PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

# Amino acid function relates to its embedded protein microenvironment: A study on disulfide-bridged cystine

**Akshay Bhatnagar,[1] Marcin I. Apostol,[2] and Debashree Bandyopadhyay[1]***

[1] Department of Biological Sciences, Birla Institute of Technology and Science, Hyderabad 500078, India

[2] ADRx. Inc. 515 Marin St., Suite 314, Thousand Oaks, California 91360

**ABSTRACT**

**In our previous study, we have shown that the microenvironments around conserved amino acids are also conserved in protein families (Bandyopadhyay and Mehler, Proteins 2008; 72:646–659). In this study, we have hypothesized that amino acids perform similar functions when embedded in a certain type of protein microenvironment. We have tested this hypothesis on the microenvironments around disulfide-bridged cysteines from high-resolution protein crystal structures. Although such cystines mainly play structural role in proteins, in certain enzymes they participate in catalysis and redox reactions. We have performed and report a functional annotation of enzymatically active cystines to their respective microenvironments. Three protein microenvironment clusters were identified: (i) buried-hydrophobic, (ii) exposed-hydrophilic, and (iii) buried-hydrophilic. The buried-hydrophobic cluster encompasses a small group of 22 redox-active cystines, mostly in alpha-helical conformations in a –C-x-x-C- motif from the Oxido-reductase enzyme class. All these cystines have high strain energy and near identical microenvironments. Most of the active cystines in hydrolase enzyme class belong to buried hydrophilic microenvironment cluster. In total there are 34 half-cystines detected in buried hydrophilic cluster from hydrolases, as a part of enzyme active site. Even within the buried hydrophilic cluster, there is clear separation of active half-cystines between surface exposed part of the protein and protein interior. Half-cystines toward the surface exposed region are higher in number compared to those in protein interior. Apart from cystines at the active sites of the enzymes, many more half-cystines were detected in buried hydrophilic cluster those are part of the microenvironment of enzyme active sites. However, no active half-cystines were detected in extremely hydrophilic microenvironment cluster, that is, exposed hydrophilic cluster, indicating that total exposure of cystine toward the solvent is not favored for enzymatic reactions. Although half-cystines in exposed-hydrophilic clusters occasionally stabilize enzyme active sites, as a part of their microenvironments. Analysis performed in this work revealed that cystines as a part of active sites in specific enzyme families or folds share very similar protein microenvironment regions, despite of their dissimilarity in protein sequences and position specific sequence conservations.**

Proteins 2016; 84:1576–1589.


**Key words: amino acid function; embedded protein microenvironment; disulphide bridged cystine; half-cystine; redox active cystine; enzyme class; enzyme active site; protein microenvironment cluster; protein dielectric medium; sequence conservation.**

## INTRODUCTION

Properties of amino acids are influenced by the surrounding dielectric medium, like water/octanol,[1] protein interior[2–4] and membrane bilayers.[5–8] The protein interior is highly heterogeneous in terms of its dielectric medium; the environment created by the complex mosaic of 20 different amino acids. This heterogeneous protein dielectric medium can be viewed as a collection of smaller dielectric media, what we refer to here as the

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: University Grants Commission; Basic Research Start-up grant, India; BITS-PILANI, research initiation Grant.

*Correspondence to: Debashree Bandyopadhyay, Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Hyderabad campus, Hyderabad 500078, India. E-mail:banerjee_debi@yahoo.com; banerjee.debi@hyderabad.bits-pilani.ac.in

Received 2 April 2015; Revised 30 June 2016; Accepted 3 July 2016
Published online 13 July 2016 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.25101

local microenvironment. Microenvironment describes the three dimensional arrangement of atoms around any given amino acid (or its functional group) within its first contact shell.[9] It has been shown earlier that certain properties of amino acids, for example $pK_a$ in titrable amino acids, can be altered due to transition between different microenvironments.[4] It has also been shown that protein microenvironments are crucial in modulating the protonation states of titratable amino acids such as arginine in voltage-gated ion channels,[10] aspartic acid in photo cycle of bacteriorhodopsin,[11] glutamic acid in hen egg white lysozyme, and others in several acid-base catalyzed hydrolysis reactions.[12] Despite of all these facts, there is a gap in knowledge of how specifically microenvironment influences the particular role of an amino acid in a protein structure. This is due to insufficient biochemical (experimental) information on individual amino acid microenvironment inside the protein. Microenvironment measurement is experimentally limited only to certain surface exposed amino acid residues which are intrinsic fluorophores or those can be tagged with extrinsic fluorophores.[13] Recently we have generated a database of microenvironments around all the amino acid side chains in the context of high-resolution protein crystal structures.[9] In that work, we have demonstrated that each amino acid side-chain is embedded in a wide range of protein microenvironment. The average hydrophobicity of the microenvironment is similar to that of the embedded amino acid side chain. 20–30% of the protein microenvironments significantly differ from that of the corresponding amino acid, those termed as "mismatched microenvironment." It has also been shown that amino acids in these mismatched microenvironments were evolved to perform specific structure or functional roles in the protein, Moreover that work has also demonstrated that the microenvironments around the conserved residues in respective protein family are also conserved. These observations triggered the question whether different sets of microenvironments are associated with different kinds of functions in an amino acid. Hypothesis based on this question has been tested here on disulfide-bridged cystine residues embedded in the protein microenvironments obtained from an updated microenvironment database. Cystine residues occur by the oxidation of two cysteine amino acids covalently linking them through a disulfide bond. This was chosen as model system because this residue has a limited and simple functional role in proteins, it either plays a structural role or in an enzymatic capacity participates in redox reactions. Out of the six broad enzyme classes (according to the nomenclature committee of International Union of Biochemistry and Molecular Biology, IUBMB)[14] cystines are directly involved in catalytic and redox reactions in the oxidoreductase (thioreductase fold)[15] and hydrolase families.[16]

Here, we analyze the previously unexplored cystine functions (here functions imply presence of a cystine in enzyme active site or within its microenvironment) in various protein microenvironments. In this study, protein microenvironments were clustered using hierarchical clustering. Functions in entire microenvironment database (mainly out of the enzymes) were curated from literature and different websites. Grouping of cystine functions based on their microenvironment clusters are discussed in the result section. The notable observation from this study is that the cystines those are part of active sites in specific enzyme classes share similar microenvironments despite of their dissimilarity in overall protein sequences.

## METHODOLOGY

### Description of microenvironment dataset around disulfide-bridged cystine residues

The current microenvironment dataset around disulfide-bridged cystine residues contains total 5084 cystines from 1303 high resolution protein crystal structures. Part of the dataset was obtained from our previous work, containing 700 disulfide-bridged cystine residues from 175 protein crystal structures.[9] Remaining 4384 cystines were curated from recent PDB[17] entries, from January 2004 to April 2016. Protein crystal structures with resolution better than 2 Å and sequence similarity $<$30% were selected. These proteins were not complexed with nucleic acid molecules. Proteins containing modified residues were not included in the dataset. The current selection resulted into 9090 protein structures. Out of this current selection, only 1128 proteins were detected with cystines from PDB header files using SSBOND keyword (see Supporting Information for PDB IDs, Table SI).

The microenvironment around each cystine was described by the amino acid fragments within a given radius, where the radius varies from one atom type to the other [Fig. 1(a)]. The microenvironment space around individual cystine was described using two parameters, "buried fraction" (BF) and hydrophobicity/hydrophilicity (rHpy).[9] Buried fraction is the fraction of the side-chain of an amino acid which is buried inside the protein and not exposed to outside environment. This parameter was computed using the program GEPOL93 that computes the solvent-excluded surface by filling the solvent-inaccessible spaces with a new set of spheres.[19] Three parameters were used in the computation, the number of triangles per sphere, maximum overlap among the new spheres and the size of the smallest sphere. Default values for these parameters, 3, 0.8, and 0.5, were used. The amino acid side-chain which is completely embedded inside a protein structure will have buried fraction value equal to one and zero when

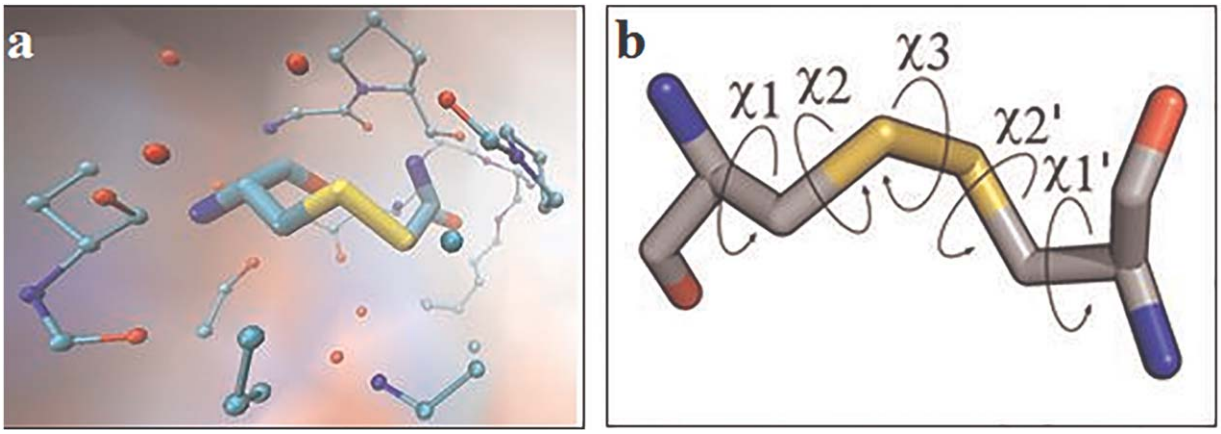


**Figure 1**

(a) Schematic representation of microenvironment around disulfide-bridged cystine residue. The cystine is shown as tube representation. The microenvironment around the cystine is shown in ball and stick representation. Oxygen atoms of water molecules are shown as isolated red balls. The protein background is shown as surface representation. The schema is created using VMD.[18] (**b**) Schematic representation of different dihedral angles observed in disulfide-bridged cystine structure. [Color figure can be viewed at wileyonlinelibrary.com]

completely exposed to the solvent. "rHpy" is a quantitative property descriptor (QPD) of microenvironment around a side-chain or fragment of an amino acid.[9] This QPD estimates the relative hydrophobicity around the amino acid fragment immersed in protein environment with respect to the same in water environment, within a given distance threshold, (that is the radius of the first contact shell around the atom). For example, the microenvironment radius described around a carbon atom is 4.475 Å.[1] From the definition of rHpy, an amino acid side-chain have maximum rHpy value of one when it is completely exposed to water. By definition, there is no absolute minimum rHpy value. Minimum rHpy value varies among proteins and that is decided by the hydrophobicity of individual protein interior. The lowest rHpy value observed in this microenvironment dataset is −0.4.

