# Chapter 10
# Structural Characterization of Mammalian Selenoproteins

**Stefano M. Marino, Vadim N. Gladyshev, and Alexander Dikiy**

**Abstract** Structural analysis of proteins is a highly informative approach to assess protein function and regulation. It can help establish catalytic mechanisms of enzymes and visualize the structural basis for their interactions with substrates and partner proteins. To date, the following mammalian selenoproteins have been structurally characterized either by X-ray crystallography or nuclear magnetic resonance spectroscopy: the 15 kDa selenoprotein, glutathione peroxidases 1, 2, 3 and 4, selenophosphate synthetase 2, selenoproteins M and W, methionine sulfoxide reductase B1, and thioredoxin reductases 1 and 3. For structural analysis of most of these proteins, the catalytic selenocysteine was mutated to cysteine or glycine, allowing high protein expression. These structures and dynamic properties of selenoproteins verified the dominance of thioredoxin fold in mammalian selenoproteins and yielded critical insights into their functions and catalytic mechanisms.

## 10.1 Structure-Function Characterization of Proteins

Numerous genome sequencing projects yielded tens of millions of protein sequences. Some of these proteins are well known and characterized; however, the majority correspond to unknown proteins (both structurally and functionally). One of the major challenges scientists face (and will continue to face in the future) is to assign functions to each protein and protein family identified through genome sequencing projects. Analysis of protein function is considerably more complicated

S.M. Marino • V.N. Gladyshev
Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

A. Dikiy (✉)
Department of Biotechnology, Norwegian University of Science and Technology, 7491 Trondheim, Norway
e-mail: alex.dikiy@biotech.ntnu.no

D.L. Hatfield et al. (eds.), *Selenium: Its Molecular Biology and Role in Human Health*,
DOI 10.1007/978-1-4614-1025-6_10, 

than gene and genome sequencing, as it requires integration of many different, yet synergistic, technologies.

Structural biology emerged as one of the core areas of the life sciences because three-dimensional structures of proteins provide many functional insights [1, 2]. Functional assignments through their structural comparison with proteins of known structure and function prove particularly valuable, since many functional similarities escape detection by sequence-based approaches [3]. The concept of determining protein structure on a genome-wide scale is called structural proteomics. Its methods include X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy on the experimental side, and molecular modeling (e.g., homology modeling, docking to identify substrates for proteins of unknown function) and functional inferences (e.g., functional associations, such as using the STRING database) on the computational side.

## 10.2 Structural Characterization of Selenoproteins

Mammalian genomes contain 24 or 25 selenoprotein genes. Their selenium (Se)-containing amino acid, selenocysteine (Sec), is co-translationally inserted into nascent proteins in response to the UGA codon, which is also the signal designating translation termination. UGA codes for Sec only if the mRNA contains a characteristic stem-loop structure, the SECIS element, in its 3′-untranslated region, and this process also requires several components of the Sec insertion machinery, including $tRNA^{[Ser]Sec}$, elongation factor, EFsec, and SECIS-binding protein 2. In bacteria, SECIS elements are located immediately downstream of UGA (i.e., within coding regions of selenoprotein genes). Thus, the majority of eukaryotic selenoproteins cannot be expressed in bacteria without changes in the coding sequences creating SECIS elements. Exceptions are selenoproteins with Sec in the C-terminal regions, such as thioredoxin reductase, which can be expressed in bacterial cells (its Sec is a C-terminal penultimate amino acid, allowing introduction of the SECIS element downstream of the stop codon). Undoubtedly, this limitation considerably affected structure analyses of selenoproteins: both X-ray crystallography and NMR spectroscopy require large amounts of proteins (at least 10–20 mg of a protein are typically needed for full structural analysis). The inability to use bacterial hosts for production of sufficient amounts of selenoproteins led to the use of their cysteine (Cys)-containing forms (or glycine-containing forms). The only exceptions include thioredoxin reductase 1 (TR1) and glutathione peroxidase 1 (GPx1), whose structures include the actual Sec.

To date, the following mammalian selenoproteins have been structurally characterized: the 15 kDa selenoprotein (Sep15), glutathione peroxidases 1 (GPx1), 2 (GPx2), 3 (GPx3), and 4 (GPx4), selenophosphate synthetase 2 (SPS2), selenoproteins M (SelM) and W (SelW), methionine sulfoxide reductase B1 (MsrB1), and thioredoxin reductases 1 (TR1, also called TrxR1 or TxnRd1) and 3 (TR3, also called TrxR2 or TxnRd2). A graphical overview of all mammalian selenoprotein structures is shown in Fig. 10.1.

