Received: 1 June 2017 | Revised: 9 November 2017 | Accepted: 10 November 2017

DOI: 10.1002/prot.25424

RESEARCH ARTICLE

WILEY PROTEINS

Characterization of cysteine thiol modifications based on protein microenvironments and local secondary structures

Akshay Bhatnagar | Debashree Bandyopadhyay

Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Hyderabad, India

Correspondence
Debashree Bandyopadhyay, Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Hyderabad Campus, Hyderabad, 500078, India.
Email: banerjee.debi@hyderabad.bits-pilani.ac.in

Abstract

We have demonstrated earlier that protein microenvironments were conserved around disulfide-bridged cystine motifs with similar functions, irrespective of diversity in protein sequences. Here, cysteine thiol modifications were characterized based on protein microenvironments, secondary structures and specific protein functions. Protein microenvironment around an amino acid was defined as the summation of hydrophobic contributions from the surrounding protein fragments and the solvent molecules present within its first contact shell. Cysteine functions (modifications) were grouped into enzymatic and non-enzymatic classes. Modifications studied were—disulfide formation, thio-ether formation, metal-binding, nitrosylation, acylation, selenylation, glutathionylation, sulfenylation, and ribosylation. 1079 enzymatic proteins were reported from high-resolution crystal structures. Protein microenvironments around cysteine thiol, derived from above crystal structures, were clustered into 3 groups—buried-hydrophobic, intermediate and exposed-hydrophilic clusters. Characterization of cysteine functions were statistically meaningful for 4 modifications (disulfide formation, thioether formation, sulfenylation, and iron/zinc binding) those have sufficient amount of data in the current dataset. Results showed that protein microenvironment, secondary structure and protein functions were conserved for enzymatic cysteine functions, in contrast to the same function from non-enzymatic cysteines. Disulfide forming enzymatic cysteines were tightly packed within intermediate protein microenvironment cluster, have alpha-helical conformation and mostly belonged to CxxC motif of electron transport proteins. Disulfide forming non-enzymatic cysteines did not belong to conserved motif and have variable secondary structures. Similarly, enzymatic thioether forming cysteines have conserved microenvironment compared to non-enzymatic cystienes. Based on the compatibility between protein microenvironment and cysteine modifications, more efficient drug molecules could be designed against cysteine-related diseases.

KEYWORDS

cysteine redox, cysteine thiol pKa, glutathionylation, metalloenzymes, nitrosylation, protein microenvironment, reactive nitrogen species (RNS), reactive oxygen species (ROS), ribosylation, selenylation

1 | INTRODUCTION

Thiol group of cysteine participates in various catalytic reactions.[1–5] Thiol group is interchangeably used with sulfydryl group, in this report. In regular protein structures, cysteines are observed in 2 predominant oxidation states, disulfide (oxidation number −1) and thiol (oxidation number −2).[6–8] In addition to these 2 redox forms, reactive cysteine thiols (R-SH) may undergo oxidation leading to several chemical modifications, like, sulfenamides [R-SN-R′]), sufinic acids [R-SO_2H], sulfonic acids [R-SO_3H], sulfinamides [R-SON-R′], and sulfonamides [R-SO_2N-R′].[3,9] These oxidations are mainly triggered within the cell by reaction oxygen species (ROS) and reactive nitrogen species (RNS), like hydrogen peroxide, organic hydroperoxides, hypohalous acids (HOX), and peroxynitrite etc.[9] Some of these oxidations are reversible whereas

others are irreversible. Cellular reductants, like thioredoxin and glutathione, facilitate reversible oxidation process. Whereas the formation of sulfenic and sulfonic acids are irreversible modifications.[3] All these cysteine modifications can be either a part of enzymatic or non-enzymatic chemical reactions. Cysteine is the second most reactive amino acid, after histidine, in enzymatic reactions.[1,4] Non-enzymatic cysteines are involved in variety of reactions, including scavenging of free radicals, pro-oxidants and heavy metal ions thus protecting the cell from oxidative damage.[5]

pKa of cysteine thiol group (default pKa value is 9.0) is largely affected by its nature of reactivity and embedded protein environment.[9,10] For example, pKa of catalytic cysteines acting as nucleophiles in *disulfide bond oxidoreductase (DSbA)* protein is as low as 3.5.[9] It was reported earlier that the pKa shift of titrable amino acids (side chains) were dependent upon embedded protein microenvironments.[1,11,12] However, limited methods are available to demonstrate the dependence of cysteine pKa values on protein micro-environments, for example, Raman spectroscopy[13] and fluorescence spectroscopy.[14]

Similarly, cysteine functions are also sensitive toward their embedded microenvironments.[2,15,16] For example, (1) thiolate group participates in catalytic action of glutathione S-transferase P protein only upon glutathionylation of the side chain,[17] (2) S-nitrosylation of cysteines in caspase-3 and caspase-9 induce apoptosis in mitochondria[18] but not in cytoplasm,[19] (3) S-Sulfenylation of cysteines in KCNA5 potassium voltage channel protein accounts for hypoxic condition and it downregulates channel protein expression,[20] (4) S-sulfenylation of cysteines in cancer cell lines was exploited to treat breast cancer.[21] Sensitivity of cysteine thiol functions and corresponding pKa values toward protein microenvironments were demonstrated only for isolated examples. However, characterization of cysteine thiol functions based on their embedded microenvironments were not attempted before.

We have earlier developed protein microenvironment quantitative descriptor that is the summation of hydrophobic contribution from all the protein fragments and solvent molecules present within the first contact shell of a particular amino acid (or its functional group).[22] Recently, we have demonstrated that protein microenvironments around -S-S- linkage of cystines are conserved, those are part of the active sites in specific enzyme classes, irrespective of their very low sequence similarity.[23] However, the study was confined to enzymatic cystines only, no references were made with respect to the non-enzymatic cystines. No literature report is yet known that correlates different functions of cysteine with their embedded protein microenvironments.

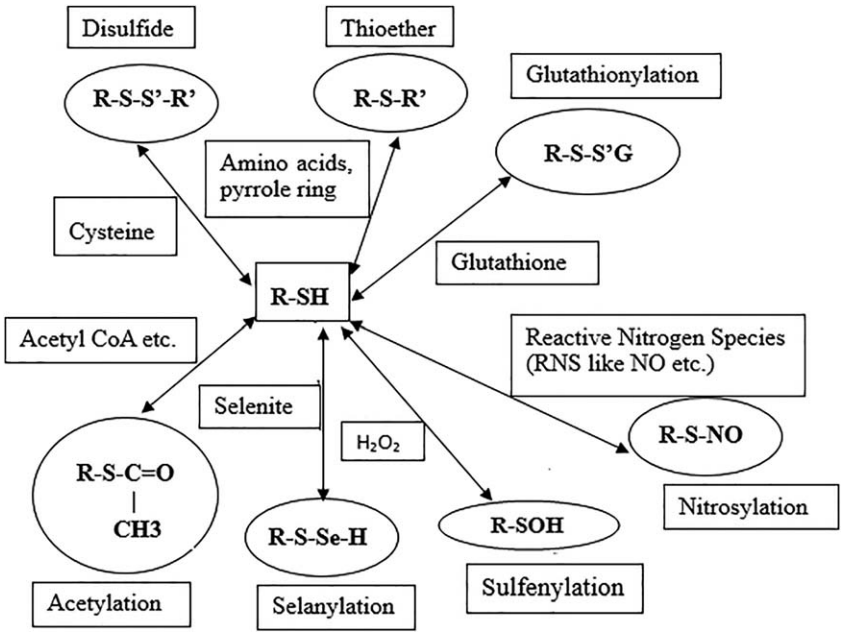
Here, we hypothesized that individual cysteine functions (metal-binding, redox and so forth) prefer particular protein microenvironment, secondary structure and specific biological function. This hypothesis has been tested on latest microenvironment dataset around cysteine thiol groups in high resolution enzyme crystal structures. We have demonstrated the cysteine thiol functions for both enzymatic and non-enzymatic classes. Similar to the functions of enzymatic cystines (disulfide-linked), individual enzymatic cysteine function has conserved microenvironment. However, non-enzymatic cysteines (of same thiol function) do not have conserved microenvironments. The possible outcome of this study will presumably help exploring more effective drugs, targeting cysteine-related disorders, based upon embedded protein microenvironments.

2 | METHODS

2.1 | Protein microenvironment calculation for cysteine residues in the current dataset

Protein microenvironment calculation around an amino acid required the 3-dimensional structure of the protein. This 3-dimensional structure can be derived from X-ray crystallography, NMR, or any other modelling techniques. In the present study, protein microenvironments around cysteine thiol groups were derived from crystal structures of enzymatic proteins in recent PDB[24] entries, dated August 2016. Selection criteria were following: (1) resolution better than 1.5 Å, (2) sequence similarity $<$30% (the least sequence similarity cutoff provided by PDB database), and (3) exclusion of modified proteins and protein-nucleic acid complexes. These selection criteria resulted into 896 enzymatic proteins with 5257 cysteine residues were selected (PDB IDs are shown in Table S1). These selection criteria were chosen to ensure structural precision and statistical adequacy of protein microenvironments around cysteine thiol groups. To ensure statistically significant number for cysteine functions, we have reviewed individual cases. For example, there were only 59 sulfenylated cysteines from 48 proteins, using resolution cutoff 1.5 Å. To make the number statistically significant for both enzymatic and non-enzymatic sulfenylated cysteines, resolution cutoff was removed; 390 sulfenylated cysteines were selected from 183 proteins. Including sulfenylated cysteines, total 1079 enzymatic proteins with 5647 cysteines were selected (Table S1).

Protein microenvironments around all the *5647* cysteines, were computed using a FORTRAN program developed earlier.[22] This calculation required following inputs—(1) 3-dimensional structure of the protein, (2) CHARMM[25] topology and parameter files, and (3) Rekker's fragmental constants of individual atom types.[26] Microenvironment calculations report 2 outputs—(1) buried fraction and (2) rHpy. Definition of these 2 output parameters follow. (1) Buried fraction was described by the normalized surface area of cysteine thiol group buried inside the protein. Values of this parameter ranged from 0.0 to 1.0. Zero buried fraction indicated that the thiol group was completely exposed to the solvent and vice versa. Buried fraction of an amino acid (or its side-chain) was computed in this FORTRAN program by calling another FORTRAN program GEPOL93.[27] (2) Second parameter, rHpy, termed as microenvironment property descriptor, described the relative hydrophobic contribution of protein and the solvent toward the cysteine thiol group within its first contact shell.[22] According to the mathematical formulation, rHpy value adopted upper limit of 1, when embedded in pure aqueous solvent. There was no lower limit for rHpy value that depended on the hydrophobicity of the protein interior. In our current dataset, lower limit of rHpy for cysteine thiol group was −0.311. Buried fraction and rHpy together constituted protein microenvironment space around cysteine thiol group.



SCHEME 1 Various oxidized forms of cysteine thiol group (R-SH), as reported in the current dataset

2.2 | Clustering of protein microenvironment space around cysteine thiol group

Protein microenvironment space around thiols was clustered using agglomerative hierarchical clustering.[28] Ward's method[29] was employed and XLSTAT software[30] was used. The microenvironment space was primarily divided into smaller bins of equal spacing [buried fraction = 0.1, rHpy = 0.1]. Agglomerative hierarchical clustering method joined smaller bins into single cluster, based on the distance proximity of the bin to the nearest cluster center. Distinct clusters were defined based on the closest proximity of a data point (buried fraction, rHpy) to a cluster center. Total 3 clusters were identified.