## Curation of cystine functionalities in protein structures from literature

Functions of 5084 disulfide-bridged cystines were extracted from published literature, PDB header files,[17] Catalytic Site Atlas (CSA)[20] and PDBSUM[38] database. Each sulfur atom from a cystine is defined as a half-cystine through-out this article. The half-cystines were defined functional when those are marked as "catalytic sites" or as "active site" in PDBSUM website.[38] Catalytic sites were defined by the Catalytic Site Atlas (CSA).[20] Active sites were defined by PDBSUM[38] from the SITE records of the PDB header files.[17] The analysis of functional half-cystines has been performed over all the available enzyme classes in the current data set. A total of 617 enzymes containing cystines were analyzed (Supporting Information Table SII).

## Clustering of microenvironment space

Clustering methods are mainly of two types: partitioning and hierarchical.[21] In this study, agglomerative clustering, a subclass of hierarchical clustering has been employed.[22] Agglomerative clustering is a type of hierarchical clustering where each observation is initially considered as a single cluster; based on the distance proximity of the nearest cluster centers, the smaller clusters combine to form large dissimilar clusters. Initially, the microenvironment space was divided into small bins with equal spacing [(BF, rHpy) = (0.1, 0.1)]. Agglomerative clustering method incorporates small microenvironment bins into larger dissimilar clusters where the Euclidean distance of each bin from cluster center is minimal with respect to other cluster centers. Ward's method[23] of agglomerative clustering was employed.[24]

## Identification of cystine residues in different structural classification of protein (SCOP) classes

Protein classes for all the 617 enzymes were attempted to extract from Structural Classification of Proteins (SCOP) database.[25] As latest SCOP database entry dated back to 2009 and our current selection dated, April 2016, SCOP classification was not available from the same database, for some PDB entries. In those cases, structural classification was done based on PDBeFold.[26] The secondary structures were calculated for individual proteins using Kabsch and Sander algorithm[27] in the Dictionary of Protein Secondary Structure (DSSP) calculation.[28]

## Analysis of cystine geometry in hydrophobic and hydrophilic microenvironments

Disulfide-bridged cystine structures are governed by their internal conformations, the five dihedral angles,

**Table I**
Description of Different Microenvironment Clusters Around Half-Cystines, $(S\text{-}S)_{1/2}$

| Cluster | Center[a] | D[b] (Å) | V[c] | Cluster-size | N[d] |
|---|---|---|---|---|---|
| Buried hydrophobic | 0.947, 0.164 | 0.212 | 0.063 | 8549 | 1071 (12.5) |
| Buried hydrophilic | 0.770, 0.482 | 0.186 | 0.043 | 1465 | 960 (66) |
| Exposed hydrophilic | 0.328, 0.720 | 0.145 | 0.031 | 154 | 142 (92) |

Clusters are defined in terms of buried fraction and rHpy using agglomerative hierarchical clustering.[21] Number of other half-cystine (partner in the disulfide) is also reported along with the normalized values in parenthesis. Clusters are arranged, according to descending order of hydrophobicity, measured by cluster center values.
[a]Centroid values of buried fraction, rHpy.
[b]Average distance to center.
[c]Within-class variance.
[d]No. of cystines present in different clusters (Normalized value in %).

namely chi1, chi2, chi3, chi1' and chi2' [Fig. 1(b)]. Based on these dihedral angles, dihedral strain energies[29] were computed for all the 5084 cystines in the current dataset. The program used to compute the disulfide dihedral strain energy (DSE) and is publicly available as a server (http://services.mbi.ucla.edu/disulfide/).

### Sequence and structural alignment of proteins in the dataset

Sequence and structural alignment were performed for two enzyme classes; oxidoreductases and hydrolases. Global sequence alignment was performed using T-COFFEE[30] and local alignment was performed using COBALT.[31] The structural alignment was performed using SALIGN.[32]

## RESULTS AND DISCUSSIONS

### Validation of the updated microenvironment dataset around disulphide ($TD^9$) group of cystine

Microenvironment dataset around disulphide group of cystine amino acid, generated based on the current PDB

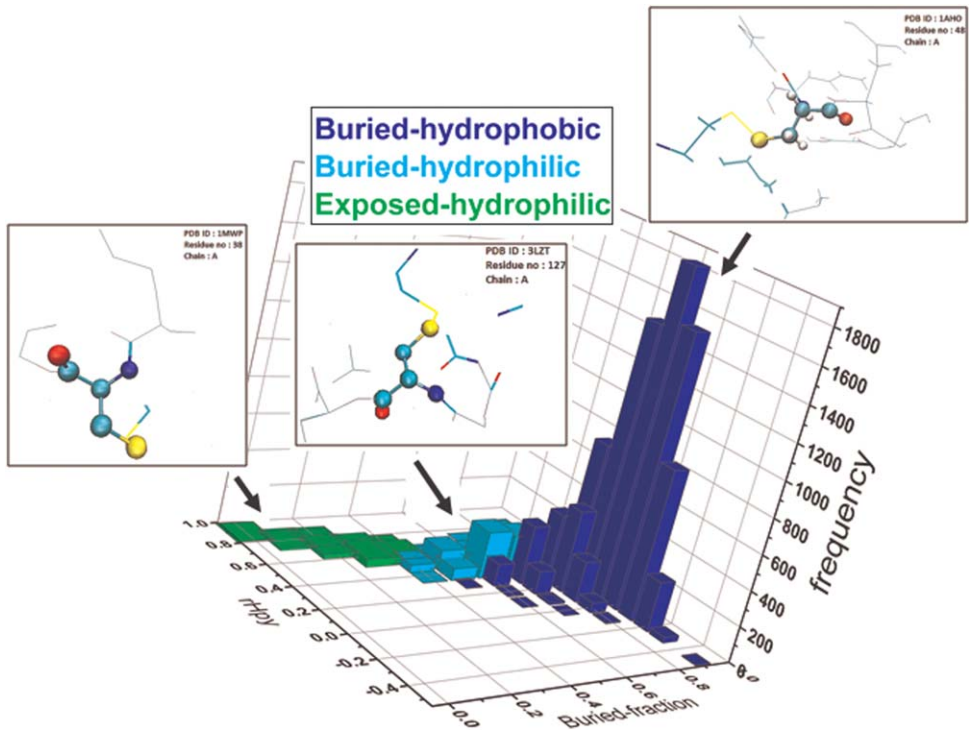


**Figure 2**
Distribution of different microenvironment clusters (obtained from hierarchical clustering) around all the half-cystines from 1478 different proteins in our updated microenvironment dataset. Microenvironments are being clustered in the space of buried fraction and rHpy values. Frequencies in each cluster are shown along z-axis. The plot was generated using Origin [Origin (OriginLab, Northampton, MA)]. Insets show ball and stick representations of representative microenvironments around half-cystines from the 3 different clusters. Thicker sticks highlight bonds within the microenvironment around the sulfur atoms and thin sticks represent the extended microenvironment around the cystine residue.

**Table II**
Distribution of Secondary Structures in Different Microenvironment Clusters, Results Obtained from DSSP Analysis

| | Buried hydrophobic | Buried hydrophilic | Exposed hydrophilic | Total |
|---|---|---|---|---|
| Helix (G+H+I) | 1976 (0.23) | 322 (0.22) | 33 (0.21) | 2331 |
| Beta strand (B+E) | 2968 (0.35) | 268 (0.18) | 32 (0.21) | 3268 |
| Turn (T) | 630 (0.07) | 183 (0.12) | 18 (0.12) | 831 |
| Coil (C) | 2251 (0.26) | 536 (0.37) | 58 (0.38) | 2845 |
| Bend (S) | 724 (0.08) | 156 (0.11) | 13 (0.08) | 893 |

The values given in parenthesis are normalized with respect to the total number of $(S\text{-}S)_{1/2}$ in a particular cluster.

entries, was validated with respect to our previously observed dataset.[9] Buried fraction and rHpy, the two parameters used throughout this work to describe microenvironment, were compared across the previous and the current datasets. Mean and standard deviation values (given within the parenthesis) for buried fraction and rHpy, in the current dataset are 0.91 (0.13) and 0.22 (0.18), respectively. These are in close agreement to the values from the previous dataset (0.92 (0.12) and 0.21 (0.16) respectively; Table III in reference 1 for TD group). Agreement of the microenvironment descriptors in current and previous datasets indicates that despite of the dataset size, overall nature of the protein microenvironments around cystines remain the same. In addition to the mean values of these property descriptors, percentage of cystines microenvironment outliers, (outliers are defined as those outside of one standard deviation with respect to the mean value) are also compared across previous and current datasets. Total number of outliers in the current dataset is 906; 8.9% of the total dataset; this value is fairly in agreement with that in the previous dataset (9.3%; Table I in reference 1).