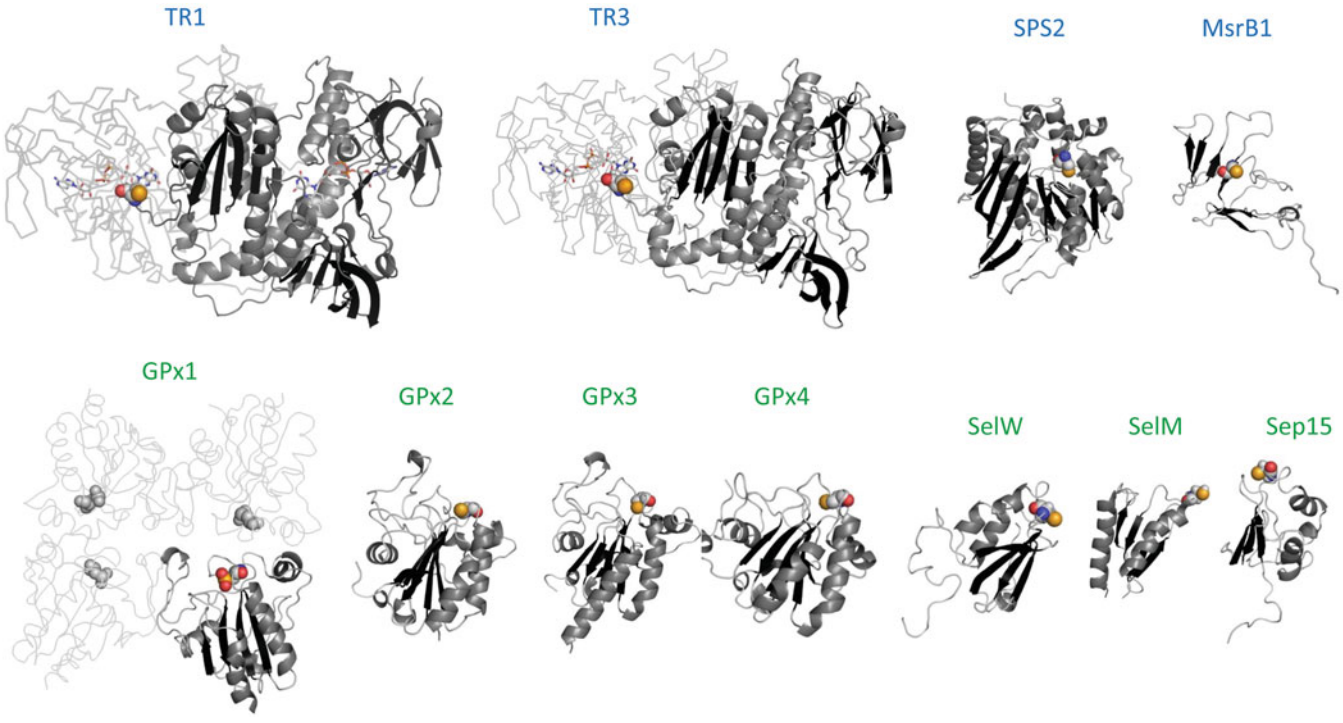


**Fig. 10.1** Three-dimensional structures of mammalian selenoproteins. Selenoprotein names are shown above each structure. Helixes are shown in *dark gray*, strands in *black* and coils in light *gray*. For each protein, the catalytic Sec/Cys is shown as space fills. Selenoproteins with the thioredoxin fold are labeled with names in *green*, and their structures are reported in the *bottom panel*. Mammalian TR1 and TR3 are homodimers. In the figure, one monomer is reported in ribbon representation and a color scheme, while the backbone atoms of the other monomer are shown in *gray*. The C-terminus of TR3 (pdb code 1zdl) was modeled based on the structure of human TR1 (3ean). The GPx1 quaternary structure is derived from the assembly (1gp1). GPx2 and GPx3 are also natural homo-tetramers (not shown in the figure). GPx1 has been crystallized with its Sec overoxidized

The most represented structural fold among selenoproteins is the thioredoxin fold (SelW, SelM, Sep15 and all four GPxs), followed by pyridine nucleotide disulfide oxidoreductases (all TRs). Below, we review each mammalian selenoprotein with a known 3D structure.

## *10.2.1 Thioredoxin Reductases*

TRs maintain their thioredoxin substrates in the reduced state and also reduce a variety of small compounds with NADPH as the electron donor. The thioredoxin system, along with the glutathione system, are the major systems of cellular redox homeostasis [4, 5]. All mammalian TRs are selenoproteins, so the entire mammalian thioredoxin system is dependent on Se.