2.3 | Computation of cysteine side-chain pKa

Computation of pKa for all the 1079 proteins, in the current dataset, was performed using PROPKA 3.1 software.[31] An in-house perl script was used to extract the pKa values of cysteines only. PROPKA predicted the pKa values of ionizable groups in proteins and protein-ligand complexes based on the 3-dimensional structure.[31,32] The default pH value used in these calculations was 7.0. The default pKa value for cysteine side-chain was 9.0. Final pKa was calculated based on the surrounding interactions with the cysteine side-chain.

As PROPKA is quick and fast method for pKa computation, we have benchmarked our cysteine pKa values by comparing PROPKA results with Molecular Dynamics based pKa method, Amber MD pKa calculation.[33,34] pKa values for 10 cysteine thiol groups were compared from 9 proteins using 2 different methods, PROPKA and AMBER MD pKa calculation. Amber MD pKa results for these cysteines were obtained from literature.[35,36] The pKa values were fairly similar for these cysteines, from 2 different methods (Table S2). The RMSD of pKa values comparing these 2 methods was 1.33. This benchmarking indicated that although PROPKA is quick and fast method, the pKa results were comparable to more accurate method, hence, this method was applied to all the 5647 cysteine residues.

2.4 | Computation of secondary structures for cysteine residues

The secondary structures for all the 1079 proteins were calculated using DSSP software[37] based on Kabsch and Sander algorithm.[38] For each protein backbone, secondary structure information was obtained from DSSP. Secondary structures per cysteine in protein were extracted using perl script. The DSSP algorithm has calculated the secondary structure based on the 3-dimensional structure. The algorithm discarded any hydrogen atom present in the input structure. New hydrogen atoms were added to the backbone nitrogen atom (N-H distance 1.0 Å) pointing opposite to the backbone C = O bond. The new hydrogen positions were optimized. The secondary structure per residue in protein was determined by the relative position of hydrogen bonds, in the most stable conformation.[39]

2.5 | Selection of enzyme classes

According to the International Union of Biochemistry and Molecular Biology (IUBMB), 6 enzyme classes exist, namely, Oxidoreductases, Transferases, Hydrolases, Lyases, Ligases and Isomerases.[40] Names of the enzyme classes corresponding to the 1079 enzymatic proteins, in the current dataset, were extracted from respective PDB header files.

2.6 | Definition of cysteine thiol functions

Cysteine thiol functions studied were—disulfide formation, thioether formation, metal binding, acylation (acetylation, palmytoylation, and

thioester formation), ribosylation, nitrosylation, alkylation, glutathionylation, selanylation, and sulfenic acid formation (Scheme 1). These functions were broadly classified into 2 groups, enzymatic and non-enzymatic.

2.7 | Classification of cysteines into enzymatic and non-enzymatic groups

2.7.1 | Enzymatic cysteines

According to Catalytic Site Atlas (CSA)[41] database, cysteines were defined as **catalytic** those directly participate at the reaction catalytic center of an enzyme. Catalytic cysteines were identified from CSA dataset through perl script. Cysteines those did not participate in enzyme catalysis, directly, but involved in stabilization of the catalytic residues, were defined as **active**. Active cysteines were identified from PdbSum[42] database through perl script.

In this report, both catalytic and active cysteines together constitute **enzymatic** cysteines.

2.7.2 | Non-enzymatic cysteines

In this report, non-enzymatic cysteines were defined as those cysteines undergoing different chemical modifications (according to the current PDB dataset or the corresponding research articles) but not reported by PdbSum or CSA database.

2.8 | Identification of chemical modifications of cysteines

Cysteine chemical modifications (Scheme 1) were mainly identified from respective research articles. Corresponding to 1079 proteins reported in the current PDB dataset, only 829 research articles were available. All these 829 research articles were downloaded. Different chemical modification names were used as "keywords" to search a particular chemical modification in research articles (corresponding to PDB files). The cysteine modifications extracted from research articles and reported by CSA or PdbSum database were termed as enzymatic. Remaining other cysteine modifications extracted from research articles but not reported by the above databases were termed as non-enzymatic.

2.8.1 | Disulfide formation

Keyword search in respective research articles resulted in identification of cysteine "X" forming disulfide bond with cysteine "Y." For example, disulfide bond formation was reported between Cys288 and Cys317 in literature[43] corresponding to PDB ID: 4JXE. However, in the PDB file (4JXE) those 2 cysteines were reported in thiol form. This was because this particular PDB contained only the reduced form of thiol and the oxidized form was reported elsewhere (PDB ID: 4MS7).

2.8.2 | Thioether formation

Keyword search resulted in identification of ligands those were capable to form thioether with Cysteine "X." These results were cross checked with the PDB header files. 32 cysteines were reported to form thioether bond in different PDB header files.

2.8.3 | Metal binding cysteines

This kind of cysteines was identified using distance cutoff from the PDB coordinate file. A cysteine thiol group within 5 Å distance of a metal ion was considered as metal-binding. 5 Å distance cutoff ensured both (1) direct coordination of the thiol group to the metal ion and (2) water mediated interaction of thiol group to the metal ion. Metal ions analyzed in the current dataset are iron, copper, cobalt, mercury, zinc, cadmium, and nickel.

Remaining cysteine modifications, namely acylation (includes acetylation, thioester, and plamytoylation), nitrosylation, ribosylation, thioesterification, selenylation, and glutathionylation, were very few in number and were extracted from literature. However, those cysteine modifications were not always observed in respective PDB files. Most likely this was because those PDB files contained only the free form of thiol and cysteine modification was reported elsewhere. For example; in PDB ID: 3A2Z a glutathione molecule was observed within 2.2 Å of the cysteine thiol group. However, no glutathionylation was directly observed in that particular PDB file. Although, the corresponding research article has reported this thiol as possible glutathionylation site.

Out of the 5647 cysteines, the functionalities were identified only for 1085 cysteines from literature and databases. Total 589 enzymatic cysteines and 496 non-enzymatic cysteines were reported in the current dataset.

2.9 | Identification of PFAM families for all the proteins in the current dataset

Proteins belonging to the same PFAM family have similar functional domains.[44,45] Information of PFAM family corresponding to each PDB file was available in UNIPROT database.[46,47] An in-house perl script was used to extract PFAM family information from UNIPROT and reported for all 1079 proteins in the current dataset (Table S3).

3 | RESULTS AND DISCUSSION

3.1 | Protein microenvironment space around cysteine thiol group clustered into 3 different categories

Microenvironment space around cysteine thiol group was widely distributed across various proteins, in the current dataset. The entire protein microenvironment space around cysteine thiol group was clustered into 3 categories according to hierarchical clustering, namely, buried-hydrophobic, intermediate and exposed-hydrophilic (Figure 1). Each cluster was named according to their average buried fraction and rHpy values (Table 1). Population of cysteine thiol group gradually decreased from buried-hydrophobic microenvironment cluster to exposed-hydrophilic micro-environment cluster. Cystine disulfide, the oxidized form of thiol, was also distributed in 3 different microenvironment clusters.[23] However, the frequencies of populations differ among these 2 redox forms. Frequency of population was defined here as number of cysteines (or cystines) in respective cluster divided by total number of cysteines (or cystines). According to the protein structure based hydrophobicity scale developed earlier, disulfide containing

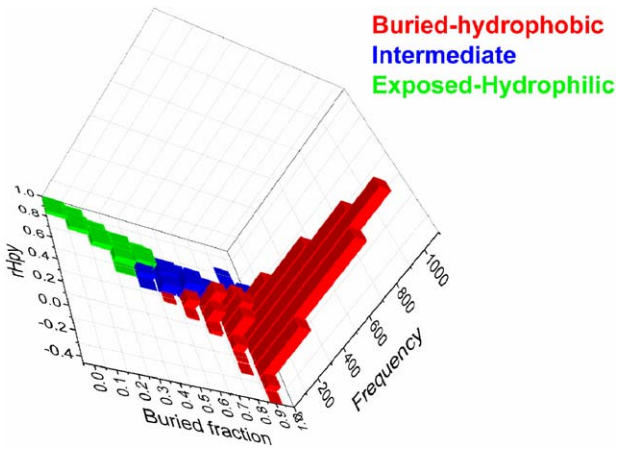


FIGURE 1 Distribution of all enzymatic cystines into 3 microenvironment clusters. *x* axis represents buried fraction and *y* axis represents rHpy. Z-axis shows the population of cysteine. Different colors represent 3 different microenvironment clusters; buried-hydrophobic (red), intermediate (blue), and exposed-hydrophilic (green) [Color figure can be viewed at wileyonlinelibrary.com]

cystine exhibited maximum hydrophobicity,[22] hence minimum frequency of population was observed in exposed hydrophilic cluster (0.015) [based on the data from table 1 of reference[23]]. On the other hand, the frequency of population for cysteine thiol group in exposed hydrophilic cluster was 0.03, twice compared to the value in disulfide group. This comparison indicated that 2 redox forms of cysteines were not equally populated in exposed hydrophilic protein microenvironments.

3.2 | Chemical modifications in enzymatic cysteines

According to the current dataset, 9 chemical modifications were identified in enzymatic cysteine group (Table 2). Out of those 9 types, 4 were more abundant and remaining others were fewer in number. To note, cysteine sulfenylation was selected using different criteria (see method section). Relative abundance of enzymatic cysteine chemical modifications was shown for different protein microenvironment clusters (Figure 2).

3.3 | Disulfide modification in enzymatic cysteines

Disulfide formation was one of the most abundant chemical modification in the current dataset. Thiol/disulfide-exchange reactions were nucleophilic substitutions of a thiol or thiolate nucleophile (RSH or RS^-) on a disulfide bond (RS-SR), which leads to oxidation of the nucleophile and reduction of the "exchanged" leaving group.[3]

Enzymatic cysteines capable of disulfide formation were mostly observed in intermediate and exposed hydrophilic microenvironment clusters (Table 2).

In intermediate microenvironment cluster, all the 38 cysteines belonged to cytochrome C (electron transport) protein family. All these 38 cysteines have either helix or turn secondary structures (Table S4). Majority of those cysteines (26 out of 38) were part of CxxC motifs

TABLE 1 Statistics of cysteine microenvironment clusters[a]

Cluster type	Average Buried fraction	AveragerHpy	Average distance to centroid (Å)	Within class variance	No. of cysteines in each cluster
Buried-hydrophobic	0.975	0.081	0.119	0.019	4062
Intermediate	0.773	0.401	0.150	0.028	1375
Exposed-hydrophilic	0.320	0.729	0.145	0.042	210

[a]Total 5647 cysteines were present in the dataset from 1079 proteins. Relative hydrophobic contribution of protein (weighted by buried fraction) and solvent (weighted by [1-buried fraction]) toward the microenvironment of individual cysteine was described by the parameter rHpy. Average rHpy value for each cluster was reported in third column. Distance between -SH group of individual cysteine and the centroid of each cluster was averaged and reported in fourth column.