## Microenvironment clusters around cystine residues distributed in high-resolution crystal structures with 25–30% sequence similarity

Microenvironment clusters around half-cystines -$(S\text{-}S)_{1/2}$- were obtained using agglomerative clustering, resulting into three categories (i) buried-hydrophobic, (ii) buried-hydrophilic, and (iii) exposed-hydrophilic (Table I). These microenvironment clusters gradually progress from protein interior to surface (decrease in BF) with increasing hydrophilic character (increase in rHpy) (Fig. 2).

Cystine residues in buried-hydrophobic cluster are observed inside the protein interior, (average buried fraction, 0.95) (Supporting Information Fig. S1). The buried-hydrophobic cluster mainly contains half-cystines which are buried deep inside the protein core and embedded in hydrophobic microenvironment (average rHpy 0.16) (Fig. 2 and Supporting Information Fig. S1). Conversely, half-cystines in the buried-hydrophilic cluster are buried inside the protein (average buried fraction, 0.77) which constitutes relatively hydrophilic protein microenvironment (average rHpy, 0.48). Half-cystines in the exposed-hydrophilic cluster are observed on the outer surface of proteins (average buried fraction of 0.33) and have relatively high hydrophilic microenvironment (average rHpy 0.72). Interestingly, cluster sizes (number of half-cystines in each cluster) gradually decreases with hydrophilicity in the clusters (Table I), indicating that cystine residues tend to be in buried hydrophobic microenvironments.[9]

Both half-cystines derived from the same disulfide bonded cystine not necessarily fall into the same microenvironment cluster. The second sulfur atom from the disulfide bond tend to stay in the same microenvironment cluster, provided the later is hydrophobic (Table I).

## Correlation between secondary, super-secondary structures and microenvironment clusters around half-cystines

The correlation between secondary structures of half-cystines and respective protein folds was known.[33] Here we aim to find out whether half-cystines with similar secondary structures (e.g., alpha helix, beta sheet, or coil) populate into similar microenvironment clusters. Microenvironment can be viewed as the local tertiary structure around a particular amino acid which might include different components of secondary structures. Therefore, there is a possibility of underlying relationship between

**Table III**
Description of Different Microenvironment Clusters Populated with -$(S\text{-}S)_{1/2}$ - from Different Enzyme Classes

| Cluster | E | Oxidoreductases[a] | Hydrolase | Lyase | Transferase | Isomerase |
|---|---|---|---|---|---|---|
| Buried hydrophobic | 3571[b] | 571 (0.16) | 2471 (0.69) | 97 (0.03) | 355 (0.1) | 59 (0.02) |
| Buried hydrophilic | 717[c] | 104 (0.15) | 520 (0.73) | 29 (0.04) | 54 (0.08) | 4 (0.01) |
| Exposed hydrophilic | 31 | 0 (0.0) | 28 (0.90) | 0 (0.0) | 2 (0.06) | 1 (0.03) |

The total numbers of -$(S\text{-}S)_{1/2}$ -, present in an enzyme class in each cluster (E), are reported. The number of -$(S\text{-}S)_{1/2}$ - in individual enzyme classes are also reported, along with their normalized values, given in parenthesis. The values are normalized with respect to the total number of half-cystines present in a particular microenvironment cluster as part of an enzymes class (E).
[a]And Electron transport proteins.
[b]18 cystines were observed as part of ligase.
[c]6 cystines were observed as a part of ligase.

**Table IV**
Number of -(S-S)$_{1/2}$ - with Different Main Chain Conformations Present in Enzyme Classes—Hydrolases, Oxidoreductases and Transferases

| Enzyme class | Total number of cysteines | Alpha helix | Beta sheet | Loop |
|---|---|---|---|---|
| Hydrolases | 3019 | 656 | 869 | 1495 |
| Oxidoreductases and675 Electron transport proteins | 209 | | 91 | 375 |
| Transferases | 411 | 77 | 124 | 210 |

super-secondary structures (protein classes) and amino acid microenvironment, particularly, around disulphide-bridged cystines which dictates protein scaffolds. It has been shown earlier that similar disulphide bonding patterns lead to similar fold, families and super-families of proteins, despite of their low sequence identity.[34] To understand the main-chain conformations of half-cystines in different microenvironments, DSSP secondary structure analyses were performed on our protein dataset. Preferences of different secondary structures around half-cystines in three different microenvironment clusters were observed and are described below (Table II).

The buried-hydrophobic cluster possesses higher percentage of beta secondary structures (and to lesser extent alpha helical structures). The buried-hydrophilic cluster predominates in coil conformations. The exposed-hydrophilic cluster mainly possesses cystines in coil conformations. These results have suggested that the half-cystines buried in hydrophobic microenvironment are mainly comprised of alpha and beta secondary structures, whereas, partly exposed half-cystines embedded in hydrophilic microenvironment, predominate in flexible conformations, like coils. Here we have attempted to identify microenvironment specific functions of cystines observed with different secondary structures. As the functions of cystines are mainly associated with enzymes, in the following section we have analyzed cystine functions for 617 enzymes from all the enzyme classes[14] present in our current dataset.

## Buried-hydrophobic microenvironment cluster hosts half-cystines from redox-active -C-x-x-C- motif

The half-cystines present in buried-hydrophobic microenvironment cluster, mainly come from (i) oxidoreductase and electron transport proteins, (ii) hydrolase, (iii) lyase, (iv) transferase, and (v) isomerase enzyme classes (Table III). Hydrolase is the most frequently

**Table V**
Buried Hydrophobic Microenvironment Cluster: Secondary Structures and Strain Energies of Redox-Active Cystines (Part of C-x-x-C motif), Present in Different Oxidoreductases Enzymes and Electron Transport Proteins

| PDB | Fold SCOP or PDBeFOLD[a] | Cystine Numbers (Chain) | Half-cystine in helix (Chain) [cluster no.] | Half-cystine in beta turn[b] (Chain) [cluster no.] | Strain Energy (KJ/mol) | BF | rHpy |
|---|---|---|---|---|---|---|---|
| 3ZIT | Thioredoxin[a] | 12 (A)−15(A) | 15 (A) [2] | 12 (A) [1] | 11.4 | 0.732 | 0.375 |
| 3GWN | a[c] | 80(A)−83(A) | 83 (A) [2] | 80 (A) [2][d] | 19.5 | 0.751 | 0.275 |
| 3GWN | a[c] | 80(B)−83(B) | 83 (B) [2] | 80 (B) [2][d] | 21.5 | 0.800 | 0.228 |
| 4HS1 | Thioredoxin | 11(A)−14(A) | 14 (A) [2] | 11 (A) [2] | 13.0 | 0.804 | 0.194 |
| 3P0K | b[e] | 155 (A)−158(A) | 158 (A) [2] | 155 (A) [2] | 8.23 | 0.805 | 0.162 |
| 3ZIT | Thioredoxin[a] | 12 (B)−15(B) | 15 (B) [2] | 12 (B) [1] | 11.6 | 0.871 | 0.33 |
| 3FZ4 | Thioredoxin[a] | 10 (A)−13(A) | 13(A) [2] | 10 (A) [2] | 10.5 | 0.926 | 0.095 |
| 1THX | Thioredoxin | 32(A)−35(A) | 35(A) [2] | 32 (A) [1] | 15.4 | 0.953 | 0.106 |
| 1JR8 | a[c] | 54(A)−57(A) | 57 (A) [1] | 54 (A) [2][d] | 19.1 | 0.960 | 0.045 |
| 1H75 | Thioredoxin | 11(A)−14(A) | 14 (A) [2] | 11 (A) [2] | 13.5 | 0.967 | 0.192 |
| 2I4A | Thioredoxin[a] | 32(A)−35(A) | 35 (A) [2] | 32 (A) [2] | 11.4 | 0.978 | 0.029 |
| 2B1L | Thioredoxin | 80(B)−83(B) | 83 (B) [2] | 80 (B) [1] | 25.8 | 0.984 | 0.149 |
| 2B1L | Thioredoxin | 80(A)−83(A) | 83 (A) [2] | 80 (A) [1] | 25.0 | 0.986 | 0.142 |
| 1FVK | Thioredoxin | 30(B)−33(B), | 33(B) [2] | 30 (A) [1] | 14.3 | 0.988 | 0.065 |
| 1FVK | Thioredoxin | 30(A)−33(A), | 33(A) [2] | 30 (A) [1] | 19.2 | 0.993 | −0.021 |
| 1KNG | Thioredoxin | 92(A)−95(A) | 95 (A) [2] | 92 (A) [2] | 19.3 | 0.996 | −0.017 |
| 1ABA | Thioredoxin | 14(A)−17(A) | 17 (A) [2] | 14 (A) [1] | 12.4 | 0.997 | 0.130 |
| 2HLS | Thioredoxin[a] | 150(B)−153(B) | 153 (B) [2] | 150 (B) [1] | 10.3 | 0.998 | 0.078 |
| 1ST9 | Thioredoxin | 73(A)−76(A) | 76 (A) [2] | 73 (A) [2] | 14.8 | 0.999 | 0.077 |
| 1ST9 | Thioredoxin | 73(B)−76(B) | 76 (B) [2] | 73 (A) [2] | 13.0 | 0.999 | 0.119 |
| 2HLS | Thioredoxin[a] | 150(A)−153(A) | 153 (A) [2] | 150 (A) [1] | 12.2 | 1.000 | 0.117 |
| 3IOS | Thioredoxin[a] | 81 (A)−84(A) | 84 (A) [2] | 81 (A) [2] | -[f] | 1,.000 | 0.015 |