### 10.2.1.1 Cytosolic TR1

Initially, an X-ray structure of rat cytosolic TR1 (truncated enzyme lacking Sec) was described [6]. More recently, a recombinant TR1 expressed in bacteria was structurally analyzed that had Sec inserted with the help of an engineered SECIS element [7]. TRs are homodimers with two subunits arranged in a "head-to-tail" manner. Their overall topology is similar to that of other pyridine nucleotide disulfide oxidoreductases, particularly glutathione reductases [6–8]. The structural features revealed by crystallographic analyses of both truncated (pdb code 1h6v) [6] and Sec-containing (pdb codes for reduced and oxidized forms are 3ean and 3eao, respectively) [7] forms are very similar: root mean square deviation (RMSD) value from backbone atom comparison is 0.4–0.5 Å. These studies showed, besides the obligatory "head-to-tail" arrangement, that the redox-active C-terminal tail of one subunit interacts with the N-terminal active site of the other subunit (Fig. 10.1). Each subunit of TR1 consists of three domains: FAD-binding (residues 1–163 and 297–367, rat TR1 numbering), NADPH-binding (residues 164–296), and the interface (residues 368–499) domains. The FAD- and NADPH-binding domains have similar structural organizations: each contains a central five-stranded parallel β-sheet and a three-stranded β-meander that is packed against the larger β-sheet. The other side of the parallel sheet is covered by several α-helices. These domains bind FAD and NADPH in a manner which is characteristic of this fold. The active disulfide is formed between Cys59 and Cys64: it is located in helix α2, within the FAD domain, a feature in common with other pyridine nucleotide disulfide oxidoreductases [8]. Interestingly, structural analyses confirmed that mammalian TR1 is more closely related to glutathione reductases than to prokaryotic TRs, in which the redox-active disulfide is located within the NADPH-binding domain. This observation is in excellent agreement with previous findings, derived from amino acid sequence analyses and evolutionary comparison of pyridine nucleotide disulfide oxidoreductases [9].

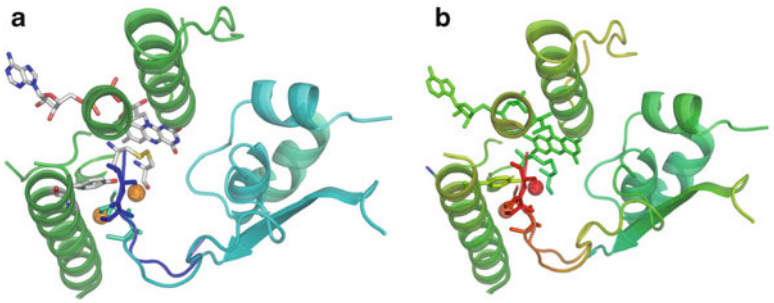


**Fig. 10.2** The active site of TRs and the mobility of the C-terminal region. (**a**) Simplified view (only amino acids within 20 Å from the catalytic Sec are shown) of the "head-to-tail" arrangement of rat TR1. The two monomers are shown in cartoon representation. The N-terminal part of one monomer is shown in *green ribbons* (the N-terminal redox center is shown in *sticks*). The C-terminal part of the second monomer is shown in *cyan colored ribbons*. The C-terminal redox center (GCUG) is shown in *sticks*, colored in *cyan* (C and U in *cis*-conformation, oxidized protein, pdb code 3eao) or in *blue* (C and U in *trans*-conformation, reduced protein, pdb code 3ean). (**b**) The same arrangement as in (**a**), but colored according to B-factors. The color scheme ranges from *red* (high mobility, reflected by high B-factor values), to *blue* (low mobility, low B-factor), passing through *yellow* and *green*

The third domain of mammalian TR1, the interface domain, contains an antiparallel five-stranded β-sheet flanked on both sides by four helices. This domain participates in subunit–subunit interaction and forms a large part of the dimer interface.

The C-terminal extension with the characteristic Gly-Cys-Sec-Gly motif carrying the essential Sec residue is a signature feature of mammalian TRs. The structural peculiarities of this C-terminal motif can be elucidated by the analysis of structures corresponding to the native protein in its reduced and oxidized forms (pdb codes 3ean and 3eao, respectively). As it appears from this analysis, the last six residues are highly flexible (Fig. 10.2) with an ensemble of possible conformations, particularly in the reduced form of the enzyme. It was suggested that the mobile C-terminal region not only serves as a third redox-active group, but it also blocks access of oxidized glutathione to the N-terminal redox Cys couple [10]. However, only weak electron density was observed for the C-terminal motif, and thus its positioning is not well determined, which is reflected by very high B-factor values of its atoms. Besides the intrinsic mobility of this region, an additional important feature is the *cis* to *trans* movement of Sec, in respect to the flanking Cys residue. In the reduced state, the functional group of Sec498 points away from the thiol of Cys497 (Fig. 10.2), whereas in the oxidized form, Sec flips on the side of Cys497 so that different redox states of TR1 (reduced or oxidized) are coupled with a significant relocation of Sec498 relative to Cys497. This movement has been proposed as a sort of shuttling from the N-terminal redox center to the exterior part of the protein. In this way, the enzyme can interact with its substrate, thioredoxin [11]. A closely located Tyr116 assists in this interaction. Based on the analysis of available X-ray structures of native rat TR1, a catalytic mechanism was proposed for this important oxidoreductase [7].