TABLE 2 Enzymatic cysteines harboring different chemical modifications

Cysteine modifications (198)	Buried hydrophobic (89)	Intermediate (84)	Exposed-hydrophilic (25)
Disulfide (54)	3	38	13
Thio-ether (51)	18	21	12
Metal-bound (87)	66	21	0
Acylation (1)	0	1	0
Acetylation (2)	0	2	0
Alkylation (1)	0	1	0
Selenylation (1)	1	0	0
Ribosylation (1)	1	0	0
Sulfenylation (306)	63	230	13

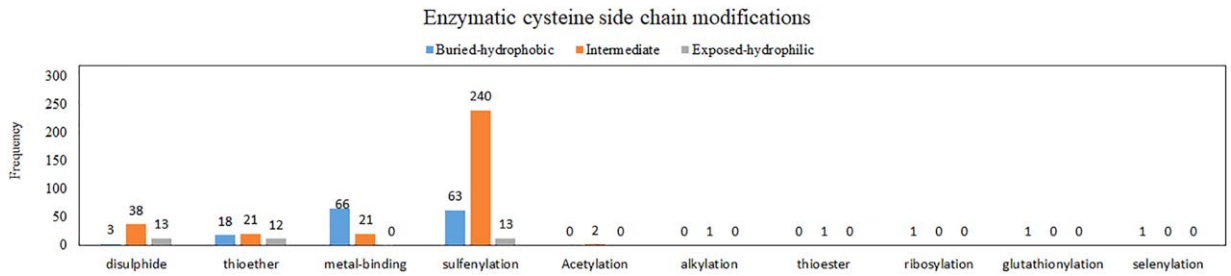


FIGURE 2 Frequencies of enzymatic cysteine side-chain modifications in different protein microenvironment clusters [Color figure can be viewed at wileyonlinelibrary.com]

from multiple chains of cytochrome C proteins (except 2CZS). Structural alignment of the CxxC motif of 26 cysteines resulted into Cα RMSD value of 0.88 Å. Low RMSD value indicated strict conservation of the motif. Both the cysteines from CxxC motif have exhibited helical secondary structure (Figure 3A). This observation was in contrast to the half-cystines in CxxC motifs of disulfides, where 1 cysteine belonged to helix and the other 1 belongs to turn.[23] Total 30 out of 38 cysteines in intermediate cluster belongs to helical conformation (Figure 3B). These 30 cysteines with helical conformation were packed within a smaller microenvironment space, (standard deviation of rHpy values is 0.09 compared 0.1 for the intermediate cluster). The computed pKa values for above-mentioned 30 cysteines (9.94 ± 1.29) were close to the default pKa value of cysteine side-chain (9.0).

In exposed-hydrophilic microenvironment cluster, all the 12 cysteines belonged to cytochrome C (Electron Transport) protein family (PDB ID: 1OFW; 2CZS), similar to intermediate microenvironment cluster. Microenvironments of these 12 cysteines were compact (in terms of low standard deviation value of rHpy, 0.07) compared to the overall spread of exposed hydrophilic cluster (standard deviation value of rHpy is 0.1). Those cysteines mainly adopt helical secondary structure (Figure 3B). Cysteines with helical conformation have lower pKa (10.76 ± 0.89) than those present in coil/bend/turn conformations (14.06 ± 1.75). Higher pKa values of cysteines, in coil/bend/turn conformations, can be presumably attributed to slight increase in hydrophobicity of the embedded microenvironments (average rHpy value for cysteines in turn confirmation is 0.613, slightly lower than the average rHpy value of potential disulfide forming cysteines in exposed-hydrophilic cluster, 0.67).

In buried-hydrophobic cluster, only 3 cysteines capable of disulfide formation were identified. Two of these cysteines, Cys288 (pKa, 8.49) and Cys317 (pKa, not computed by PROPKA), form disulfide bond among themselves in *AMSH like protease* (PDB ID: 4JXE). This protein (PDB ID: 4JXE) controls the destruction of cell-surface receptors on endosomal sorting complexes.[43] The sulfur-sulfur distance between these 2 cysteines was 4.4 Å in this crystal structure. Although this distance was too large to form disulfide bond, rotameric change in Cys288 side-chain enabled 2 cysteine thiols to come close and oxidize into disulfide (PDB ID: 4MS7).[43] Cys288, although deeply buried within the protein core (buried fraction value of 0.90), has moderate hydrophilic micro-environment (rHpy, 0.23). Upon oxidation, there was slight decrease in hydrophobicity (and subsequent increase in hydrophilicity, as indicated by higher rHpy value) of the protein micro-environment around Cys 288 and slight increase in hydrophobicity value around Cys317 (Table 3). Changes in protein micro-environment around Cys288 was evident from the superimposed structures of oxidized and reduced forms of the protein (Figure 4A).

Third example was Cys13 from protein *methionine sulfoxide reductase A (MsrA) of Mycobacterium tuberculosis* (PDB ID: 1NWA). This Cys13 residue was capable to form disulfide bond with Cys154 present in exposed-hydrophilic micro-environment cluster. Sulfur atoms from these 2 cysteines were 6.8 Å apart in this crystal structure. However, rotameric change of Cys13 side-chain allowed 2 thiol groups to oxidize

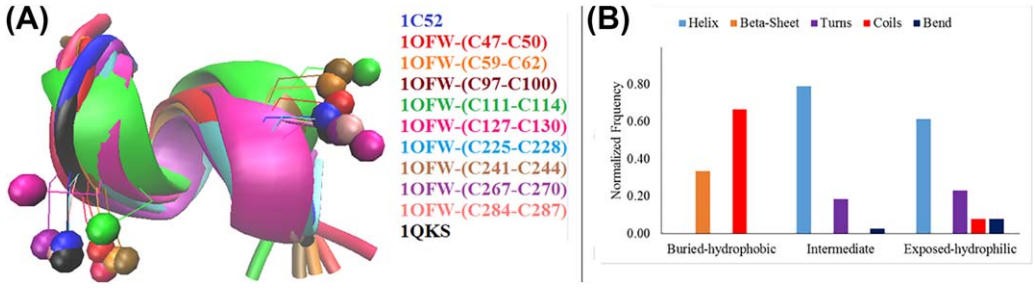


FIGURE 3 Disulfide modification in enzymatic cysteines: **A**, Helical structure of CxxC motifs, superimposed from 11 chains in 3 proteins (1 chain of PDB ID: 1C52, 9 of 1OFW, and 1 chains of 1QKS). Different chains are depicted by different colors. Helical structure of the motif is shown by cartoon representations. Cysteine sulfur atoms of the CxxC motif are depicted by ball representation. **B**, Normalized frequencies of disulfide forming cysteines in different microenvironment clusters and their related secondary structures. Normalization is done with the total number of enzymatic cysteines undergoing disulfide formation [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Changes in protein micro-environments around disulfide forming enzymatic cysteines before and after oxidation[a]

Residue No.	PDB ID; reduced form	BF	rHpy	MENV cluster	PDB ID; oxidized form	BF	rHpy	MENV cluster
288	4JXE	0.94	0.23	Buried-hydrophobic	4MS7	0.90	0.30	Buried-hydrophobic
317	4JXE	0.95	0.33	Buried-hydrophobic	4MS7	0.99	0.24	Buried-hydrophobic
13	1NWA	0.89	0.16	Buried-hydrophobic	2IEM	0.97	0.33	Buried-hydrophobic
154	1NWA	0.39	0.67	Exposed-hydrophilic	2IEM	0.79	0.25	Buried-hydrophobic

[a]Different properties, Buried fractions (BF), rHpy, and micro-environment cluster names, are shown for both reduced and oxidized forms.

into disulfide (PDB ID: 2IEM). Oxidation of these 2 thiols induced significant changes in the protein micro-environments of both of the cysteines (Table 3). Changes in the protein micro-environment, upon oxidation, around Cys 154 were more drastic as compared to Cys13 (Figure 4B). Oxidized form of Cys 154 was embedded in buried-hydrophobic protein micro-environment in contrast to the reduced form, embedded in exposed-hydrophilic micro-environment. The redox reaction involving Cys13 and Cys 154 in *Mycobacterium tuberculosis* was prerequisite for catalytic activity of MsrA protein.[48]

Above 3 examples suggested that potential disulfide forming cysteines were crucial for reductase and protease activities when embedded in buried-hydrophobic microenvironment cluster. Subsequent oxidation of these cysteines induced considerable amount of changes in respect protein microenvironments. Protein micro-environments around these 3 cysteines, in reduced form, were very similar to each other (standard deviation in rHpy value for these 3 was 0.08, in contrast to 0.13, the standard deviation of rHpy value in buried hydrophobic cluster).

Average pKa values for cysteines, present in buried-hydrophobic and intermediate clusters were lower compared to those present in exposed-hydrophilic cluster (Table 4A). According to Henderson-Hasselbalch equation, lower pKa values have indicated higher ratio of thiolate to thiol concentration. In this case, the thiolate concentration of cysteines present in buried-hydrophobic and intermediate clusters were 10-fold (pKa difference >1.0) higher than those present in exposed-hydrophilic cluster. Oxidation of thiolate to disulfide is easier compared to oxidation of thiol to disulfide.[49] Hence potential disulfide forming enzymatic cysteines were more easily oxidized in buried (hydrophobic or intermediate) protein microenvironments, compared to exposed-hydrophilic protein microenvironments.

Above results have explored the oxidation process of cysteine thiol to cystine disulfide (at enzyme reaction centers) in the light of protein microenvironments.

3.4 | Thio-ether modification in enzymatic cysteines

Thioether modification on proteins, also known as lanthionine, were formed by the elimination of 1 sulfur atom from disulfide bond (-C-S-S-C-) and produced carbon-sulfur-carbon (-C-S-C-) linkage.[50] Base-catalyzed thioether formation was observed in protein structures. Many of these thio-ether forming cysteines were part of CxxC motifs, as found in the current dataset—(1) $C_{18}xxC_{21}$ of PDB: 1W2L; (2) $C_{45}xxC_{48}$ of PDB: 3CP5; (3) $C_{54}xxC_{57}$ of PDB: 3M97; (4 $C_{37}xxC_{40}$ of PDB: 1E29; (5) $C_{35}xxC_{38}$ of PDB: 1M1Q; (6) $C_{15}xxC_{18}$ of PDB: 1M1Q, and (7) $C_{152}xxC_{155}$ of PDB: 3A9F. In all the above proteins, CxxC motifs were strictly conserved (Cα RMSD 0.34), despite of their overall low sequence similarity (26.7; multiple sequence alignment by Kalign 2.0[51]). Cysteines present at similar

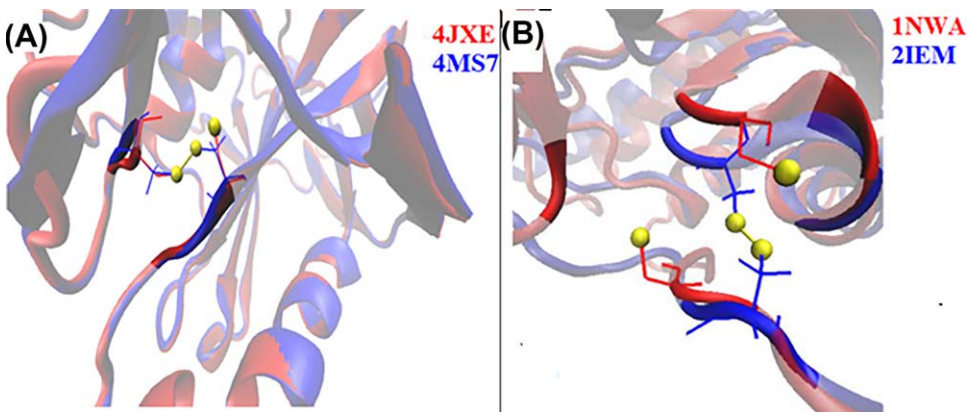


FIGURE 4 Changes in enzymatic cysteine microenvironment from thiol (reduced form) to disulfide (oxidized form) in **A** sst2 catalytic domain of AMSH-like protease; Cys288 and Cys317 residues in reduced (-SH) form (PDB ID: 4JXE) and oxidized (-S-S-) form (PDB ID: 4MS7) are shown **B** in protein methionine sulfoxide reductase A; Cys13 and Cys154 residues shown in reduced (-SH) form (PDB ID: 1NWA) and in oxidized (-S-S-) form (PDB ID: 2IEM) are shown. In figures, **A** and **B**, reduced protein is shown in red and oxidized protein in blue. Sulfur atoms are shown in yellow; reduced thiol form of sulfur is represented by a free yellow ball and oxidized disulfide form is represented by 2 yellow balls connected by a yellow stick