Residue numbers of half-cystines corresponding to each cystine are shown (column 3). Respective cluster numbers are given in parenthesis. Buried hydrophobic cluster is referred to as 2 and buried hydrophilic cluster is referred to as 1 here. Folds of individual residues are mentioned either from SCOP[25] and PDBeFOLD.[26]
[a]Obtained from PDBeFold.[26]
[b]Unless otherwise mentioned; according to DSSP program.[27,28]
[c]Represents Four-bundle-up-and-down bundle fold.
[d]Half-cystine present in Gamma turn instead of beta turn.
[ee]Represents spectrin repeat like fold.
[f]Strain energy cannot be calculated by the "*Disulfide Bond Dihedral Angle Energy Server.*"[35]

**Table VI**
Half-Cystines (First Out of the Pair of Half-cystines from a Disulphide Shown in the Table) Present in Different Enzyme Active Site or its Embedded Microenvironment, Reported from Buried-Hydrophilic Cluster Protruding Toward Protein Surface (Buried-Fraction $< 0.78$ and rHpy $> 0.40$)

| PDB ID | cystine (chain) | 2° structures | Sequence conservations[a] | Strain energy (KJ/mol) | Active residues in 4.5 Å region in half-cystine | BF | rHpy |
|---|---|---|---|---|---|---|---|
| Cystines in active site in hydrolases | | | | | | | |
| 3EDH | 65(A)−64(A) | Coil-bend | High-high | 13 | C64 | 0.438 | 0.602 |
| 1K7C | 232(A)−214(A) | COIL-helix[b] | Low-high | 6.3 | C232 | 0.471 | 0.575 |
| 2FHF | 643(A)−644(A) | bend-bend | High-medium | 9.8 | C643,T642,H607 | 0.512 | 0.547 |
| 3B8Z | 376(A)−371(A) | Turn-beta[c] | Medium-low | 13.4 | C376,L370 | 0.528 | 0.567 |
| 4XOJ | 124(A)−225(A) | Coil-helix | Medium-medium | 10.9 | C124,A125 | 0.531 | 0.538 |
| 1QNR | 284(A)−334(A) | Helix-helix | Medium-low | 5.5 | C284 | 0.534 | 0.512 |
| 3B8Z | 376(B)−371(B) | Turn-beta | Medium-low | 13.2 | C376,L370 | 0.545 | 0.553 |
| 3EQA | 473(A)−246(A) | Coil-helix | High-low | 15.4 | C473 | 0.572 | 0.534 |
| 3HHI | 215(B)−154(B) | Coil-helix | High-high | 6.3 | C154 | 0.586 | 0.521 |
| 3KUV | 73(A)−73(B) | Beta-beta | Low-low | 14.5 | C73,v74 | 0.596 | 0.547 |
| 3EDH | 64(A)−65(A) | Bend-coil | High-high | 13 | C64 | 0.598 | 0.415 |
| 3TBJ | 45(A)−28(A) | Turn-coil | High-high | 8.9 | C45 | 0.602 | 0.546 |
| 4HWX | 31(A)−46(A) | Beta-helix | High-high | 7.6 | C31,A32 | 0.605 | 0.474 |
| 1LNI | 96(B)−7(B) | Coil-beta | High-low | 10.7 | C96 | 0.622 | 0.505 |
| 2WBF | 636(X)−593(X) | Turn-helix | High-high | 5.3 | C636,R635 | 0.638 | 0.477 |
| 1LNI | 7(A)−96(A) | Beta-coil | Low-high | 12 | C7 | 0.639 | 0.515 |
| 3ARX | 121(A)−116(A) | Beta-coil | High-high | 15.5 | C121 | 0.645 | 0.424 |
| 1G6X | 14(A)−38(A) | Coil-bend | High-high | 12.7 | C14 | 0.649 | 0.42 |
| 3WQB | 301(A)−326(A) | Turn-coil | High-high | 8.7 | C326 | 0.678 | 0.526 |
| 3TBJ | 44(A)−149(A) | Turn-helix | High-high | 4 | C45 | 0.71 | 0.44 |
| 4DO4 | 127(B)−158(B) | Turn-turn | High-high | 14.8 | C127,D80 | 0.718 | 0.476 |
| 4CPY | 336(B)−317(B) | Turn-coil | High-high | 12.3 | C336 | 0.721 | 0.406 |
| 4DO4 | 127(A)−158(A) | Turn-turn | High-high | 14.5 | C127,D79 | 0.724 | 0.47 |
| 3A21 | 604(B)−585(B) | Coil-beta | Medium-low | 14.7 | C604,E577 | 0.74 | 0.468 |
| 3ZFP | 69(A)−168(A) | Turn-coil | High-high | 15.1 | C69 | 0.74 | 0.509 |
| 3A21 | 604(A)−585(A) | Coil-beta | Medium-low | 14.9 | C604,E577 | 0.749 | 0.46 |
| 4Y5L | 255(A)−290(A) | Turn-beta | High-high | 14 | C255 | 0.759 | 0.406 |
| Cystines in the active site of Oxidoreductases | | | | | | | |
| 1ZK7 | 465(A)−464(A) | Bend-bend | Medium-medium | 23.4 | C465,A466,C464 | 0.484 | 0.534 |
| 1ZK7 | 464(A)−465(A) | Turn-coil | Medium-medium | 23.4 | C464, | 0.503 | 0.566 |
| 4OZ7 | 4(B)−10(B) | Coil-coil | NA[d]-NA[d] | 16.4 | C4,C10,S5 | 0.528 | 0.568 |
| 4OZ7 | 10(B)−4(B) | Coil-coil | NA[d]-NAd | 16.4 | C10,C4,S5,P8 | 0.583 | 0.543 |
| 5DQY | 62(A)−69(A) | Bend-coil | NA[d]-NA[d] | 9.3 | C62,Q63,S67 | 0.586 | 0.49 |
| 4OZ7 | 10(A)−4(A) | Coil-coil | NA[d]-NA[d] | 17.3 | C10 | 0.595 | 0.534 |
| 4OZ7 | 4(A)−10(A) | Coil-coil | NA[d]-NA[d] | 17.3 | C4 | 0.597 | 0.476 |
| 4NTC | 148(A)−145(A) | Turn-coil | NA[d]-NA[d] | 14.8 | C148,H144 | 0.643 | 0.435 |
| 2Q0L | 136(A)−133(A) | Helix-coil | High-medium | -[e] | C136,Q289 | 0.649 | 0.51 |
| 2Q0L | 133(A)−136(A) | Coil-helix | Medium-high | -[e] | C133,T132,C136, K288,Q289 | 0.658 | 0.537 |
| Cystines in the active site of transferases | | | | | | | |
| 4WMA | 385(A)−356(A) | Turn-coil | NA[d]-NA[d] | 20.4 | C385,W358 | 0.587 | 0.473 |
| 2APC | 115(A)−145(A) | Bend-turn | Medium-medium | 8.6 | C115,A114 | 0.622 | 0.419 |
| Cystines in the microenvironment of active site residues of hydrolases | | | | | | | |
| 2VB1 | 6(A)−127(A) | Helix-turn | High-high | 5.1 | E7,I124 | 0.367 | 0.719 |
| 3RLG | 53(A)−201(A) | Turn-turn | High-high | 14.4 | N200 | 0.447 | 0.652 |
| 3LUM | 269(C)−297(C) | Coil-helix | High-high | 14.1 | P249 | 0.557 | 0.476 |
| 3KUV | 73(B)−73(A) | Beta-beta | Low-low | 14.5 | C73,V74 | 0.567 | 0.539 |
| 3C1U | 96(A)−88(A) | Beta-helix | Medium-high | 9.8 | Y97 | 0.588 | 0.444 |
| 3NKQ | 801(A)−413(A) | Turn-bend | High-high | 3.6 | S800 | 0.617 | 0.462 |
| 1KNM | 119(A)−100(A) | COIL-Beta | High-high | 13.2 | Y117 | 0.62 | 0.489 |
| 1EB6 | 117(A)−177(A) | Turn-Coil | Medium-high | 30 | H118 | 0.624 | 0.409 |
| 1G66 | 2(A)−79(A) | Coil-Bend | High-high | 6 | S1,S74 | 0.631 | 0.554 |
| 2NLR | 69(A)−64(A) | Beta- Beta | Medium-medium | 19.8 | H69 | 0.635 | 0.467 |
| 2D1Z | 425(A)−406(A) | Coil-beta | Medium-medium | 13.2 | Y423 | 0.645 | 0.437 |
| 3TRS | 101(D)−18(D) | Coil-bend | High-high | 13.9 | G15 | 0.646 | 0.458 |
| 2WJ9 | 13(B)−125(A) | Bend-coil | Low-medium | 7.3 | C129 | 0.647 | 0.524 |
| 3EQN | 77(A)−73(A) | Coil-bend | Medium-high | 13.4 | D78 | 0.65 | 0.491 |
| 2XXL | 197(A)−188(A) | Beta-beta | High-low | 18.2 | D187 | 0.671 | 0.476 |
| 3LUM | 250(A)−277(A) | Turn-turn | High-high | 2.9 | P249 | 0.673 | 0.439 |
| 3LUM | 250(C)−277(C) | Turn-turn | High-high | 3 | P249 | 0.675 | 0.437 |
| 3LUM | 250(B)−277(B) | Turn-turn | High-high | 3.2 | P249 | 0.676 | 0.437 |