#### 10.2.1.2 Mitochondrial TR3

Crystallographic structures of the Sec to Cys variant of mouse mitochondrial TR3 in its reduced and oxidized forms (pdb codes are 1zdl and 1zkq, respectively) were determined by molecular replacement [12] using rat cytosolic TR1 as a probe [6]. This fact by itself indicates that the structures of TR1 and TR3 are rather similar. Indeed, the RMSD value for comparison of protein backbones is 0.79 Å, indicating that the main structural features are well maintained in both enzymes (Fig. 10.1), particularly in relation to the overall architecture of the active sites. The Gly-Cys-Sec-Gly motif in the TR3 structure was modeled into the active site of the enzyme [12]. As expected, this model showed that the C-terminal region is positioned in the active site of the opposing subunit, as is in the case of TR1 [6, 7]. According to this model, the Se atom of selenocysteine is nearly equidistant from the sulfur of Cys86 (numbering based on mouse TR3 sequence) and the epsilon nitrogen of His497* (asterisk denotes that this residue belongs to the other subunit of TR3). Similarly, the sulfur of Cys522* is adjacent to the epsilon nitrogen of His143. The model of the C-terminal tail of TR3 is consistent with Sec523* interacting with the active site dithiol/disulfide via Cys86. After reduction of the selenenylsulfide, the resulting selenolate attacks the disulfide of the oxidized thioredoxin. Overall, the catalytic mechanism of TR3 is very similar to that proposed for TR1 [7, 12].

### *10.2.2 SPS2*

Incorporation of Sec into proteins requires the generation of an intermediate reactive Se donor compound, selenophosphate [13, 14]. Selenophosphate synthetases (SPSs) are enzymes responsible for the synthesis of selenophosphate by activating selenide with ATP. These proteins are strictly conserved in organisms that utilize Se, i.e., organisms that use Sec or selenouridine. Animals possess two SPSs (SPS1 and SPS2); however, only the latter is capable of synthesizing selenophosphate [15, 16]. Interestingly, in most animals, SPS2 is itself a selenoprotein. It exhibits a di-kinase activity wherein selenide and ATP are converted to selenophosphate, orthophosphate, and AMP. Mutational analyses revealed that Cys17 (or Sec13, numbering of *Escherichia coli* and *Aquifex aeolicus* sequences, respectively) and Lys20 (Lys16) are essential for SPS activity. In the case of SPS2, structural analysis was only possible with the Cys mutant [17]. Nevertheless, these studies provided important insights into the structure and function of this protein. SPS2 exhibits a mixed α/β fold, which is typical of PurM superfamily members. It consists of two halves represented by N-terminal and C-terminal domains (residues 1–156 and 157–336, respectively) (pdb code 2yye) [17]. The N-domain is composed of a six-stranded mixed β-sheet flanked by three α-helices and two $3^{10}$-helices. The C-terminal domain has a seven-stranded mixed β-sheet, eight α-helices, and three $3^{10}$-helices. A steep 30 Å-long channel is formed between the two domains, which seems to serve as the binding/catalytic site for the substrates. The channel is covered by the N-terminal turn-rich segment (residues 1–45), which shows high intrinsic flexibility,