TABLE 4 Average (μ) and standard deviation values (σ) (shown in parenthesis) of protein microenvironment (rHpy) and pKa for enzymatic cysteines, embedded in different micro-environment clusters for following chemical modifications - a) disulphide formation, b) thio-ether formation c) metal-binding and d) S-Sulfenylation. Numbers of cysteines present in different micro-environment clusters are shown beside the cluster names

(A)			
	Buried-hydrophobic (3)	Intermediate (38)	Exposed-hydrophilic (24)
rHpy (μ,σ)	0.24 (0.08)	0.44 (0.10)	0.67 (0.07)
pKa (μ,σ)	10.14 (1.7)	10.04 (1.2)	11.69 (1.8)
(B)			
	Buried-hydrophobic (18)	Intermediate (21)	Exposed-hydrophilic (12)
rHpy (μ,σ)	0.12 (0.09)	0.41 (0.09)	0.72 (0.1)
pKa (μ,σ)	10.73 (1.63)	10.63 (1.71)	10.41(0.68)
(C)			
	Buried-hydrophobic (66)	Intermediate (21)	Exposed-hydrophilic (0)
rHpy (μ,σ)	0.11 (0.11)	0.37 (0.09)	0
pKa (μ,σ)	8.1 (2.20)	9.10 (3.0)	0
(D)			
	Buried-hydrophobic (63)	Intermediate (230)	Exposed-hydrophilic (13)
rHpy (μ,σ)	0.14 (0.08)	0.38 (0.11)	0.71 (0.10)
pKa (μ,σ)	11.4 (1.45)	10.6 (1.48)	9.37 (0.46)

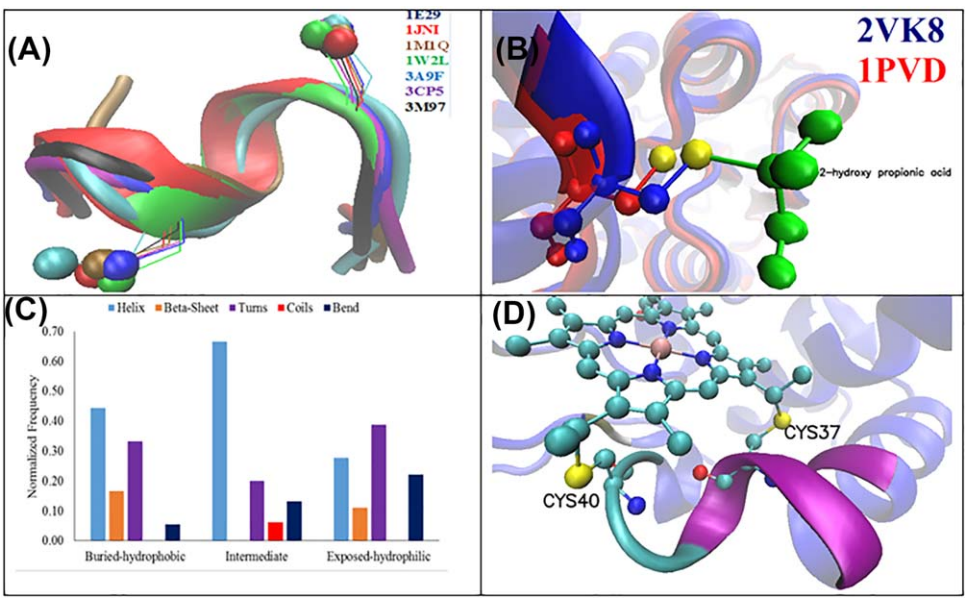


FIGURE 5 Thioether modification in enzymatic cysteines: **A**, helix turn conformation of CxxC motifs, superimposed from 7 proteins, **B**, protein conformational changes due to thiol to thio-ether (C-S-C) oxidation in *pyruvate decarboxylase* of Cys221. The protein in oxidized form (-C221-S-C-) is ligated to 2-hydroxy propionic acid (PDB ID: 2VK8); Oxidized form represented by blue cartoon representation and reduced form (PDB ID: 1PVD) by red cartoon. The sulfur atoms of Cys221, both in reduced and oxidized forms, are represented by yellow balls and remaining part of the cysteine in ball and stick representation. The 4.5 Å region around cysteine is highlighted. The Cα RMSD of the 2 proteins is 0.26 Å, **C**, Normalized frequencies of thioether forming cysteines in different microenvironment clusters and their related secondary structures, **D**, cysteine microenvironments around Cys 37 and Cys 40 (both part of CxxC motif) forming 2 separate thioether bonds with heme group in cytochrome protein (PDB ID:1E29). Cys 37 in helix conformation and Cys 40 in turn conformation [Color figure can be viewed at wileyonlinelibrary.com]

functional domains belong to same PFAM family, for example, cysteines from cytochrome C oxidase (PDB ID: 1W2L), cytochrome C-552 (PDB ID:3M97), cytochrome C (PDB ID:3CP5), all belong to PFAM family, PF00034 (Table S5). Cysteines in those motifs either belonged to helix or turn conformations (Figure 5A). According to the crystallographic data, 32 cysteines in the current dataset were part of thio-ether (-C-S-C-) linkage and 19 cysteines were capable to form thio-ether (according to literature) (Table S6). Two ligands were reported, in the current dataset, to form thio-ether with cysteines, (1) 2-hydroxy propionic acid and (2) heme group.

Functions of few thio-ether (-C-S-C-) linkages were illustrated below. Example 1: enzymatic Cys221 formed covalent bond (bond length of 1.82 Å) with 2-hydroxy propionic acid and contributed to (non-oxidative) decarboxylation of pyruvate, as a part of the *regulatory Site of Yeast Pyruvate Decarboxylases* (PDB ID: 2VK8).[52] Superimposition of oxidized (-C-S-C- in PDB ID: 2VK8) and reduced (-SH) (PDB ID: 1PVD) forms of Cys221 in pyruvate decarboxylase protein, indicated that the oxidized form was buried in slightly more hydrophobic microenvironment (rHpy, 0.16) compared to that of reduced form (rHpy, 0.227) (Figure 5B). Example 2: Cys18 forms a thio-ether bond with heme group in oxidized structure of tetraheme Cytochrome C protein (PDB ID: 1M1Q).[53] Example 3: Cys 45 and 48 form thio-ether bond with heme group in cytochrome C of *Rhodothermus marinus* (PDB ID: 3CP5).[54]

All the above 51 cysteines were more or less equally distributed through different protein micro-environment clusters (relatively less in exposed-hydrophilic microenvironment) (Table 2). Cysteines were more tightly packed within buried-hydrophobic cluster (standard deviation of rHpy values for 18 such cysteines was, 0.06), followed by those in intermediate cluster (standard deviation of rHpy for 20 such cysteines is 0.08). Compactness of thio-ether forming cysteine microenvironments was least in exposed-hydrophilic cluster (standard deviation of rHpy value for 12 such cysteine is 0.15, higher than that of exposed-hydrophilic cluster). Variation in secondary structures was fairly uniform across the microenvironment clusters (Figure 5C).

Overall pKa values for thio-ether forming cysteine side chains in different protein microenvironment clusters were more or less the same (Table 4B). pKa values for these 51 cysteines were computed by PROPKA.[31] PROPKA has a special provision to compute pKa for cysteines, part of CxxC motifs in thioredoxin family, even if those were in oxidized forms.

Although the overall pKa values were more or less the same, variations were observed when secondary structures of cysteines were taken in conjugation with different protein microenvironments (Table S6). Example 1: cysteines in turn conformations conjugated with buried-hydrophobic protein microenvironments have lower pKa values (9.2 ± 0.2) except 1 (Cys37 of PDB 1E29; pKa value 14.8). This Cys37 was partly buried (buried fraction 0.845) and moderately hydrophobic (rHpy 0.172), embedded within the heme group (Figure 5D). Example 2: Cysteines in intermediate cluster conjugated with bend conformation exhibit lower pKa (9.4 ± 0.3) value compared to the average pKa of the cluster (10.6 ± 1.7). Example 3: cysteines in buried hydrophobic cluster conjugated with turn conformation have higher pKa values (11.1 ± 1.6) compared to the average pKa value of the cluster (9.2 ± 0.2). Above 3 examples indicate that specific combinations of protein microenvironments and secondary structures influence the pKa value of the cysteines those form thio-ether within in protein structures.

3.5 | Metal binding modification in enzymatic cysteines

Total 87 cysteines were identified in the current dataset (Table 2) those coordinate to different metal ions either via direct coordination or via water mediated interactions and participate in catalytic reactions. Both of these metal binding modes were biologically important, as demonstrated for nucleic acid-Magnesium interactions.[55,56] Cysteines present at similar functional domains belong to same PFAM family, for example (1), cysteines from Ferridoxin (PDB ID: 1KRH) and carbon monoxide deaminase (PDB ID:1N62) belong to PFAM family PF00111; example (2), cysteines from cytosine deaminase (PDB ID: 1P60), cytidine deaminase (PDB ID: 2FR5), and blastidine deaminase (PDB ID:2Z3H), all belong to PFAM family, PF0383 (Table S5).

Sixty-six metal-binding cysteines were embedded in buried-hydrophobic cluster. According to the sulfur-metal distances (2.29 ± 0.3), cysteines embedded in buried-hydrophobic microenvironment directly coordinate to the metal ions (Table S7). pKa values for these cysteines were computed using PROPKA. pKa of 30 such cysteines were not reported (NA in Table S7) by PROPKA. According to the formulation, PROPKA ignored pKa computations for amino acids those were bound to ligands or ions and reported odd pKa value (99.99) for disulfide links.[31] Overall computed pKa values for enzymatic metal-binding cysteines (excluding above-mentioned 30 cysteines) were lower (8.4 ± 2.9) than the cysteines involved in other chemical modifications (Table 4C). Metal coordinated cysteines were predominant in thiolate form, as evident from their sulfur-metal distances and low or anomalous pKa values (reported as NA in Table S7). Lowering of cysteine side-chain pKa values in presence of positively charged Lysine and Arginine side chains were also reported earlier.[57] In our current dataset, metal-ligated cysteines were also located within the vicinity of positively (*albeit* $pH \leq 7$) charged residue, like Histidine (Figure 6A-C). Majority of these metal-binding cysteines adopted coil conformations (Figure 6D). Metal binding enzymatic cysteines present in oxidoreductase enzyme class, in general, exhibited lower range of pKa values (8.1 ± 2.2). In buried hydrophobic cluster, lowest pKa value for cysteine side-chain was exhibited for Cys214 of PDB: 1MXR (Table S7). Exceptions were also observed—cysteines from oxidoreductase family embedded in very hydrophobic (rHpy values negative) microenvironments exhibit higher pKa values; for example, Cys45 of PDB: 1IOR, Cys291 of PDB ID:3V4K, Cys68 of PDB ID:2W3Q (Table S7).

Twenty-one metal-binding cysteines embedded in intermediate micro-environment cluster, have overall larger pKa values (9.1 ± 3.0) compared to the average pKa value in buried-hydrophobic cluster. Some of these cysteine side chains exhibited pKa values as low as 4.89 in Cys288 of PDB: 3V4K (Table S7). These cysteines mostly belong to hydrolase enzyme class (10 out of 21) in contrast to buried hydrophobic cluster (those mainly belong to oxidoreductase enzyme class).