**Table VI**
(Continued)

| PDB ID | cystine (chain) | 2° structures | Sequence conservations[a] | Strain energy (KJ/mol) | Active residues in 4.5 Å region in half-cystine | BF | rHpy |
|---|---|---|---|---|---|---|---|
| 2WJ9 | 13(A)−125(B) | Coil-coil | Low-medium | 7.4 | C129 | 0.677 | 0.456 |
| 3LUM | 250(D)−277(D) | Turn-turn | High-high | 3.7 | P249 | 0.679 | 0.434 |
| 3LZT | 6(A)−127(A) | Helix- coil | High-high | 5.78 | R128 | 0.68 | 0.441 |
| 4M1U | 56(F)−3(F) | Coil-beta | Medium-medium | 7.6 | S54 | 0.68 | 0.526 |
| 2XXL | 197(B)−188(B) | Beta-bend | High-low | 18.2 | D187 | 0.69 | 0.499 |
| 1LBU | 81(A)−3(A) | Bend-bend | Medium-medium | 11.1 | D80 | 0.691 | 0.43 |
| 3A21 | 647(B)−628(B) | Coil-beta | Low-medium | 17.1 | G626,W645 | 0.693 | 0.432 |
| 4HYQ | 53(A)−28(A) | Coil-helix | High-high | 10.8 | S54 | 0.693 | 0.467 |
| 1I71 | 1(A)−78(A) | Beta-Coil | High-high | 7.5 | P79 | 0.699 | 0.44 |
| 4H04 | 589(B)−564(B) | Coil-beta | High-high | 11.4 | N588 | 0.703 | 0.557 |
| 4H04 | 589(A)−564(A) | Coil-beta | High-high | 11.5 | N588 | 0.705 | 0.489 |
| 3TRS | 101(B)−18(B) | Coil-coil | High-high | 15.6 | G15 | 0.71 | 0.455 |
| 4EMN | 291(C)−355(C) | Helix-helix | High-high | 8.1 | E292,C355 | 0.711 | 0.41 |
| 2IFR | 148(A)−141(A) | Beta-beta | Low-medium | 13.9 | E139 | 0.72 | 0.402 |
| 1Y43 | 101(B)−18(B) | Coil-coil | High-high | 11.8 | E19 | 0.735 | 0.436 |
| 2D1Z | 842(B)−823(B) | Coil-beta | High-High | 12.3 | H843,Y840 | 0.737 | 0.436 |
| 2FHF | 644(A)−643(A) | Bend-beta | Medium-high | 9.8 | T642,H607 | 0.74 | 0.401 |
| 3B8Z | 371(B)−376(B) | Beta-beta | Low-medium | 13.2 | L370,C376 | 0.749 | 0.487 |
| 3A21 | 562(A)−543(A) | Coil-beta | Medium-low | 13.3 | N563,W560 | 0.752 | 0.464 |
| 2I9A | 50(D)−131(D) | Bend-bend | High-high | 8.1 | H129 | 0.753 | 0.436 |
| 2XU3 | 65(A)−22(A) | Turn-helix | High-high | 7.7 | G23 | 0.757 | 0.403 |
| 4HYQ | 28(A)−53(A) | Helix-coil | High-high | 10.8 | S54 | 0.766 | 0.421 |
| 3B8Z | 371(A)−376(A) | Beta-turn | Low-medium | 13.4 | L370,C376 | 0.785 | 0.4 |
| Cystine in microenvironment of active site residues in Transferases | | | | | | | |
| 4WTP | 264(A)−218(A) | Coil-helix | Medium-high | 12.7 | K221,A262 | 0.642 | 0.402 |
| 4EBY | 93(A)−25(A) | Beta-coil | High-high | 8.3 | F98 | 0.64 | 0.461 |
| Cystine in microenvironment of active site residues in Lyases | | | | | | | |
| 4Q8K | 365(A)−377(A) | Coil-helix | High-high | 9.1 | K364 | 0.636 | 0.47 |
| 3KBR | 131(A)−210(A) | Helix-beta | High-medium | 7.7 | H213,P214,N215 | 0.642 | 0.542 |
| Cystine in microenvironment of active site residues in Electron transport proteins | | | | | | | |
| 2Z8Q | 48(A)−21(A) | Helix-high | Low-coil | 11.3 | L20 | 0.601 | 0.543 |
| Cystine in microenvironment of active site residues in Isomerases | | | | | | | |
| 3O22 | 89(A)−186(A) | Beta-bend | High-high | 6.1 | F83 | 0.257 | 0.75 |

Secondary structures and sequence conservation of each half-cystines were reported along with the strain energy of the cystine disulphide bonds. PDB IDs are arranged according to the increasing values of buried fraction (second last column). Amino acids and their positions involved in enzyme active sites are mentioned in third last column.

[a]Sequence conservation obtained from PdbSum.[38]

[b]Includes 3(10)-helix, alpha-helix and π-helix secondary structures calculated by DSSP.[27,28]

[c]Includes beta sheet, beta bridge and beta strand secondary structures calculated by DSSP program.[27,28]

[d]Sequence conservation was not reported by "*The ConSurf Server.*"[36]

[e]Strain energy cannot be calculated by the "*Disulfide Bond Dihedral Angle Energy Server.*"[35]

observed enzyme class in this cluster, as this enzyme class constitutes the maximum population size. Conversely, oxidoreductase and electron transport proteins were exclusively present in buried clusters, either hydrophobic or hydrophilic (0 half-cystines were observed in exposed-hydrophilic cluster from oxidoreductase and electron-transport proteins). In the following part of this section, we have attempted to answer the question why oxidoreductase and electron transport proteins are selective toward buried clusters only. Out of the 675 half-cystines in oxidoreductase and electron-transport proteins, 209 belong to alpha helical secondary structure (Table IV). Twenty two cystines in alpha helical conformations were part of redox-active C-x-x-C- motif (Table V). 18 out of 22 cystines, belongs to thioredoxin fold. All these 22 cystines have higher strain energies compared to remaining disulphides in buried-hydrophobic cluster (average value of strain energy in buried-hydrophobic cluster is 11.13 KJ/mol). The above redox active half-cystines with alpha helical conformations in thioredoxin fold span a small region within buried-hydrophobic cluster, with few exceptions (Table VI). One prominent exception is, Cys15 (chain A) of a thioredoxin like mutant protein (PDB ID:3ZIT) that significantly differ from other thioredoxin proteins due to the presence of a Thr53 residue (not conserved) within the microenvironment of -C-x-x-C- motif.[37] The second Cys from –C-x-x-C- motif belongs to beta turn secondary structure for all the above thioredoxin folds (as obtained from PDBSUM[38]) (Table V). The active half-cystines, part of four-helical bundle folds, possess the second Cys in gamma turn secondary structure. It has been proposed earlier that redox active

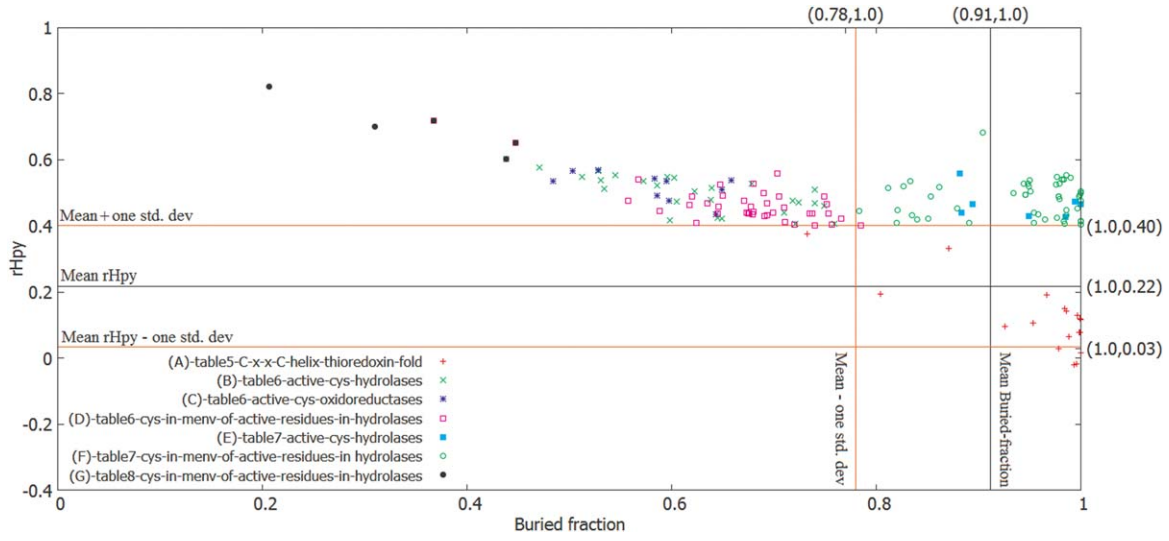


**Figure 3**

Microenvironments of half-cystines depicted in buried fraction and rHpy space–(a) (red diamond) as a part of -C-x-x-C- motif in thioredoxin fold with alpha helical geometry (values given in Table V); (b) (green cross) as a part of the active sites in hydrolase enzyme class present in buried hydrophilic cluster protruding toward protein surface (values given in Table VI); (c) (blue star) as a part of the active sites in oxidoreductase enzyme class present in buried hydrophilic cluster protruding toward protein surface (values given in Table VI); (d) (magenta square) as a part of the microenvironment around the active sites in hydrolase enzyme class present in buried hydrophilic cluster protruding toward protein surface (values given in Table VI); (e) (cyan filled square) as a part of the active sites in hydrolase enzyme class present in buried hydrophilic cluster protruding toward protein interior (values given in Table VII); (f) (green circle) as a part of microenvironment surrounding the active sites in hydrolase enzyme class present in buried hydrophilic cluster protruding toward protein interior (values given in Table VII); (g) (black filled circle) as a part of the active sites in hydrolase enzyme class present in exposed hydrophilic cluster (values given in Table VIII). Mean values of buried fraction and rHpy are shown by black lines. One standard deviation values with respect to buried faction and rHpy are shown in orange. [Color figure can be viewed at wileyonlinelibrary.com]

cystines in oxidoreductases prefer alpha helical conformation due to their easy cleavage.[39] To analyze the residues involved in -C-x-x-C- motif of thioredoxin fold, sequence and structural alignment of these proteins (listed in Table V) were performed. Global sequence alignment (using T-COFFEE[30]) revealed 37% of sequence identity among all these 13 sequences. The local sequence alignment shows the conservation of -C-x-x-C motif in all these proteins, with an exception of one protein, PDB ID: 1FVK (Supporting Information Fig. S2). However, structural alignment of the above full proteins resulted into C-alpha RMSD value of 4.4 Å in the structurally aligned regions (as obtained from SALIGN[32]) It was reported earlier that the redox active cystines, part of -C-x-x-C- motif in thioredoxin folds, were mainly observed in beta-alpha-beta fold region.[15] Superimposition of beta-alpha-beta regions from above 13 proteins shows Cα RMSD of 1.99 Å (Fig. 3). The Cα RMSD of the superimposed cystine residues in those proteins is 1.4 Å. The Cα RMSD of the microenvironment around the cystines in -C-x-x-C- motif of thioredoxin fold is 1.7 Å. The novel finding in this work is that these half-cystines at the active site share near-identical microenvironment as a consequence of having a common fold (Fig. 4).