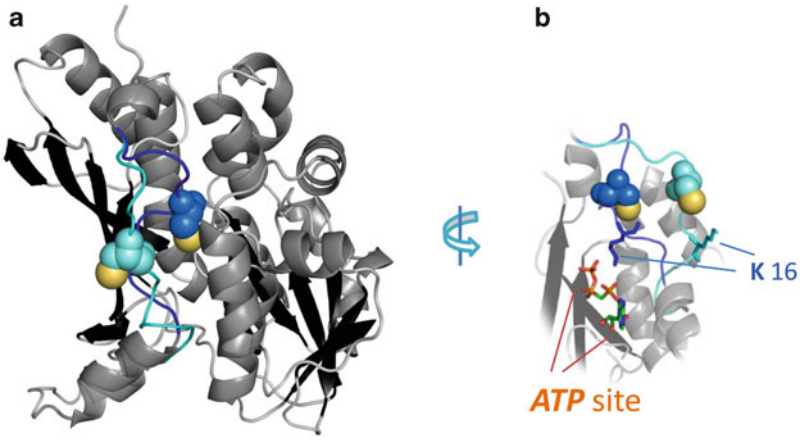


**Fig. 10.3** Mobility of the N-terminal region of SPS2. (**a**) The structure of SPS2 is shown in *cartoon*, with the catalytic Sec shown in space fills. (**b**) Detailed view of the active site (*blue* colored loop, with Sec in space fills) and of the ATP (AMP-PP, represented in *sticks*) binding site. The catalytic N-terminal loop can shuttle between two "extreme" positions: (1) inward, going towards the ATP binding site, and (2) outward, reaching the solvent and going further away from the ATP, as shown by the *cyan colored loop* in (**a**) and (**b**). A fundamental feature of the shuttling trajectory is that it moves Sec13 and Lys16 (numbering of *Aquifex aeolicus* SPS2, pdb code 2yye) in and out of the phosphate donor pocket. These residues are essential for catalysis: Lys16 stabilizes the interaction with ATP, and Sec is the catalytic nucleophile [15]

with large portions of this segment being disordered. The catalytically crucial residues Sec/Cys13 and Lys16 reside on this mobile segment. During the movement, these residues can approach ATP accommodated in the channel (Fig. 10.3). Interaction of the charged nitrogen of Lys16 provides a crucial binding for one of the negatively charged phosphate groups of ATP. According to the proposed reaction mechanism, the ability of Sec/Cys to move would be an essential feature of catalysis, as Sec could attack hydrogen selenide from the outside and then, after the Sec(Se)-Se bond is formed, bring in the newly bound Se, which subsequently nucleophilically attacks one of the phosphodiester bonds of ATP. As in the case of TRs, the structural analysis of SPS2 provided very valuable information on the reaction mechanism of this selenoprotein.

### *10.2.3 MsrB1*

Methionine sulfoxide reductases (Msr) are thiol-based oxidoreductases in which either Cys or Sec function as catalytic residues. These enzymes are classified into two families, MsrA and MsrB, according to their substrate specificity. MsrA catalyzes the reduction of the s-form of methionine sulfoxide (Met-s-SO), whereas MsrB can only reduce the R-form (Met-R-SO). These two enzyme families reduce methionine sulfoxides in proteins, but MsrA can also reduce free Met-s-SO. Msrs

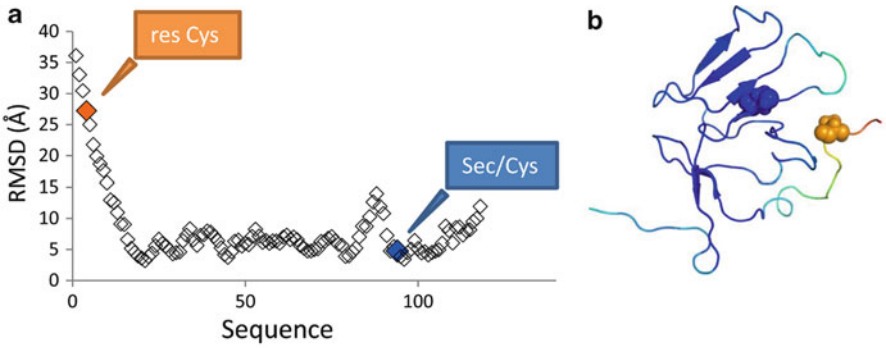


**Fig. 10.4** Mobility of the N-terminal region of MsrB1. (**a**) The average RMSD, calculated for carbon α atoms, in the pairwise comparison of all structures present in the NMR ensemble of mouse MsrB1 (18). The N-terminal region, including Cys4 (the resolving Cys, "res Cys" in the figure), is considerably more mobile than the rest of the protein. (**b**) Visual representation of the values calculated in (**a**), as plotted on the structure of mouse MsrB1 (shown in *cartoon*; color code: *red* higher mobility, *blue* lower mobility, *green* intermediate). The N-terminal resolving Cys is shown in space fill

are important repair proteins that protect cells against oxidative stress and have been implicated in delaying the aging process and protecting against neurodegeneration [18, 19]. One of four mammalian MsrBs, MsrB1 (also called SelR or SelX), is a selenoprotein.