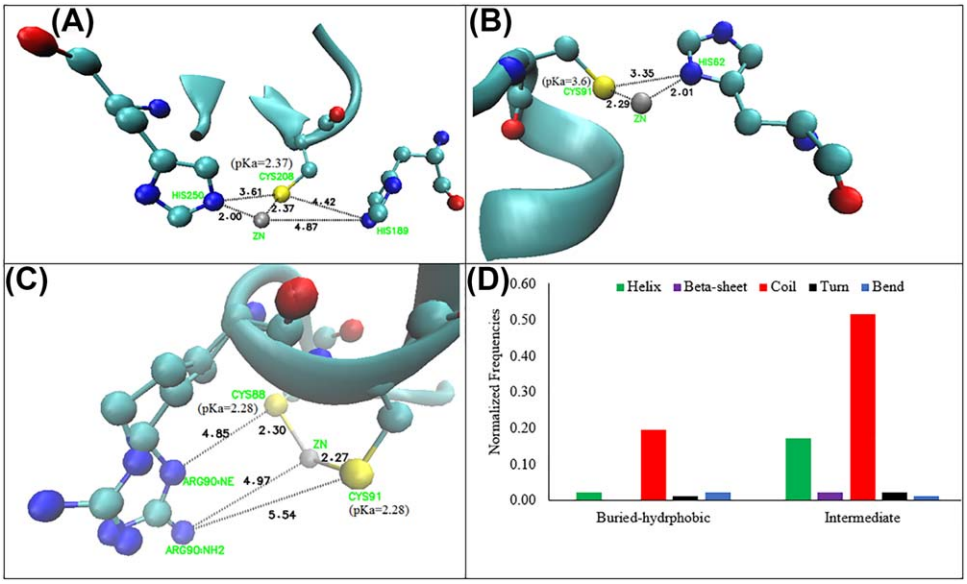


FIGURE 6 Cysteines with low pKa values having thiol group coordinated to metal ions and surrounded by Histidine residues. Cysteines are represented by yellow ball and metal ion by grey ball. Histidines are represented as ball and stick. Remaining part of the protein microenvironment around the cysteine thiol is depicted by cartoon representation for PDB ID: **A** 4HL2, **B** 1P60, **C** 2Z3H, and **D**, normalized frequencies for metal binding cysteines in different protein microenvironments and their related secondary structures. (As exposed-hydrophilic cluster has zero population, it is excluded from the figure) [Color figure can be viewed at wileyonlinelibrary.com]

3.6 | S-sulfenylation modification in enzymatic cysteines

Sulfenylation of cysteine is a post-translational modification mainly triggered via reactive oxygen species (ROS), like, hydrogen peroxide (H_2O_2) or reactive nitrogen species (RNS), like, peroxynitrite ($ONOO^-$). Total 306 enzymatic cysteines with S-Sulfenylation were identified in the current dataset. Out of these 306 sulfenylated cysteines, 63 were present in buried hydrophobic microenvironment cluster, 230 in intermediate cluster and 13 in exposed hydrophilic cluster (Table 2). The predominant secondary structures of S-Sulfenylated cysteines in intermediate microenvironment cluster were helix or coil (Figure 7A). However, the enzymatic sulfenylated cysteines were mostly part of beta-coil-helix motif (Figure 7B). Earlier report has also indicated the presence of S-sulfenylated cysteines in beta-coil-helix motif.[36] Helix or coil conformations were presumed to be energetically more susceptible toward reaction in comparison to beta conformation.[58,59] Hence, sulfenylated cysteines, in helix or coil conformation, were supposed to be more reactive. The reactive sulfenylated cysteines undergo further chemical modifications, namely, sulfinic acid formation, acetylation, disulfide bridge formation, and glutathionylation. Due to high reactivity, cysteine sulfenylation has direct or indirect implication in many redox-mediated biological pathways, like oxidative stress, regulatory effect on cardiovascular diseases,[60] acetylation in Krebs cycle,[61] S-glutathionylation in HMP shunt pathway.[62] One example has been cited from the current dataset, reactive S-sulfenylated-Cys438 in malate synthase A (PDB ID: 3CV2), facilitates binding acetyl group to the enzyme pocket.[63] This Cys438 was embedded in intermediate protein microenvironment (rHpy value of 0.69).

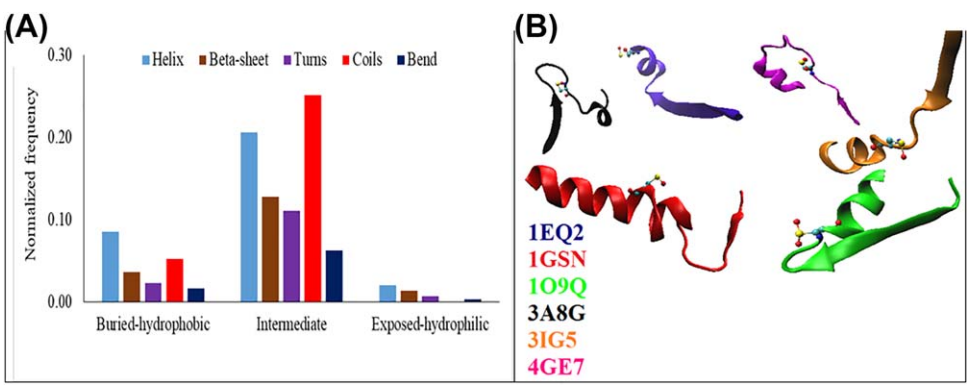


FIGURE 7 Protein microenvironment around cysteine sulfenylation for **A**, normalized frequencies of cysteines in different microenvironment clusters and their related secondary structures and **B**, beta-turn-helix motif (cartoon representation) containing S-Sulfenylated cysteines (yellow ball) [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Non-enzymatic cysteines harboring different chemical modifications

Cysteine modifications (501)	Buried hydrophobic (344)	Intermediate (122)	Exposed-hydrophilic (35)
Disulfide (17)	12	2	3
Thio-ether (16)	5	8	3
Metal-binding (367)	289	64	14
Nitrosylation (2)	2	0	0
Acetylation (4)	2	1	1
Palmitoylation (1)	0	1	0
Alkylation (2)	0	2	0
Glutathionylation (4)	3	1	0
Thioester (3)	1	2	0
Sulfenylation (85)	30	41	14

Cysteines present at similar functional domains belong to same PFAM family. S-Sulfenylated cysteines belong to many PFAM family, mainly, PF0069, PF01625, PF01965, and PF00578 (Table S5). Most of the S-Sulfenylated cysteines, in this dataset, belong to PFAM domain PF01965, that correspond to DJ-1/PfPI family. This family includes the protease PfpI and the corresponding domain belongs to transcriptional regulators. Earlier report has also shown *S*-sulfenylated cysteines belonging to PFAM family PF01965.[36]

3.7 | Acetylation modification in enzymatic cysteines

In the current dataset, 2 enzymatic cysteines were reported to undergo acetylation (PDB IDs: 4L8A, 5HWO). Both of the cysteines, capable to undergo acetylation, were embedded in intermediate microenvironment cluster (Table S8). These cysteines mainly belonged to transferase enzyme family combined with helical secondary structure (Figure 2).

3.8 | Selenylation modification in enzymatic cysteines

Selenylation of cysteines were triggered by selenites (SeO_3^{2-}) present in trace amount, mainly in bacteria[64] and plants.[65] Selenites were often involved in formation of reactive oxygen species (ROS) and subsequent cell damage. Cysteine selenylation provides a mechanism to remove selenite induced oxidative stress in cell. In the current dataset, only 1 selenylation is reported, Cys388 from a dehydrogenase protein present in eubacterium, *oligotropha carboxidovorans*. Both selenylated (PDB ID: 1N62) and free thiol (PDB ID: 1QJ2) form of the protein were identified. Changes in the protein microenvironment upon selenylation was moderate, rHpy of oxidized form was 0.305 and that of reduced form was 0.204.

3.9 | Ribosylation modification in enzymatic cysteines

Mono-ADP ribosylation was a protein modification that occurred at a number of amino acids, including cysteine.[66] Cysteine was the ADP-ribose acceptor cite of pertussis toxin-catalyzed ADP-ribosylation.[67] In the current dataset only 1 protein was identified with potential ribosylation site on cysteine (PDB ID: 1TKE, Cys182). This cysteine was located in buried hydrophobic microenvironment (buried fraction 0.86 and rHpy 0.14).

3.10 | Glutathionylation modification in enzymatic cysteines

Post-translational *S*-glutathionylation occurred through the reversible addition of a proximal donor of glutathione to thiolate anions of cysteines in target proteins. Glutathionylation altered the molecular mass, charge, and structure-function of sulfhydryl groups thus prevented it from over oxidation or proteolysis. Both the forward and the reverse reactions, catalyzed by glutathione *S*-transferase P and glutaredoxin, respectively, have created a functional cycle that could regulate certain protein functional clusters, including those involved in redox-dependent cell signaling.[17] In the current dataset, only 1 cysteine (PDB ID: 3RHB, Cys 29) was identified, that is deeply buried (buried fraction 1.0) in buried hydrophobic (rHpy 0.083) microenvironment.

3.11 | Chemical modification with non-enzymatic cysteines in current dataset

Apart from enzymatic reactions, cysteines have undergone similar chemical modifications during non-enzymatic processes. Nine chemical modifications were observed in the current dataset (Table 5). Maximum number of non-enzymatic cysteines were observed in buried hydrophobic microenvironment (Figure 8). Here we have compared the major (in terms of the numbers in the current dataset) non-enzymatic cysteine modifications (disulfide formation, thioether formation, metal binding, and S-Sulfenylation) with similar cysteine modifications in enzymatic reactions.

3.12 | Disulfide modification in non-enzymatic cysteines

Disulfide forming non-enzymatic cysteine population was maximum in buried hydrophobic microenvironment cluster (12 out of 17; Table 5). In case of disulfide forming enzymatic cysteines, maximum cysteine

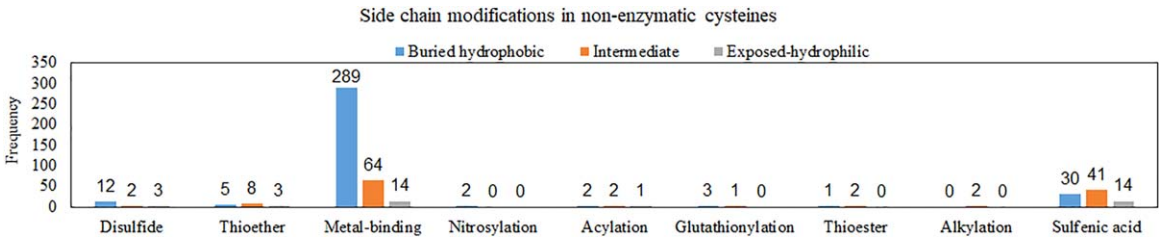


FIGURE 8 Frequencies of non-enzymatic cysteine side-chain modifications in different protein microenvironment clusters [Color figure can be viewed at wileyonlinelibrary.com]

population was observed in intermediate cluster, (38 out of 54; Table 3). Those 12 non-enzymatic cysteines (from buried hydrophobic cluster) belonged to diverse proteins, amidase, thioredoxin, ATP synthase, ferrodoxin, and so forth. In contrast, most of the disulfide forming enzymatic cysteines were part of CxxC motif from cytochrome C proteins (Figure 4A). Majority of these non-enzymatic cysteines adopted coil or turn conformation (Table S9). However, cysteines, in coil conformations, exhibited large standard deviation in rHpy values (0.115). pKa values (10.43 ± 3.52) for these cysteines, in coil conformations, exhibited larger fluctuations compared to those of disulfide forming enzymatic cysteines. This comparative analysis of disulfide forming enzymatic and non-enzymatic cysteines showed that the non-enzymatic disulfides have less conserved microenvironments and pKa values, presumably, owing to their diversity in protein functions.