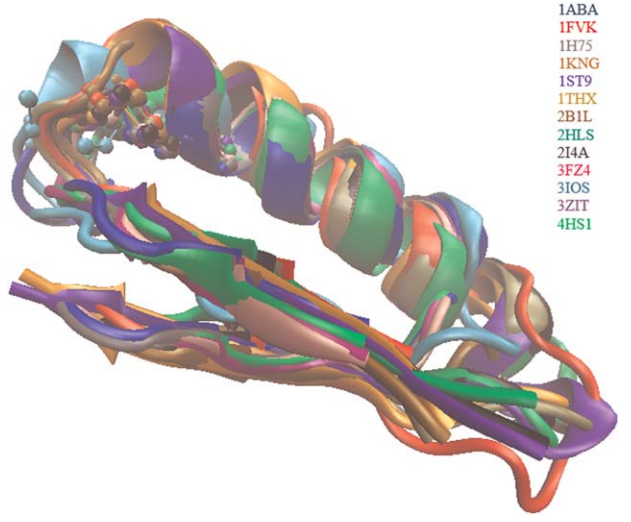


**Figure 4**

Superimposed beta-alpha-beta region from 13 proteins with thioredoxin fold. Half-cystine from different proteins in this aligned regions are depicted by different colors. Disulphide bridges are shown by spheres connected by sticks. Beta-alpha-beta regions are represented through cartoons. [Color figure can be viewed at wileyonlinelibrary.com]

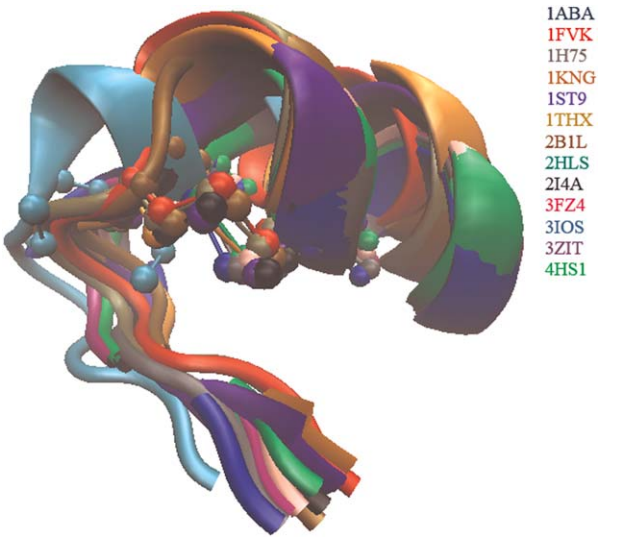


**Figure 5**

Conservation of microenvironment around disulphide bridge (-S-S-) in -C-x-x-C- motif from 13 proteins in alpha-helical conformations of thioredoxin folds. Microenvironment in each case have alpha-helical conformation on one side and beta turn on the other side. -C-x-x-C motif in all these proteins are encompassed within the cleft created by alpha helix and beta turn region. Disulphide bond is represented as spheres connected by sticks. The cartoon diagram represents alpha helix and beta turn. [Color figure can be viewed at wileyonlinelibrary.com]

### Buried-hydrophilic cluster hosts half-cystines from enzyme active sites or from surrounding microenvironments of the catalytic (or active) sites

Buried-hydrophilic microenvironment cluster contradicts with the intrinsic hydrophobic character of disulphide-bridged cystine residues. The average buried fraction and rHpy values of cystines in the total dataset (total 10,168 half-cystines) are 0.91 and 0.22 respectively. Respective one standard deviation values for buried fraction and rHpy are 0.13 and 0.18 (Fig. 5). Cystines outside one standard deviation values of buried fraction and rHpy were defined to have mismatched microenvironment.[9] Cystines from buried hydrophilic cluster, encountering mismatched microenvironment (total 963 half-cystines), are mostly exposed toward the surface of the protein (buried fraction $<0.78$ and rHpy $>0.40$). Out of these, 445 belongs to enzymes; mostly from hydrolase enzyme class (332 half-cystines, constituting 74% of the enzyme class in buried hydrophilic cluster). Total 39 half-cystines were detected as part of the enzyme active sites and 47 in the microenvironments around enzyme active sites (Table VI). Two-third of the active half-cystines in hydrolases have high sequence conservation according to ConSurf Server.[40] There were only three active half-cystines in hydrolases with low sequence conservation. No active half-cystines in oxidoreductases exhibited low sequence conservation (Table VI). In terms of the secondary structures of active half-cystines, those from oxidoreductases mainly adopted coil or turn conformations and exposed toward the protein surface (Fig. 5). Hydrolases, being the largest class of enzyme in buried-hydrophilic cluster, secondary structures of half-cystine pairs (from a disulphide) have wider variations in comparison to those in oxidoreductases. However, active half-cystines from hydrolases cluster together into smaller sub-groups (Fig. 5). The one toward the protein interior is more densely packed. All the half-cystines from this sub-group share similar secondary structure with their counter-parts in the disulphide (Table VI). Remaining half-cystines from hydrolases are more scattered (toward exposed region) in buried-fraction—rHpy space (Fig. 5), that can presumably be attributed to two different secondary structures of half-cystines in a disulphide (Table VI).

Apart from the half-cystines protruding toward protein surface (BF $<0.78$ and rHpy $>0.4$) in buried hydrophilic cluster, we have also explored the functionalities of half-cystines embedded in hydrophilic microenvironment protruding toward protein interior (buried fraction $>0.78$ and rHpy $>0.40$). These cystines also exhibited catalytic activities but to lesser extent compared to the cystines those discussed in the above paragraph. Total 652 such half-cystines were obtained. Out these 337 belong to enzyme classes. Only seven half-cystine were part of active site and 63 were reported as part of the microenvironment around the active sites (Table VII). One compact sub-group of half-cystines from microenvironment of the active site was observed (buried fraction $\sim 1$) (Fig. 5). Within this sub-group, almost all the half-cystines share the same secondary structures with its counterpart in the disulphide (Table VII). This observation is similar to our observation in the previous paragraph for densely packed sub-group (Table VI and Fig. 5). This analysis has shown that the cystines in the microenvironment of catalytic residues, on average, have high strain energies compared to the average strain energy in this microenvironment cluster (11.0 KJ/mol). Despite of low or medium sequence conservation in some of the half-cystines, all these half-cystines share very similar microenvironment (Fig. 5). One of the reasons for low sequence identity of these cystine residues is their occurrences in irregular secondary structures (coil, turn etc.) in contrast to cystines in -C-x-x-C- motif of thioredoxin fold (fully conserved). It has been shown in Table V that the cystines in -C-x-x-C- motifs from oxidoreductase or electron transport proteins mainly occupy alpha-helical structure to facilitate easy cleavage of the disulfide bond, with high strain energy values. These observations together suggest that cystines within the microenvironment of catalytic residues of hydrolase enzyme do not require easy cleavage of the disulfide bond as in thioredoxin folds.

