxin 3 (PICOT) in iron homeostasis and hemoglobin maturation, Mol. Biol. Cell 24 (2013) 1895–1903.
[56] C. Johansson, A.K. Roos, S.J. Montano, R. Sengupta, P. Filippakopoulos, K. Guo, F. von Delft, A. Holmgren, U. Oppermann, K.L. Kavanagh, The crystal structure of human GLRX5: iron-sulfur cluster co-ordination, tetrameric assembly and monomer activity, Biochem. J. 433 (2011) 303–311.
[57] C. Camaschella, A. Campanella, L. De Falco, L. Boschetto, R. Merlini, L. Silvestri, S. Levi, A. Iolascon, The human counterpart of zebrafish shiraz shows sideroblastic-like microcytic anemia and iron overload, Blood 110 (2007) 1353–1358.
[58] M. Murakami, K. Nakashima, D. Kamei, S. Masuda, Y. Ishikawa, T. Ishii, Y. Ohmiya, K. Watanabe, I. Kudo, Cellular prostaglandin E2 production by membrane-bound prostaglandin E synthase-2 via both cyclooxygenases-1 and -2, J. Biol. Chem. 278 (2003) 37937–37947.
[59] C.-Y. Chiang, C.-C. Pan, H.-Y. Chang, M.-D. Lai, T.-S. Tzai, Y.-S. Tsai, P. Ling, H.-S. Liu, B.-F. Lee, H.-L. Cheng, et al., SH3BGRL3 protein as a potential prognostic biomarker for urothelial carcinoma: a novel binding partner of epidermal growth factor receptor, Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 21 (2015) 5601–5611.
[60] Weingarten, L. Identification of novel cytosolic thioredoxin-1 target proteins in mammalian cells by mechanism-based kinetic trapping. Available online: https://archiv.ub.uni-heidelberg.de/volltextserver/8923/(accessed on Sep 5, 2018).
[61] Claudia Urbainsky, Nölker Rolf, Marcel Imber, Lübken Adrian, Jörg Mostertz, Falko Hochgräfe, Jose Godoy, Eva-maria Hanschmann, and Christopher Horst Lillig nucleoredoxin-dependent targets and processes in neuronal cells, Oxid. Med. Cell. Longev. (2018).
[62] A. Jiménez, C. Johansson, J. Ljung, J. Sagemark, K.D. Berndt, B. Ren, G. Tibbelin, R. Ladenstein, T. Kieselbach, A. Holmgren, et al., Human spermatid-specific thioredoxin-1 (Sptrx-1) is a two-domain protein with oxidizing activity, FEBS Lett. 530 (2002) 79–84.
[63] A. Miranda-Vizuete, J. Ljung, A.E. Damdimopoulos, J.A. Gustafsson, R. Oko, M. Pelto-Huikko, G. Spyrou, Characterization of Sptrx, a novel member of the thioredoxin family specifically expressed in human spermatozoa, J. Biol. Chem. 276 (2001) 31567–31574.
[64] C.M. Sadek, A.E. Damdimopoulos, M. Pelto-Huikko, J.A. Gustafsson, G. Spyrou, A. Miranda-Vizuete, Sptrx-2, a fusion protein composed of one thioredoxin and three tandemly repeated NDP-kinase domains is expressed in human testis germ cells, Genes Cells Devoted Mol. Cell. Mech. 6 (2001) 1077–1090.
[65] C.M. Sadek, A. Jiménez, A.E. Damdimopoulos, T. Kieselbach, M. Nord, J.-A. Gustafsson, G. Spyrou, E.C. Davis, R. Oko, F.A. van der Hoorn, et al., Characterization of human thioredoxin-like 2. A novel microtubule-binding thioredoxin expressed predominantly in the cilia of lung airway epithelium and spermatid manchette and axoneme, J. Biol. Chem. 278 (2003) 13133–13142.
[66] Y. Lu, X. Zhao, G. Luo, G. Shen, K. Li, G. Ren, Y. Pan, X. Wang, D. Fan, Thioredoxin-like protein 2b facilitates colon cancer cell proliferation and inhibits apoptosis via NF-κB pathway, Cancer Lett. 363 (2015) 119–126.
[67] A. Jiménez, W. Zu, V.Y. Rawe, M. Pelto-Huikko, C.J. Flickinger, P. Sutovsky, J.-A. Gustafsson, R. Oko, A. Miranda-Vizuete, Spermatocyte/spermatid-specific thioredoxin-3, a novel Golgi apparatus-associated thioredoxin, is a specific marker of aberrant spermatogenesis, J. Biol. Chem. 279 (2004) 34971–34982.
[68] D. Wang, X. De Deken, M. Milenkovic, Y. Song, I. Pirson, J.E. Dumont, F. Miot, Identification of a novel partner of duox: EFP1, a thioredoxin-related protein, J. Biol. Chem. 280 (2005) 3096–3103.
[69] R. Shaheen, K. Szymanska, B. Basu, N. Patel, N. Ewida, E. Faqeih, A. Al Hashem, N. Derar, H. Alsharif, M.A. Aldahmesh, et al., Characterizing the morbid genome of ciliopathies, Genome Biol. 17 (2016) 242.
[70] C. Harz, N. Ludwig, S. Lang, T.V. Werner, V. Galata, C. Backes, K. Schmitt, R. Nickels, E. Krause, M. Jung, et al., Secretion and immunogenicity of the meningioma-associated antigen TXNDC16, J. Immunol. Baltim. Md 1950 193 (2014) 3146–3154.
[71] W. Jeong, Y. Jung, H. Kim, S.J. Park, S.G. Rhee, Thioredoxin-related protein 14, a new member of the thioredoxin family with disulfide reductase activity: implication in the redox regulation of TNF-alpha signaling, Free Radic. Biol. Med. 47 (2009) 1294–1303.
[72] I. Pader, R. Sengupta, M. Cebula, J. Xu, J.O. Lundberg, A. Holmgren, K. Johansson, E.S.J. Arnér, Thioredoxin-related protein of 14 kDa is an efficient L-cystine reductase and S-denitrosylase, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 6964–6969.
[73] C. Thorpe, K.L. Hoober, S. Raje, N.M. Glynn, J. Burnside, G.K. Turi, D.L. Coppock, Sulfhydryl oxidases: emerging catalysts of protein disulfide bond formation in eukaryotes, Arch. Biochem. Biophys. 405 (2002) 1–12.