Recently, a family of NMR solution structures of mouse selenoprotein MsrB1, in which Sec95 was replaced with Cys, was reported (pdb code 2kv1) [20]. The structure is characterized by an overall β-fold consisting of eight all antiparallel β-strands. The structure is of ellipsoidal shape and consists of two antiparallel β-sheets. The first β-sheet is three-stranded, forming the backside of the structure, while the second sheet has five strands forming the front side. The active site is situated in the second β-sheet. The N-terminal region contains two hinge sections defined by Gly8-Gly9 and Pro18-Gly19 pairs followed by the backside β-sheet consisting of strand $\beta_1$ (residues 19–23), $\beta_8$ (residues 100–104), and $\beta_2$ (residues 28–30) in the listed order. The C-terminal flexible region comes out of the middle of the backside β-sheet. The front side β-sheet is connected in the following order: $\beta_3$ (45–48), $\beta_7$ (93–96), $\beta_6$ (77–82), $\beta_5$ (66–72), and $\beta_4$ (55–60). This sheet forms the hydrophobic core through residues Leu67, Val69, Phe94, and Ile96, linking to the backside β-sheet hydrophobic Tyr21, Phe31, and Phe103 at the bottom of the structure. The top part of MsrB1 is held together through tetrahedral structural zinc, in which the metal ion is bound to the protein matrix by Cys23, Cys26, Cys71, and Cys74. A disordered region consisting of 13 residues (amino acids 31–44) between $\beta_2$ and $\beta_3$ strands connects their front side and backside β-sheets of the protein.

MsrB1 mobility data indicate that its N- and C-terminal regions are largely unstructured (since they are intrinsically flexible), with the N-terminus being particularly mobile (Fig. 10.4). It was suggested that the increased mobility of these

regions is needed during catalysis to form an intermediate selenenylsulfide [20]. It is also possible that flexibility of these regions promotes interactions between MsrB1 and its substrates.

On the basis of the MsrB1 structure, mobility studies and analysis of interactions of the enzyme with its substrate and inhibitor, the N-terminus was proposed to play a crucial role in MsrB1 structure and function: its unstructured nature and hydrophobic compatibility with the active site [20] allow it to move toward the reaction center in response to substrate binding. Therefore, MsrB1 can regulate the completion of its catalytic cycle by modulating access of the resolving Cys4 to the active site. Thus, even though the high mobility is not related to Sec (as in TR and SPS2, i.e., it relates to the N-terminal resolving Cys), high flexibility seems to be a common structural feature associated with Se-based enzymatic catalysis.

### *10.2.4 GPxs*

GPxs are critical components of the mammalian system that controls hydroperoxide levels [21]. These enzymes reduce peroxides at the expense of reduced glutathione and/or other reductants. Mammalian GPxs are classified into several subfamilies, which are numbered consecutively from 1 to 8 [22, 23]. GPx1, GPx2, GPx3, and GPx4 are invariantly Se-containing enzymes (i.e., they possess the active site Sec nucleophile). GPx6 is a selenoprotein in most mammals, but in many rodents it is a Cys-containing protein. GPx5, GPx7, and GPx8 are Cys-containing proteins in all mammals. X-ray structures were determined for native bovine GPx1 (pdb code 1gp1), Sec to Gly mutants of human GPx1 (2f8a), GPx3 (2r37) and GPx4 (2gs3), and Sec to Cys mutants of human GPx2 (2he3) and GPx4 (2obi). All GPxs have a rather similar structure; therefore, in the following paragraphs, we focus on just one of them, human GPx4, which is one of the best characterized GPxs.

The structural model characterizes GPx4 (Fig. 10.1) as a monomeric protein composed of 170 amino acids. In contrast, GPx1, GPx2, and GPx3 are tetramers (Fig. 10.1, note that the structure of only one the fully assembled tetramer is available, i.e., the structure of GPx1). Structures of all GPxs show a typical thioredoxin fold [24, 25] consisting of four α-helices on the protein surface and seven β-strands, five of which are clustered to form a central β-sheet (Fig. 10.1). The catalytic triad consisting of Sec46, Gln81, and Trp136 is positioned on the protein surface [25, 26].

Sequence alignments of GPxs indicated strict conservation of the catalytic triad: mutations of any of these residues impaired enzymatic activity [26, 27]. Additionally, a fourth residue (Asn137, numbering based on human GPx4) is also strictly conserved in GPxs and was proposed to assist in catalysis, forming the catalytic tetrad [22]. This amino acid is found in close proximity to the catalytic thiol/selenol of GPxs.

Structural analysis confirmed the importance of the tetrad residues in the modulation of GPx activity by establishing direct interactions (e.g., H-bonds) with the catalytic Cys (or Sec) [22]. Indeed, it was proposed that this H-bond network is responsible not only for structure stabilization, but also for lowering the p$K_a$ of

the catalytic residue [22]. However, this hypothesis does not appear to justify the observation that Sec-containing GPxs, in spite of the intrinsic acidity of Sec (i.e., Sec p$K_a$, estimated in the range 5–6, is considered to be ~3 pH units lower than that of Cys), show complete conservation of all residues belonging to the catalytic tetrad. Thus, it is still not completely clear why the tetrad residues are strictly conserved in GPxs, regardless of whether they do or do not contain the very reactive Sec (i.e., Sec does not need the H-bond network with tetrad residues to lower its p$K_a$).