**Table VII**
Half-Cystines (First Out of the Pair of Half-Cystines From a Disulphide Shown in the Table) Present in Different Enzyme Active Site or its Embedded Microenvironment, Reported from Buried-Hydrophilic Cluster Protruding Toward Protein Interior (Buried-Fraction > 0.78 and rHpy > 0.40)

| PDB ID | Cystine (chain) | 2° structures | Sequence conservations[a] | Strain energy (KJ/mol) | Active residues in 4.5 Å region in half-cystine | BF | rHpy |
|---|---|---|---|---|---|---|---|
| Active cystines in hydrolases | | | | | | | |
| 4ZA3 | 206(A)−189(A) | Coil-beta[b] | High-high | 8.4 | C206 | 0.882 | 0.559 |
| 2QTW | 358(B)−323(B) | Turn-helix[c] | High-high | 11.4 | C358 | 0.883 | 0.439 |
| 2JJB | 539(C)−533(C) | coil-bend | Low-medium | 19.4 C539, P534, A544, C541 | 0.894 | 0.465 | |
| 1G66 | 52(A)−46(A) | Helix-coil | Low-medium | 17.1 T13, G47, Q49, S50, Q91 | 0.949 | 0.43 | |
| 1G66 | 179(A)−147 | Bend-coil | Low-low | 13.6 D175, T146, D172, A173, | 0.985 | 0.427 | |
| 1G66 | 46(A)−52(A) | Coil-Helix | Medium-low | 17.1 T13, G47, Q49, S50, Q91, S174, Y177 | 1 | 0.464 | |
| 1GPI | 225(A)−245(A) | Beta-Beta | High-high | 16.3 | G207 | 0.994 | 0.472 |
| Cystines in microenvironment of active site residues in Hydrolases | | | | | | | |
| 2D1Z | 342(A)−323(A) | Coil-beta | Medium-low | 11.4 | Y340, H343 | 0.783 | 0.445 |
| 4YEO | 127(A)−6(A) | Coil-helix | High-high | 3.8 | R125 | 0.812 | 0.514 |
| 3WMT | 297(B)−308(B) | Coil-coil | High-high | 8.9 | L296 | 0.82 | 0.408 |
| 3E2V | 303(B)−331(B) | Helix-coil | High-high | 13 | E332 | 0.821 | 0.447 |
| 3E2V | 303(A)−331(A) | Helix-coil | High-high | 12 | E332 | 0.827 | 0.518 |
| 1XKG | 183(A)−145(A) | Coil-helix | High-high | 8.5 | R184 | 0.833 | 0.536 |
| 3WMT | 297(A)−308(A) | Coil-bend | High-high | 8.2 | L296 | 0.835 | 0.432 |
| 2X5X | 36(A)−85(A) | Bend-turn | Medium-medium | 12 | S35 | 0.84 | 0.419 |
| 3O0D | 43(A)−47(A) | Coil-helix | High-high | 16.8 | N42 | 0.851 | 0.421 |
| 2QTW | 323(B)−358(B) | Helix-turn | High-high | 11.4 | C358 | 0.853 | 0.489 |
| 2XXL | 349(B)−314(A) | Turn-helix | High-high | 14.7 | T348 | 0.862 | 0.517 |
| 5AR6 | 64(A)−71(A) | Coil-beta | High-high | 17 | N70 | 0.879 | 0.451 |
| 3VPL | 159(A)−189(A) | Turn-helix | High-medium | 15.7 | Y158 | 0.891 | 0.409 |
| 2JJB | 539(B)−539(B) | Coil-coil | Medium-low | 24.6 | P534, A544, C540 | 0.904 | 0.681 |
| 1R0R | 16(I)−35(I) | Helix-helix | High-high | 7.7 | F37 | 0.934 | 0.498 |
| 3WMT | 308(A)−297(A) | Bend-coil | High-high | 8.2 | L296 | 0.944 | 0.526 |
| 3WMT | 308(B)−297(B) | Coil-coil | High-high | 8.9 | L296 | 0.946 | 0.494 |
| 3WMT | 308(B)−297(B) | Coil-coil | High-high | 8.9 | L296 | 0.946 | 0.494 |
| 1KNM | 76(A)−59(A) | Coil-Beta | High-high | 14.9 | W77 | 0.948 | 0.527 |
| 3B7E | 318(B)−336(B) | Coil-bend | High-high | 23.4 | D387 | 0.949 | 0.519 |
| 3B7E | 318(A)−336(A) | Coil-bend | High-high | 22.3 | D387 | 0.95 | 0.538 |
| 2XXL | 314(B)−349(A) | Helix-turn | High-high | 14.7 | T348, V347 | 0.951 | 0.505 |
| 3MWQ | 197(A)−204(A) | Coil-beta | NA[d]-NA[d] | 14.8 | D253 | 0.954 | 0.439 |
| 3TRS | 18(B)−101(B) | Coil-coil | High-high | 15.6 | G15 | 0.954 | 0.408 |
| 2X5X | 85(A)−36(A) | Turn-bend | Medium-medium | 12 | D84 | 0.958 | 0.434 |
| 3LUM | 277(D)−250(A) | Turn-turn | High-high | 2.9 | Q262, A263 | 0.965 | 0.419 |
| 2UWA | 218(B)−226(B) | Beta-coil | High-low | 9.8 | T219 | 0.976 | 0.524 |
| 2D1Z | 382(A)−365(A) | Coil-beta | Medium-medium | 14.1 | E357, R359 | 0.977 | 0.548 |
| 1LLF | 60(A)−97(A) | Coil-coil | High-high | 4.5 | R361 | 0.977 | 0.423 |
| 3NKQ | 73(A)−86(A) | Coil-coil | High-medium | 7.3 | N53 | 0.978 | 0.487 |
| 3NKQ | 73(A)−86(A) | Coil-coil | High-medium | 7.3 | N53 | 0.978 | 0.487 |
| 1KLI | 91(L)−102(L) | Turn-beta | High-high | 16.7 | E94, S89 | 0.979 | 0.481 |
| 2UWA | 218(C)−226(C) | Beta-coil | High-low | 7.2 | T219 | 0.979 | 0.526 |
| 2UWA | 218(A)−226(A) | Beta-coil | High-low | 8.0 | T219 | 0.982 | 0.539 |
| 2JJB | 533(A)−539(A) | Bend-coil | Medium-low | 21.6 | P534, A544, C539 | 0.983 | 0.54 |
| 2JJB | 533(D)−539(D) | Bend-coil | Medium-low | 23.6 | P534, A544, C541 | 0.983 | 0.54 |
| 3B8Z | 342(A)−394(A) | Helix-beta | High-high | 23.3 | C394 | 0.983 | 0.413 |
| 2ODP | 655(A)−685(A) | Coil-turn | High-high | 7.5 | R694 | 0.984 | 0.407 |
| 4DO4 | 49(A)−42(A) | Bend-turn | High-high | 10.6 | C42, I40 | 0.986 | 0.553 |
| 3PMS | 231(A)−252(A) | Helix-coil | High-high | 6.4 | P246 | 0.986 | 0.447 |
| 4O36 | 65(B)−72(B) | Coil-beta | High-high | 14.6 | Q69, N71, C72 | 0.986 | 0.432 |
| 4DO4 | 49(B)−42(B) | Bend-turn | High-high | 11.4 | C42, I40 | 0.99 | 0.544 |
| 3PMS | 208(A)−204(A) | Bend-beta | High-high | 15 | Y161 | 0.997 | 0.452 |
| 2GMN | 181(B)−201(B) | Coil-coil | High-high | 29.7 | H101, H177 | 0.998 | 0.489 |
| 2GMN | 181(A)−201(A) | Coil-coil | High-high | 29.8 | H101, H177 | 0.999 | 0.493 |
| 3WWC | 111(A)−99(A) | Turn-helix | High-high | 14.2 | S66 | 1 | 0.498 |
| 4UZ1 | 413(A)−432(A) | Helix-beta | High-high | 16.1 | R409, H412 | 1 | 0.503 |
| 3EQN | 698(A)−692(A) | Bend-helix | High-high | 7.5 | F684 | 1 | 0.414 |
| 3EQN | 692(B)−692(B) | Helix-helix | High-high | 7.9 | F684 | 1 | 0.414 |
| 4BDX | 33(A)−5(A) | Beta-beta | NA[d]-NA[d] | 6.2 | L4 | 1 | 0.476 |
| 3LZT | 127(A)−6(A) | Coil-helix | high-High | 5.8 | R128 | 1 | 0.402 |
| Cystine in the microenvironment of Ligase | | | | | | | |

**Table VII**
(Continued)

| PDB ID | Cystine (chain) | 2° structures | Sequence conservations[a] | Strain energy (KJ/mol) | Active residues in 4.5 Å region in half-cystine | BF | rHpy |
|---|---|---|---|---|---|---|---|
| 2PHN | 248(B)−244(B) | Helix-helix | High-medium | 12.8 | C244 | 0.889 | 0.501 |
| 2PHN | 248(A)−244(B) | Helix-helix | High-medium | 12.8 | C244 | 0.895 | 0.474 |
| Cystine in the microenvironment of Llyases | | | | | | | |
| 1Y7W | 221(A)−31(A) | Coil-helix | High-high | 11.1 | L216 | 1 | 0.566 |
| 1Y7W | 221(B)−31(B) | Coil-helix | High-high | 11.4 | L216 | 1 | 0.596 |
| Cystine in the microenvironment of Transferases | | | | | | | |
| 5AJO | 473(A)−456(A) | Coil-beta | NA[d]-NA[d] | 12.4 | Q451 | 0.804 | 0.43 |
| 4WMA | 356(A)−385(A) | Coil-turn | NA[d]-NA[d] | 20.3 | N384 | 0.955 | 0.444 |
| 5FOE | 414(A)−407(A) | Coil-coil | High-high | 9.7 | T1045, I1022 | 0.996 | 0.525 |
| Cystine in microenvironment of Oxidoreductases | | | | | | | |
| 3NT1 | 36(B)−47(B) | Turn-beta | Low-medium | 10.9 | Y55 | 0.828 | 0.418 |

Secondary structures and sequence conservation of each half-cystines were reported along with the strain energy of the cystine disulphide bonds. PDB IDs are arranged according to the increasing values of buried fraction (second last column). Amino acids and their positions involved in enzyme active sites are mentioned in third last column.
[a]Sequence conservation obtained from PdbSum.[38]
[b]Includes beta sheet, beta bridge and beta strand secondary structures calculated by DSSP program.[27,28]
[c]Includes 3(10)-helix, alpha-helix and π-helix secondary structures calculated by DSSP program.[27,28]
[d]Sequence conservation is not available by ConSurf server.[36]

### Exposed-hydrophilic microenvironment cluster hosts half-cystines as a part of enzyme active sites or catalytic sites

Physico-chemical properties (in terms of buried fraction and rHpy values) of this microenvironment cluster are most dissimilar with respect to the embedded cystine residue. Average buried fraction and rHpy values for this cluster are 0.328 and 0.720 respectively (Table I), in contrast to 0.91 and 0.22 for current cystine microenvironment dataset. As this microenvironment cluster is extremely mismatched to the overall hydrophobicity of cystine, very less data points were recorded in this cluster, compared to other two clusters (Table I). There were only 31 half-cystines recorded as part of enzymes, 28 out of those were hydrolases (Table III). Out of these 28, only 5 half cystines were detected as a part of the microenvironment of active site in hydrolases (Table VIII). No half-cystines were reported as part of the active sites in hydrolases, in contrary to those in buried hydrophilic cluster. This presumably indicates that half-cystines occupying the exposed surface of the protein rarely participate in direct enzymatic reactions, however, those are engaged in protecting (via microenvironment) the enzyme active sites. All these half-cystines are highly scattered in the buried fraction, rHpy space (Fig. 5 and Table VIII).