Average pKa value for disulfide forming non-enzymatic cysteines, embedded in exposed-hydrophilic micro-environment cluster, was (9.72 ± 1.46). The average pKa value in buried-hydrophobic microenvironment was (11.60 ± 2.34). According to Henderson-Hasselbalch equation, at a given pH, concentration of base (thiolate) to acid (thiol) ratio increased with decrease in pKa. Hence, above observation suggested approximately 100 times increase in thiolate to thiol concentration when cysteine was embedded in exposed-hydrophilic protein microenvironments compared to buried-hydrophobic microenvironments. Oxidation of thiolate to disulfide was easier compared to oxidation of protonated thiol to disulfide.[49] Hence, cysteines embedded in exposed-hydrophilic microenvironment were more susceptible toward oxidation, in non-enzymatic processes. In literature, it was proposed that cysteine residues exposed on protein surfaces were mainly intra-mitochondrial thiol, involved in protection against oxidative damage.[5]

However, disulfide forming enzymatic cysteines embedded in exposed-hydrophilic micro-environment cluster have higher pKa values (11.69 ± 1.8) compared to those in intermediate micro-environment cluster (10.04 ± 1.2), suggesting 45 times increase in thiolate to thiol concentration of enzymatic cysteine when embedded in intermediate protein microenvironment. Thus, enzymatic disulfide forming cysteines are less susceptible toward oxidation when embedded in exposed-hydrophilic micro-environments.

From these above observations (based on pKa values) it can be inferred that exposed-hydrophilic protein microenvironment cluster facilitated easier oxidation to non-enzymatic cysteines. Whereas, the same exposed-hydrophilic cluster disfavored easy oxidation (because of high pKa values) of enzymatic cysteines (to form disulfide). Moreover, enzymatic cysteines were more easily converted to disulfide (inferred based on low pKa values) at enzyme active sites when embedded within intermediate microenvironment clusters.

3.13 | Thioether modification in non-enzymatic cysteines

Thioether forming non-enzymatic cysteine population was maximum in intermediate cluster (8 out of 16), similar to those of thioether forming enzymatic cysteines. These 8 cysteines mainly have beta sheet conformation (Table S10) and were present in variety of proteins, namely, oxidase, ribokinase, membrane proteins, histone lysine demethylase, and so forth. These cysteines were widely distributed throughout the microenvironment space, indicating less conservation of the microenvironments. Secondary structures of thioether forming non-enzymatic cysteines were different from those participated in enzymatic reactions. In contrast, thioether forming enzymatic cysteines, either have helix or turn secondary structure, mostly ligated to heme as a part of CxxC motif in cytochrome C proteins (Figure 5C). As non-enzymatic cysteines were not ligated to heme through CxxC motif, structural constraints of turn and helix were absent.

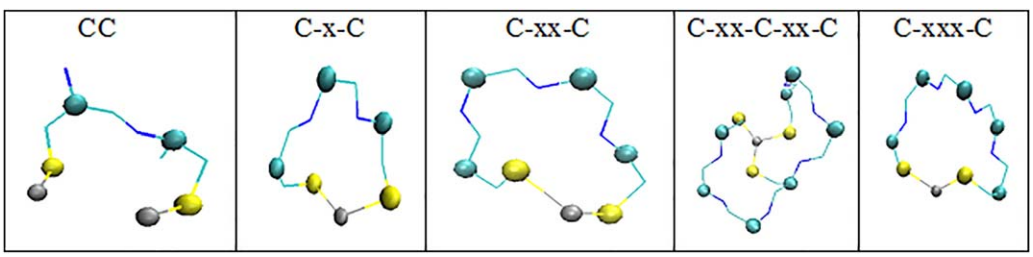


FIGURE 9 Schematic of different metal binding cysteine motifs [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 6 Different metal binding cystiene motifs

Metal	Motif	No. of protein	No. of cysteines	Functions of the proteins involved	Avg. Buried fraction (Std. dev)	Avg. rHpy (Std. dev)
Fe	CC	2	8 (6)#	Electron transport in respiration and removes superoxide anion (reduces stress)	0.86 (0.13)	0.18 (0.14)
Fe	C-x-C	5	22 (22)#	Electron transport in photosynthesis	0.95 (0.03)	0.11 (0.12)
Fe	C-xx-C	34	114 (112)#	Electron transport chain in photosynthesis and respiration, 2 Dehydrogenase, DNA glycosylase, oxygenase, isoprenoid and pyrimidine synthesis, formylation of glycine (posttranslational)	0.86 (0.12)	0.19 (0.14)
Fe	C-xxx-C	4	22 (22)#	iron-sulfur proteins involved in electron transfer in photosynthesis	0.85 (0.20)	0.26 (0.19)
Fe	C-xx-C-xx-C	5	30 (24)#	iron-sulfur proteins involved in photosynthesis and dehalogenases in respiration	0.85 (0.09)	0.14 (0.13)
Zn	CC	12	42	Anti-apoptotic, nucleotide and small molecule synthesis, DNA binding, regulation of cell cycle	0.86 (0.17)	0.21 (0.19)
Zn	C-x-C	19	60	Small molecule synthesis and degradation, nucleic acid binding	0.86 (0.17)	0.21 (0.17)
Zn	C-xx-C	110	628	Synthesis and degradation of nucleotides, amino acids and t-RNA	0.89 (0.18)	0.18 (0.18)
Zn	C-xxx-C	5	14	Transcription and cross bridging of myosin protein	0.88 (0.23)	0.21 (0.23)
Zn	C-xx-C-xx-C	3	18	Glucose oxidation and cell differentiation	0.92 (0.19)	0.15 (0.17)
Mn	CC	1	2	Pyrenoid biosynthesis	0.86 (0.19)	0.27 (0.27)
Cu	C-x-C	1	4	Chaperone	1.00 (0.00)	0.15 (0.17)
Cu	C-xx-C	1	4	Chaperone	0.60 (0.47)	0.44 (0.44)
Cu	C-xxx-C	1	2	Cytochrome oxidase	1.00 (0.00)	0.22 (0.12)
Hg	C-xx-C	1	2	Chaperone	0.63 (0.37)	0.51 (0.26)
Hg	C-xxx-C	1	4	Ribonucleotide reductase	0.95 (0.05)	0.07 (0.13)

number of cysteines involved in electron transfer are given in parenthesis.

3.14 | Metal binding modification in non-enzymatic cysteines

Maximum number of non-enzymatic cysteines were part of metal coordination (Table 5). Most of these cysteines (289) belong to buried-hydrophobic microenvironment cluster (Table S11), similar to those of metal-binding enzymatic cysteines. These cysteines embedded in buried-hydrophobic cluster mostly belong to hydrolase enzyme family. Most of these metal binding non-enzymatic cysteines (199 out of 367) have adopted coil or turn conformations, similar to those of metal binding enzymatic cysteines (Table S11). Overall pKa values (9.32 ± 2.96) for 367 non-enzymatic cysteines were comparable to the default thiol pKa value (9.0). Relative comparison of metal binding cysteines in enzymatic and non-enzymatic groups revealed their similar microenvironments, secondary structures and pKa values. Above observation indicated no difference in metal coordination during enzymatic or non-enzymatic reactions.

Although no differences were observed between protein microenvironments of metal binding enzymatic and non-enzymatic cysteines, characteristic features were observed for specific metal binding motifs. To increase the significance of statistical analysis, size of the dataset was increased by incorporating crystal structures with resolutions better than 2.0 Å (in contrast to existing dataset with resolution cutoff 1.5 Å). In the enlarged dataset, total 976 cysteines were identified as metal binding, extracted from 205 proteins. There were 5 different metal binding cysteine motifs, CC, CxC, CxxC, CxxCxxC, and CxxxC (Figure 9), associated with 7 different metal ion. Different metal-binding cysteine motifs, variations in their embedded protein microenvironments, individual cysteine functions and related protein functions were reported (Table 6). Microenvironments around iron binding and zinc

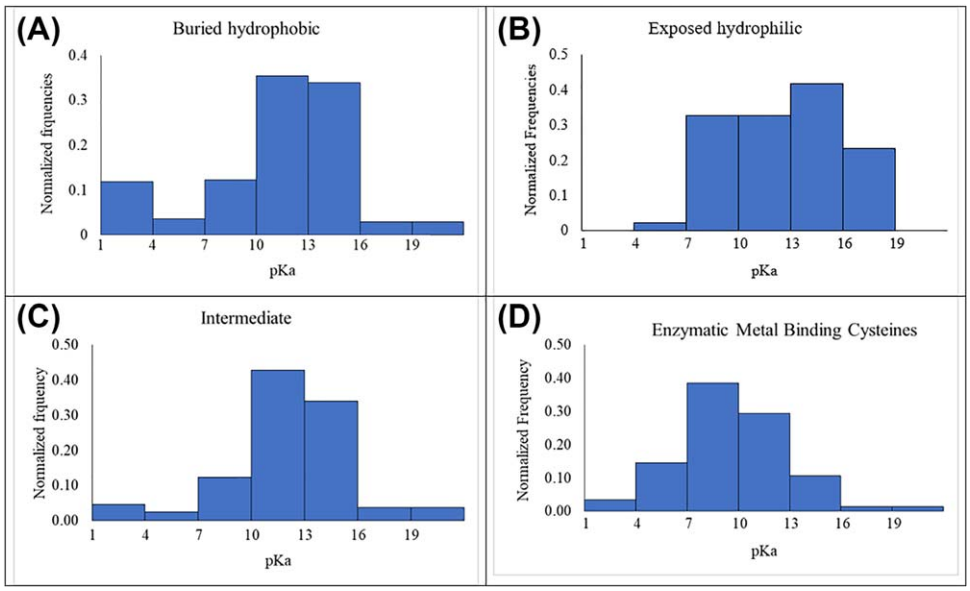


FIGURE 10 Normalized frequencies of cysteine thiol pKa values present in different protein microenvironment clusters—(1) buried hydrophobic, (2) intermediate and (3) exposed hydrophilic, and in (4) metal binding enzymatic cysteines. Normalization done with respect to the total number of cysteines in each cluster [Color figure can be viewed at wileyonlinelibrary.com]

binding motifs were discussed below due to their abundance in the current dataset.

Iron binding cysteine motifs have similar functions, mostly evolved to transport electrons in cellular respiration and photosynthesis (Table 6). Standard deviation of rHpy values for all iron binding cysteines was 0.14. Majority of the iron binding cysteines adopted coil or turn secondary structures. In contrast zinc binding cysteine motifs were involved in variety of functions (Table 6; Table S12). Cysteines involved in zinc binding motifs were more scattered throughout the protein microenvironment space, compared to iron binding cysteine motifs. Standard deviation of rHpy values for zinc binding cysteine motifs was 0.19, less compact than that of iron binding cysteine motifs.

Above observations suggested that protein micro-environments were conserved around iron-binding cysteines those participated in similar biological function, namely, electron transport. Evolution of similar amino acid functions in near identical protein microenvironments presumably facilitate different proteins to perform same function.