Irrespective of the overall structural similarity, some important differences between various Sec-containing GPxs were found. In particular, comparison of their structures revealed differences between GPx4 and other GPxs. GPx4 lacks an internal stretch of about 20 amino acids, which is present in all other GPxs [24]. This sequence forms a surface exposed loop (loop 1) that does not exist in GPx4. Interestingly, this loop structure lines the active site of other GPxs and partially shields Trp136, a constituent of the catalytic tetrad. It may be concluded that the occurrence of this loop structure limits the accessibility of complex lipid substrates to the active site of GPxs. Instead, the lack of this structural element in GPx4 may allow efficient binding of these molecules in the active site.

### *10.2.5 SelW*

Rdx is a recently identified family of selenoproteins and Cys homologs [28]. Mammalian members of this family include selenoproteins SelW, SelT, SelH, SelV, and a Cys-containing Rdx12 protein. Functions are not known for any of these proteins. Through sequence analysis and structural modeling, it was predicted that these proteins possess a thioredoxin-like fold, suggesting a possible catalytic and redox-related function for their Sec (Cys) residues.

Among Rdx proteins, the best characterized is SelW. It is a 9 kDa selenoprotein, abundant in muscle and brain [29–32]. The structure of the Sec13Cys variant of mouse SelW was determined by high resolution NMR spectroscopy (pdb code 2npb, Fig. 10.1) [33]. It consists of a four-stranded β-sheet with two extended α helices and a short $3^{10}$ helix, all located on one side of the β-sheet. Besides these secondary structures, the protein contains two external loops and a type II turn. The molecule is characterized by a $\beta_1$(3–9)-$\alpha_1$(15–28)-$\beta_2$(34–40)-$\beta_3$(48–52)-$\beta_4$(54–60)-$3^{10}$(61–63)-$\alpha_2$(70–88) secondary structure pattern, wherein $\beta_1$ and $\beta_2$ are parallel strands forming a classical $\beta_1$-$\alpha_1$-$\beta_2$ motif, typical of thioredoxin fold proteins. The axis of the $\alpha_2$-helix is at an approximately 45° angle with respect to the axis of the $\alpha_1$-helix as a consequence of the β-sheet topology. The CxxU motif is located in the loop (residues 10–14) between $\beta_1$ and $\alpha_1$. The second loop (residues 40–47) separates $\beta_2$ and $\beta_3$ strands. $\beta_3$ and $\beta_4$ strands form a β-hairpin with a type II turn (residues 52–54) in between. The decreased structural resolution observed for the two loops may be a consequence of increased mobility of these regions with respect to the rest of the protein. Considering the already described role of high flexibility for other selenoproteins (TRs, SPS2, MsrB1), the latter observation suggests that, also in SelW, structural mobility has a functional role. Currently, however, the function of

SelW is unclear, with no experimental evidence supporting a model for its catalytic mechanism. Nevertheless, structural and computational analyses suggested a functional association with 14-3-3, wherein SelW might control the redox state of this protein [28, 33, 34].

### *10.2.6 Sep15 and SelM*

The structures of Cys versions of two thioredoxin-like selenoproteins, Sep15 and SelM (Fig. 10.1), were determined using NMR spectroscopy [35] (pdb codes 2a4h and 2a2p, respectively). These proteins have Sec-containing CxxC-like motifs, suggesting that they may function as redox proteins within the endoplasmic reticulum. Both Sep15 and SelM contain a central α/β domain composed of three α-helices (α1–3) and a mixed parallel/antiparallel four-stranded β-sheet (β1–4). Structure-based multiple sequence alignments illustrate that SelM has a short N-terminal extension that precedes strand β1 and a flexible C-terminal extension after helix α3. In contrast, Sep15 has an elongated Cys-rich N-terminal extension before strand β1 and a shorter C-terminal extension after helix α3 that does not adopt a regular secondary structure. The highly flexible C-terminal regions of SelM (residues 121–145) and Sep15 (residues 150–178) may participate in binding protein substrates or other redox proteins and may assume a defined conformation after such interactions.

Sep15 and SelM are structurally related to the Rdx family members in having an overall thioredoxin-like fold and lacking the thioredoxin helix following the $\beta_1$-$\alpha_1$-$\beta_2$ structural motif and two functionally relevant external loops. The second loop appears to be shorter in Sep15 and SelM, similarly to that in SelW. One difference between Rdx proteins on the one side and Sep15 and SelM on the other is that the $\alpha_1$-helix in the latter proteins is divided into two smaller α-helices with a kink in between.