3.15 | S-sulfenylation modification in non-enzymatic cysteines

There were 85 non-enzymatic sulfenylated cysteines. 30 cysteines were in buried-hydrophobic, 40 in intermediate cluster and 15 in exposed-hydrophilic cluster. Slight increase in proportion of non-enzymatic sulfenylated cysteines (30/85 = 0.35) was observed in buried-hydrophobic cluster compared to enzymatic sulfenylated cysteines (63/306 = 0.21) in the same cluster. This indicated that non-enzymatic sulfenylated cysteines were more often present in buried protein microenvironment compared to enzymatic cysteines. For example; S-sulfenylation was reported for Cys322 in SusD protein of human gut microbiota (PDB ID: 3CKC), however, no enzymatic function was known. This Cys322 was embedded in hydrophobic microenvironment created by 3 Trp residues, W96, W98, and W320 (rHpy, 0.06). This hydrophobic cleft facilitates starch binding and activate the protein function.[68] Literature report has indicated that non-enzymatic sulfenylated cystines were mostly involved in oxidation to form stable disulfide bond via formation of thiolate intermediate.[69] Non-enzymatic S-Sulfenylated cysteines mainly adopt coil or turn conformation in buried hydrophobic protein microenvironment; and helix conformation in intermediate and exposed hydrophilic microenvironment (Table S13). In contrast, S-Sulfenylated enzymatic cysteines preferred helical conformation in buried-hydrophobic protein microenvironment as well (Figure 7A).

3.16 | Other chemical modifications of non-enzymatic cysteines

Non-enzymatic cysteines have more variety of chemical modifications compared to enzymatic cysteines, namely, nitrosylation, palmytoylation, glutathionylation, and thioesterification (Tables 2 and 5). However, the number of modifications in the current dataset were too less to draw any statistical inference. These chemical modifications on cysteines were more non-specific compared to those in enzymatic reactions, still has important consequences in the biochemical pathways. Example 1, Cys83 in *Dimethylarginine dimethylaminohydrolase* protein (PDB ID: 2C6Z) has undergone trans-nitrosylation reaction by S-nitroso-L-homocysteine inhibitor under crystallization conditions.[70] S-ntiroso-L-homocysteine inhibitor formed a covalent thiosulphoximide bond with cysteine273 in the same protein. Example 2, Cysteine87 of Glutathioredoxin (PDB ID: 3RHB) was one of the site for glutathionylation outside the active site of the enzyme and involved in reduction of glutathionylated protein.[71] Details of the microenvrionments around non-enzymatic cysteine modifications (other than disulfide, thioether, metal-binding, and S-Sulfenylation) were reported in Table S14.

3.17 | Effects of embedded protein microenvironments and other factors (thiol modifications and so forth) on cysteine thiol pKa

In order to understand the effects of protein microenvironment on cysteine side-chain pKa values, we have studied the distribution of cysteine pKa values (computed by PROPKA) in 3 different microenvironment clusters (Figure 10). The observations were counter-intuitive. The intuition was, the cysteine pKa value (negative logarithm of proton dissociation constant of the cysteine -SH group) should reduce in presence of higher dielectric medium (qualitatively comparable to high rHpy value), according to Coloumb's law. However, buried hydrophobic cluster has maximum number of cysteines with low pKa values (Figure 10A) and exposed hydrophilic cluster has zero number of cysteines with low pKa values (pKa ranging from 1 to 4) (Figure 10B). This observation indicated that cysteine pKa values do not decrease proportionally with the hydrophilicity of the embedded protein microenvironment. Rather, cysteine pKa values were more related to its respective functions. For example, metal binding cysteines were reported with low pKa values (Figure 10D) and most of the metal binding cysteines (75% of the metal binding cysteines) were located in buried hydrophobic microenvironment cluster (Table S11). Second example was about enzymatic cysteine thiols undergoing disulfide formation. Enzymatic cysteine thiols exhibited lower pKa when susceptible toward oxidation to form disulfide and present in buried-hydrophobic or intermediate microenvironments (compared to exposed-hydrophilic region) (Table 4A). Above 2 examples explained why buried hydrophobic cluster was populated with significant number of cysteines with low pKa values.

Presumably, lowering of cysteine thiol pKa inside proteins was result of competitive interactions between strong force (e.g., formation of coordinate bonds between thiolate and metal ions) versus weak force (non-covalent electrostatic interaction, like, effect of protein dielectric medium or protein microenvironment). Strong forces predominate over the weak force and exhibit changes in thiol pKa (with respect to the default pKa value).

Thus, cysteine thiol pKa values inside proteins are governed by multiple factors, chemical modifications, secondary structures (discussed above), and protein microenvironments. Details study of pKa modulation for titrable amino acid side-chain by protein structures will be reported elsewhere.

4 | CONCLUSIONS

Protein microenvironment played a crucial role in modulating the structure and catalytic activity of disulfide-bridged cysteine.[22,23] In this work, we have demonstrated the functions (chemical modifications) of cysteine thiol (reduced form of disulfide-bridged cystine) side chains embedded in different protein microenvironments and secondary structures. Biological functions of cysteine analyzed in the current dataset were disulfide formation, thio-ether formation, metal-binding, nitrosylation, acylation, selenylation, glutathionylation, sulfenylation, and ribosylation. Based on hierarchical clustering, 3 microenvironment clusters were identified, buried-hydrophobic, intermediate, and exposed-hydrophilic. Comparison of same cysteine modifications across enzymatic and non-enzymatic groups (enzymatic and non-enzymatic groups were defined in the method section) indicated clear difference in their embedded protein microenvironments and secondary structures. Enzymatic cysteines analyzed in the current dataset, those have undergone a particular chemical modification (disulfide or thioether formation), mostly belonged to similar structural motifs from same enzyme class and embedded in similar microenvironments and secondary structures. For example, disulfide forming enzymatic cysteines mostly belonged to CxxC motif of electron transport family. Those cysteines preferred helical conformation and mostly confined in intermediate cluster. However, disulfide forming non-enzymatic cysteines did not belong to any specific structural motifs and distributed over various protein types. Similar observations were also made for enzymatic and non-enzymatic thioether forming cysteines. Based on the pKa calculations it was observed that enzymatic cysteine thiols were more susceptible to oxidation (to disulfide) when embedded in buried-hydrophobic or intermediate microenvironments. In contrast, the non-enzymatic cysteine thiols were more susceptible to oxidation (to disulfide) when embedded in exposed-hydrophilic microenvironments. Cysteine modifications, like, glutathionylation, ribosylation, nitrosylation and selenylation and so forth were extremely important in redox balance, cellular signaling, heavy-metal scavenging and so forth. However, due to limited data in our current dataset, no specific conclusions can be made for those modifications. Details for those modifications will be reported elsewhere. Protein microenvironments around iron-binding cysteine motifs were relatively more conserved than zinc binding cysteine motifs. Above observations suggested that iron binding cysteines contribute to same function (electron transport) in different proteins.

Initial hypothesis of this work was—individual cysteine functions preferred specific protein microenvironment, secondary structure and biological function. This hypothesis was proved for enzymatic cysteine functions only. The key finding was—enzymatic cysteine functions (and their related pKa values) from different proteins strongly preferred specific combination of protein microenvironments, secondary structures and biological functions. The key finding can be used to manipulate biological and pharmacological reactions involving thiol containing compounds (like, glutathione and thioredoxin)[72] or to design pharmacologically important exogenous anti-oxidant molecules.[73] Earlier report has suggested (post-translational) modifications (sulfenylation, glutathionylation, and so forth) of cysteine residue (Cys797 residue of EGFR protein) affect drug pharmacology (drugs involved—Dacomitinib, Afatinib, Gefitinib, Erlotinib, Cetuximab, Lapatinib, Panitumumab, Vandetanib, and so forth).[74] Another report has shown that cysteine rich engineered proteins can regulate hypoxic condition within the cell.[2,75] Cysteine-based drug and metal chelator, meso-2,3-dimercaptosuccinic acid (DMSA), was clinically approved to reduce lead levels in blood stream.[76] Conversion of cysteine to Cys-sulfenic acid by cysteine dioxygenase reduced the levels of cysteine in the cellular environment.[77] Here, we suggest that drug candidatures against NADPH oxidase (NOX) protein family can be improved by chemically modifying the cysteine residues involved. The existing drugs against NOX family,

those have undergone clinical trials, were GKT136901, GKT137831,[78] and VAS3947.[79] Further possibilities, modification of protein microenvironments around critical thiols involved in mitochondrial ROS generation might help to design effective drugs against mitochondrial disorders.

ACKNOWLEDGMENTS

The authors sincerely acknowledge Dr. Marcin I Apostol, ADRx Inc, California, USA for thoroughly reading the manuscript and giving his valuable comments. D.B. acknowledge University Grants Commission: Basic Research Start-up Grant, India to support this research work; Grant Number is: 20-4(21)/2012(BSR).

ORCID

Debashree Bandyopadhyay http://orcid.org/0000-0003-4131-907X

REFERENCES

[1] Harris TK, Turner GJ. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life*. 2002;53(2):85–98.

[2] Giles NM, Watts AB, Giles GI, Fry FH, Littlechild JA, Jacob C. Metal and redox modulation of cysteine protein function. *Chem Biol*. 2003;10(8):677–693.

[3] Jacob C, Giles GI, Giles NM, Sies H. Sulfur and selenium: the role of oxidation state in protein structure and function. *Angew Chemie Int Ed*. 2003;42(39):4742–4758.

[4] Gutteridge A, Thornton JM. Understanding nature's catalytic toolkit. *Trends Biochem Sci*. 2005;30(11):622–629.

[5] Requejo R, Hurd TR, Costa NJ, Murphy MP. Cysteine residues exposed on protein surfaces are the dominant intramitochondrial thiol and may protect against oxidative damage. *FEBS J*. 2010;277(6):1465–1480.

[6] Jones W. *Inorganic Chemistry*. Philadelphia, USA: Blakiston; 1949.

[7] Pauling L. *General Chemistry*., New York, USA: Dover Publications Inc.; 1988.

[8] Woo HA, Jeong W, Chang T-S, et al. Reduction of cysteine sulfinic acid by sulfiredoxin is specific to 2-cys peroxiredoxins. *J Biol Chem*. 2005;280:3125–3128.

[9] Klomsiri C, Karplus PA, Poole LB. Cysteine-based redox switches in enzymes. *Antioxid. Redox Signal*. 2011;14(6):1065–1077.

[10] Houk J, Singh R, Whitesides GM. Measurement of thiol-disulfide interchange reactions and thiol pKa values. *Methods Enzymol*. 1987;143:129–140.

[11] Mehler EL, Guarnieri F. A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys. J*. 1999;77:3–22.

[12] Shan J, Mehler EL. Calculation of pK(a) in proteins with the microenvironment modulated-screened coulomb potential. *Proteins*. 2011;79(12):3346–3355.

[13] Raso SW, Clark PL, Haase-Pettingell C, King J, Thomas GJ. Distinct cysteine sulfhydryl environments detected by analysis of Raman S-H markers of Cys→Ser mutant proteins11Edited by P. E. Wright. *J Mol Biol*. 2001;307:899–911.

[14] Lakowicz J. *Principles of Fluorescence Spectroscopy*. New York, USA: Plenum Press; 1983.

[15] Bhandary B, Marahatta A, Kim H-R, Chae H-J. An involvement of oxidative stress in endoplasmic reticulum stress and its associated diseases. *Int J Mol Sci*. 2012;14:434–456.

[16] Wu C, Belenda C, Leroux J-C, Gauthier MA. Interplay of chemical microenvironment and redox environment on thiol-disulfide exchange kinetics. *Chemistry*. 2011;17(36):10064–10070.

[17] Grek CL, Zhang J, Manevich Y, Townsend DM, Tew KD. Causes and consequences of cysteine S-glutathionylation. *J Biol Chem*. 2013;288(37):26497–26504.

[18] Zeigler MM, Doseff AI, Galloway MF, et al. Presentation of nitric oxide regulates monocyte survival through effects on caspase-9 and caspase-3 activation. *J Biol Chem*. 2003;278:12894–12902.

[19] Mannick JB, Schonhoff C, Papeta N, et al. S-Nitrosylation of mitochondrial caspases. *J Cell Biol*. 2001;154(6):1111–1116.

[20] Svoboda LK, Reddie KG, Zhang L, et al. Redox-sensitive sulfenic acid modification regulates surface expression of the cardiovascular voltage-gated potassium channel Kv1.5. *Circ Res*. 2012;111:842–853.

[21] Seo YH, Carroll KS. Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies. *Proc Natl Acad Sci USA*. 2009;106:16163–16168.

[22] Bandyopadhyay D, Mehler EL. Quantitative expression of protein heterogeneity: Response of amino acid side chains to their local environment. *Proteins*. 2008;72(2):646–659.

[23] Bhatnagar A, Apostol M, Bandyopadhyay ID. Amino acid function relates to its embedded protein microenvironment: a study on disulfide-bridged cystine. *Proteins*. 2016;84(11):1576–1589.

[24] Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*. 2000;28:235–242. http://www.rcsb.org. Accessed March 20, 2014.

[25] Brooks BR, Brooks CL, Mackerell AD, et al. CHARMM: The biomolecular simulation program. *J Comput Chem*. 2009;30:1545–1614.

[26] Rekker *The Hydrophobic Fragmental Constant*. Amsterdam: Elsevier; 1977.

[27] Pascual-Ahuir JL, Silla E, Tunon I. GEPOL: an improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *J Comput Chem*. 1994;15:1127–1138.

[28] Tryon DE, Bailey RC. *Cluster Analysis*. New York, USA: McGraw-Hill; 1973.

[29] Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. *J Classif*. 2014;31(3):274–295.

[30] Addinsoft. XLSTAT 2014, http://www.xlstat.com. Data analysis and statistics software for Microsoft Excel. 2014.

[31] Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*. 2005;61(4):704–721.

[32] Davies MN, Toseland CP, Moss DS, Flower DR. Benchmarking pK(a) prediction. *BMC Biochem*. 2006;7:18.

[33] Walker RC, Crowley MF, Case DA. The implementation of a fast and accurate QM/MM potential method in Amber. *J Comput Chem*. 2008;29(7):1019–1031.

[34] Aliev AE, Kulke M, Khaneja HS, Chudasama V, Sheppard TD, Lanigan RM. Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics. *Proteins*. 2014;82(2):195–215.

[35] Awoonor-Williams E, Rowley CN. Evaluation of methods for the calculation of the pKa of cysteine residues in proteins. *J Chem Theory Comput*. 2016;12(9):4662–4673.

[36] Defelipe LA, Lanzarotti E, Gauto D, Marti MA, Turjanski AG, Wade RC. Protein topology determines cysteine oxidation fate: the case of sulfenyl amide formation among protein families. *PLOS Comput Biol*. 2015;11(3):e1004051

[37] Joosten RP, Te Beek TAH, Krieger E, et al. A series of PDB related databases for everyday needs. *Nucleic Acids Res*. 2011;39(Database):D411–D419.

[38] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577–2637.

[39] Touw WG, Baakman C, Black J, et al. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res*. 2015;43(Database issue):D364–D368.

[40] Bairoch A. The ENZYME database in 2000. *Nucleic Acids Res*. 2000; 28(1):304–305.

[41] Furnham N, Holliday GL, de Beer TAP, Jacobsen JOB, Pearson WR, Thornton JM. The catalytic site atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res*. 2014;42(D1): D485–D489.

[42] de Beer TAP, Berka K, Thornton JM, Laskowski RA. PDBsum additions. *Nucleic Acids Res*. 2014;42(Database issue):D292–D296.

[43] Shrestha RK, Ronau JA, Davies CW, et al. Insights into the mechanism of deubiquitination by JAMM deubiquitinases from cocrystal structures of the enzyme with the substrate and product. *Biochemistry*. 2014;53(19):3199–3217.

[44] Sammut SJ, Finn RD, Bateman A. Pfam 10 years on: 10 000 families and still growing. *Brief Bioinform*. 2008;9(3):210–219.

[45] Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–D285.

[46] The UniProt ConsortiumActivities at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. 2014;42(11):D191–D198.

[47] U. Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45:D158–D169.

[48] Taylor AB, Benglis DM, Dhandayuthapani S, Hart PJ. Structure of *Mycobacterium tuberculosis* methionine sulfoxide reductase A in complex with protein-bound methionine. *J Bacteriol*. 2003;185(14): 4119–4126.

[49] Winther JR, Thorpe C. Quantification of thiols and disulfides. *Biochim Biophys Acta*. 2014;1840(2):838–846.

[50] Zhang Q, Schenauer MR, McCarter JD, Flynn GC. IgG1 thioether bond formation in vivo. *J Biol Chem*. 2013;288(23):16371–16382.

[51] Lassmann T, Frings O, Sonnhammer ELL. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res*. 2009;37(3):858–865.

[52] Kutter S, Weiss MS, Wille G, Golbik R, Spinka M, König S. Covalently bound substrate at the regulatory site of yeast pyruvate decarboxylases triggers allosteric enzyme activation. *J Biol Chem*. 2009;284:12136–12144.

[53] Leys D, Meyer TE, Tsapin AS, Nealson KH, Cusanovich MA, Van Beeumen JJ. Crystal structures at atomic resolution reveal the novel concept of electron-harvesting; as a role for the small tetraheme cytochrome c. *J Biol Chem*. 2002;277:35703–35711.

[54] Stelter M, Melo AMP, Pereira MM, et al. A novel type of monoheme cytochrome c: biochemical and structural characterization at 1.23 A resolution of rhodothermus marinus cytochrome *c*. *Biochemistry*. 2008;47(46):11953–11963.

[55] Bandyopadhyay D, Bhattacharyya D. Different modes of interaction between hydrated magnesium ion and DNA functional groups: database analysis and ab initio studies. *J Biomol Struct Dyn*. 2003; 21:447–458.

[56] Robertson MP, Scott WG. The structural basis of ribozyme-catalyzed RNA assembly. *Science*. 2007;315(5818):1549–1553.

[57] Van Laer K, Oliveira M, Wahni K, Messens J. The concerted action of a positive charge and hydrogen bonds dynamically regulates the pKa of the nucleophilic cysteine in the NrdH-redoxin family. *Protein Sci*. 2014;23(2):238–242.

[58]]Maiorov V, Abagyan R. Energy strain in three-dimensional protein structures. *Fold Des*. 1998;3(4):259–269.

[59] Simone AD, Berisio R, Zagari A, Vitagliano L. Limited tendency of alpha-helical residues to form disulfide bridges: a structural explanation. *J Pept Sci*. 2006;12:740–747.

[60] Pan J, Carroll KS, Martin BR. Chemical biology approaches to study protein cysteine sulfenylation. *Biopolymers*. 2014;101(2):165–172.

[61] Voet D, Voet JG, Pratt CW. *Principles of Biochemistry*. Hoboken, NJ: Wiley; 2008.

[62] Clancy RM, Levartovsky D, Leszczynska-Piziak J, Yegudin J, Abramson SB. Nitric oxide reacts with intracellular glutathione and activates the hexose monophosphate shunt in human neutrophils: evidence for S-nitrosoglutathione as a bioactive intermediary. *Proc Natl Acad Sci USA*. 1994;91:3680–3684.

[63] Lohman JR, Olson AC, Remington SJ. Atomic resolution structures of *Escherichia coli* and *Bacillus anthracis* malate synthase A: comparison with isoform G and implications for structure-based drug discovery. *Protein Sci*. 2008;17(11):1935–1945.

[64] Bebien M, Chauvin JP, Adriano JM, Grosse S, Verméglio A. Effect of selenite on growth and protein synthesis in the phototrophic bacterium *Rhodobacter sphaeroides*. *Appl Environ Microbiol*. 2001;67: 4440–4447.

[65] Zhao B, Zhang J, Yao J, Song S, Yin Z, Gao Q. Selenylation modification can enhance antioxidant activity of *Potentilla anserina* L. polysaccharide. *Int J Biol Macromol*. 2013;58:320–328.

[66] McDonald LJ, Moss J. Enzymatic and nonenzymatic ADP-ribosylation of cysteine. *Mol Cell Biochem*. 1994;138(1–2):221–226.

[67] West RE, Moss J, Vaughan M, Liu T, Liu TY. Pertussis toxin-catalyzed ADP-ribosylation of transducin. Cysteine 347 is the ADP-ribose acceptor site. *J Biol Chem*. 1985;260:14428–14430.

[68] Koropatkin NM, Martens EC, Gordon JI, Smith TJ. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure*. 2008;16(7): 1105–1115.

[69] Beedle AEM, Lynham S, Garcia-Manyes S. Protein S-sulfenylation is a fleeting molecular switch that regulates non-enzymatic oxidative folding. *Nat Commun*. 2016;7:12490

[70] Frey D, Braun O, Briand C, Vasák M, Grütter MG. Structure of the mammalian NOS regulator dimethylarginine dimethylaminohydrolase: a basis for the design of specific inhibitors. *Structure*. 2006;14 (5):901–911.

[71] Couturier J, Stroher E, Albetel A-N, et al. Arabidopsis chloroplastic glutaredoxin C5 as a model to explore molecular determinants for iron-sulfur cluster binding into glutaredoxins. *J Biol Chem*. 2011; 286:27515–27527.

[72] Dickinson DA, Forman HJ. Cellular glutathione and thiols metabolism. *Biochem Pharmacol*. 2002;64:1019–1026.

[73] Dao VT-V, Casas AI, Maghzal GJ, et al. Pharmacology and clinical drug candidates in redox medicine. *Antioxid Redox Signal*. 2015;23: 1113–1129.

[74] WaniNagata RA, Murray BW. Protein redox chemistry: post-translational cysteine modifications that regulate signal transduction and drug pharmacology. *Front Pharmacol*. 2014;5:224

[75] Cymes G, Grosman DC. Engineered ionizable side chains, In: Ahern C, Pless S, eds. *Novel Chemical Tools to Study Ion Channel Biology*. Advances in Experimental Medicine and Biology, vol 869. Springer, New York, NY. 2015: 5–23.

[76] Graziano JH. Conceptual and practical advances in the measurement and clinical management of lead toxicity. *Neurotoxicology*. 1993;14(2–3):219–223.

[77] Davies CG, Fellner M, Tchesnokov EP, Wilbanks SM, Jameson GNL. The cys-tyr cross-link of cysteine dioxygenase changes the optimal pH of the reaction without a structural change. *Biochemistry*. 2014; 53(50):7961–7968.

[78] Monika S, Richard J, Suzel D, Patti H, Gülsah G, G KU. Mucosal reactive oxygen species are required for antiviral response: role of duox in influenza A virus infection. *Antioxid Redox Signal*. 2014;20:2695–2709.

[79] Stielow C, Catar RA, Muller G, et al. Novel Nox inhibitor of oxLDL-induced reactive oxygen species formation in human endothelial cells. *Biochem Biophys Res Commun*. 2006;344(1):200–205.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Bhatnagar A, Bandyopadhyay D. Characterization of cysteine thiol modifications based on protein microenvironments and local secondary structures. *Proteins*. 2018;86:192–209. https://doi.org/10.1002/prot.25424