Sep15, SelM, and SelW lack charged and hydrophobic residues within the two external loops, which appears to be a characteristic feature of these proteins. On the other hand, thioredoxin, Sep15 and SelM, in contrast to SelW, do not have the aromatic cluster. Since the fold of all these proteins is roughly the same, it is possible to exclude the role of aromatic residues in maintaining their topology. Although the structures of SelM and Sep15 are available, their functions remain unknown.

## 10.3 Conclusions

The structures reviewed in this chapter clearly show the many significant contributions that structural biology brought to the Se field. In some cases, advancements in selenoprotein function would not have been possible in the absence of structural data. By employing a variety of tools, both experimental (X-ray crystallography, NMR spectroscopy) and computational (modeling, docking), structural biologists provided important insights into selenoprotein functions, physico-chemical properties and catalytic mechanisms. As an example, only through structural approaches it

was possible to detect (and characterize) the crucial role of protein mobility in the modulation of catalytic properties, as well as in the control of substrate specificity within selenoproteins. Altogether, structural biology provided substantial contributions to the current understanding of this very important class of proteins. Structural characterization of the remaining mammalian selenoproteins is the major objective for structural biologists involved in the Se field and, more broadly, redox biology.

**Acknowledgements** This work was supported by National Institutes of Health grants (to VNG) and by the NT Faculty, NTNU (to AD).

## References

1. Zarembinski TI, Hung LW, Mueller-Dieckmann HJ et al (1998) Proc Natl Acad Sci USA 95:15189
2. Cort JR, Koonin EV, Bash PA et al (1999) Nucleic Acids Res 27:4018
3. Teichmann SA, Chothia C, Gerstein M (1999) Curr Opin Struct Biol 9:390
4. Fernandes AP, Holmgren A (2004) Antioxid Redox Signal 6:63
5. Holmgren A, Lu J (2010) Biochem Biophys Res Commun 396:120
6. Sandalova T, Zhong L, Lindqvist Y et al (2001) Proc Natl Acad Sci USA 98:9533
7. Cheng Q, Sandalova T, Lindqvist Y et al (2009) J Biol Chem 284:3998
8. Schulz GE, Schirmer RH, Sachsenheimer W et al (1978) Nature 273:120
9. Zhong L, Arnér ES, Ljung J et al (1998) J Biol Chem 273:8581
10. Gladyshev VN, Jeang KT, Stadtman TC (1996) Proc Natl Acad Sci USA 93:6146
11. Turanov AA, Kehr S, Marino SM et al (2010) Biochem J 430:285
12. Biterova EI, Turanov AA, Gladyshev VN et al (2005) Proc Natl Acad Sci USA 102:15018
13. Veres Z, Tsai L, Scholz TD et al (1992) Proc Natl Acad Sci USA 89:2975
14. Glass RS, Singh WP, Jung W et al (1993) Biochemistry 32:12555
15. Persson BC, Böck A, Jäckle H et al (1997) J Mol Biol 274:174
16. Xu XM, Carlson BA, Irons R et al (2007) Biochem J 404:115
17. Itoh Y, Sekine S, Matsumoto E et al (2009) J Mol Biol 385:1456
18. Weissbach H, Etienne F, Hoshi T et al (2002) Arch Biochem Biophys 397:172
19. Kim HY, Gladyshev VN (2007) Biochem J 407:321
20. Aachmann FL, Sal LS, Kim HY et al (2010) J Biol Chem 285:33315
21. Burk RF, Hill KE (1993) Annu Rev Nutr 13:65
22. Tosatto SC, Bosello V, Fogolari F et al (2008) Antioxid Redox Signal 10:1515
23. Brigelius-Flohé R (1999) Free Radic Biol Med 27:951
24. Scheerer P, Borchert A, Krauss N et al (2007) Biochemistry 46:9041
25. Martin JL (1995) Structure 3:245
26. Maiorino M, Aumann KD, Brigelius-Flohé R et al (1995) Biol Chem Hoppe Seyler 376:651
27. Maiorino M, Aumann KD, Brigelius-Flohé R et al (1998) Ernährungswiss 37:118
28. Dikiy A, Novoselov SV, Fomenko DE et al (2007) Biochemistry 46:6871
29. Beilstein MA, Vendeland SC, Barofsky E et al (1996) J Inorg Biochem 61:117
30. Gu QP, Sun Y, Ream LW et al (2000) Mol Cell Biochem 204:49
31. Sun Y, Gu QP, Whanger PD (2001) J Inorg Biochem 84:151
32. Loflin J, Lopez N, Whanger PD et al (2006) J Inorg Biochem 100:1679
33. Aachmann FL, Fomenko DE, Soragni A et al (2007) J Biol Chem 282:37036
34. Musiani F, Ciurli S, Dikiy A (2010) J Proteome Res 10:968
35. Ferguson AD, Labunskyy VM, Fomenko DE et al (2006) J Biol Chem 281:3536