## CONCLUSION

It was already known that similar folds lead to similar functions (enzyme activities) in proteins.[41–43] In this study we have shown that cystines from different enzymes, evolved within similar microenvironments, when present at the active site of same enzyme class. Our underlying aim was to test the hypothesis that cystines with similar functions (enzyme activities) from

**Table VIII**
Half-Cystines (First Out of the Pair of Half-Cystines from a Disulphide Shown in the Table) Present in Embedded Microenvironment of Active Sites in Hydrolase Enzyme Class, Reported from Exposed-Hydrophilic Cluster

| PDB ID | cystine (chain) | 2° structures | Sequence conservations[a] | Strain energy (KJ/mol) | Active residues in 4.5 Å region in half-cystine | BF | rHpy |
|---|---|---|---|---|---|---|---|
| 1ZGX | 96(B)−7(B) | Coil-coil | High-low | 11.4 | K94 | 0.206 | 0.821 |
| 2G58 | 96(A)−84(A) | Helix[b]-helix | High-high | 10.8 | K74 | 0.31 | 0.699 |
| 2VB1 | 6(A)−127(A) | Helix-helix | High-high | 5.1 | E7, I124 | 0.367 | 0.719 |
| 3EDH | 65(A)−64(A) | Coil-coil | High-high | 13 | C64 | 0.438 | 0.602 |
| 3RLG | 53(A)−201(A) | Turn-turn | High-high | 14.4 | N200 | 0.447 | 0.652 |

Secondary structures and sequence conservation of each half-cystines were reported along with the strain energy of the cystine disulphide bond. PDB IDs are arranged according to the increasing values of buried fraction (second last column). Amino acids and their positions involved in enzyme active sites are mentioned in third last column.
[a]Sequence conservation obtained from PdbSum.[38]
[b]Includes 3(10)-helix, alpha-helix and π-helix secondary structures calculated by DSSP program.[27,28]

different proteins will belong to similar microenvironment clusters. To this end, by using hierarchical clustering method, we have identified three different microenvironment clusters: I) buried-hydrophobic, ii) buried-hydrophilic and exposed-hydrophilic, and correlated these with their enzymatic functions. In our analysis, special emphasis has been given to cystines from enzyme classes. As demonstrated here, cystines from -C-x-x-C- motifs in Oxidoreductase enzyme class all have very similar microenvironment, that is, buried and hydrophobic. The catalytic cystines from the hydrolase enzyme class always prefer partly exposed hydrophilic microenvironment (that is buried hydrophilic cluster). Other cystines from hydrolase enzymes which participate in stabilizing catalytic or active sites were detected in both buried hydrophilic and exposed hydrophilic microenvironment clusters. Despite of low to medium sequence conservation in some of the cystines, active cystines from hydrolase enzymes share similar microenvironments. This report illustrates that cystine residues share similar microenvironments at active sites of specific enzyme classes despite of variation in their sequence similarities and sequence position conservations, thus validating the working hypothesis. We believe this conclusion should be further verified for other amino acids, particularly, titratable amino acids, like Aspartic acid, Glutamic acid, Arginine and so forth Titratable amino acids are expected to be more sensitive toward change in microenvironment due to alteration in their protonation states. Hence, microenvironment modulated switching of protonation states in titratable amino acids would be of interest in major biochemical reactions, like photosynthesis. For example, carbon dioxide fixing enzyme, Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCo), involves seven charged residues in its active site. It is known that switching some of these protonation states could lower the transition state, thus assist to prevent backward reaction and trap more carbon dioxide.[44] This can be further verified by modifying the protein structures with altered protonation states of charged residues and test the capacity of carbon dioxide fixation.

Amino acids in heterogeneous protein microenvironments are analogous to amino acids in different solvents with variable dielectric media. As solubility, bond dissociation energy and spectral properties of cystine vary from hydrophilic to hydrophobic solvents, the same is also expected when cystine is transferred from hydrophilic part of the protein microenvironment to its hydrophobic part. This could be exploited to guide experimentation in to the local dielectric medium within protein microenvironments. We have shown here that disulfide-bridged cystine molecule is a good model system to examine the effect of various dielectric medium on S-S bond dissociation energy and can be extended for experimental verifications.

## ACKNOWLEDGMENTS

Authors sincerely acknowledge Dr. Marcin Apostol, ADRx. Inc. Thousand Oaks, California, USA, for kindly providing the program for dihedral strain energy calculations.

## REFERENCES

1. Rekker RF. The hydrophobic fragmental constant. Amsterdam: Elsevier; 1977.
2. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
3. Wimley WC, Creamer TP, White SH. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry 1996;35:5109–5124.
4. 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.
5. MacCallum JL, Bennett WFD, Tieleman DP. Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. J Gen Physiol 2007;129:371–377.
6. McIntosh TJ, Simon SA. Bilayers as protein solvents: role of bilayer structure and elastic properties. J Gen Physiol 2007;130:225–227.
7. White SH. Membrane protein insertion: the biology-physics nexus. J Gen Physiol 2007;129:363–369.
8. Wolfenden R. Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. J Gen Physiol 2007;129:357–362.
9. Bandyopadhyay D, Mehler EL. Quantitative expression of protein heterogeneity: response of amino acid side chains to their local environment. Proteins 2008;72:646–659.
10. Jiang Y, Ruta V, Chen J, Lee A, Mackinnon R. The principle of gating charge movement in a voltage-dependent K+ channel. Nature 2003;423:42–48.
11. Moukhametzianov R, Klare JP, Efremov R, Baeken C, Göppner A, Labahn J, Engelhard M, Büldt G, Gordeliy VI. Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. Nature 2006;440:115–119.
12. Harris TK, Turner GJ. Structural basis of perturbed pK a values of catalytic groups in enzyme active sites. IUBMB Life 2002;53:85–98.
13. Ray S, Bhattacharyya M, Chakrabarti A. Conformational study of spectrin in presence of submolar concentrations of denaturants. J Fluoresc 2005;15:61–70.
14. Bairoch A. The ENZYME database in 2000. Nucleic Acids Res 2000; 28:304–305.
15. Branden C, Tooze J. Introduction to protein structure, 2nd ed. New York: Garland Publishing Taylor and Francis group; 1991.
16. Lundstrom-Ljung J, Holmgren A. Prolyl hydrolase, protein disulfide isomerase, and other structurally related proteins. New York: CRC Press; 1998.
17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
18. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38, 27–28.
19. 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.
20. 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: D485–D489.
21. Hartigan JA. Clustering algorithms. New York: Wiley; 1975.
22. Tryon RC, Bailey DE. Cluster analysis. New York: McGraw-Hill; 1973.

23. Ward J. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963;58:236–244.
24. Addinsoft. http://www.xlstat.com XLSTAT 2014, Data analysis and statistics software for Microsoft Excel; Paris, France, 2014.
25. Andreeva A, Howorth D, Chandonia J-M, Brenner SE, Hubbard TJP, Chothia C, Murzin AG. Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 2008;36: D419–D425.
26. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr Sect D Biol Crystallogr 2004;60:2256–2268.
27. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
28. Joosten RP, te Beek TAH, Krieger E, Hekkelman ML, Hooft RWW, Schneider R, Sander C, Vriend G. A series of PDB related databases for everyday needs. Nucleic Acids Res 2011;39:D411–D419.
29. Schmidt B, Ho L, Hogg PJ. Allosteric disulfide bonds. Biochemistry 2006;45:7429–7433.
30. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res 2015;43: W580–W584.
31. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics 2007;23: 1073–1079.
32. Braberg H, Webb BM, Tjioe E, Pieper U, Sali A, Madhusudhan MS. SALIGN: a web server for alignment of multiple protein sequences and structures. Bioinformatics 2012;28:2072–2073.
33. Harrison PM, Sternberg MJ. The disulphide beta-cross: from cystine geometry and clustering to classification of small disulphide-rich protein folds. J Mol Biol 1996;264:603–623.
34. Chuang C-C, Chen C-Y, Yang J-M, Lyu P-C, Hwang J-K. Relationship between protein structures and disulfide-bonding patterns. Proteins 2003;53:1–5.
35. Katz BA, Kossiakoff A. The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. J Biol Chem 1986;261:15480–15485.
36. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 2005;33:W299–W302.
37. Røhr ÅK, Hammerstad M, Andersson KK. Tuning of thioredoxin redox properties by intramolecular hydrogen bonds. PLoS One 2013;8:e69411.
38. de Beer TAP, Berka K, Thornton JM, Laskowski RA. PDBsum additions. Nucleic Acids Res 2014;42:D292–D296.
39. Simone ADE, Berisio R, Zagari A, Vitagliano L. Limited tendency of α -helical residues to form disulfide bridges: a structural explanation. J Pept Sci 2006;12:740–747.
40. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 2010;38:W529–W233.
41. Chen S, Bahar I. Mining frequent patterns in protein structures: a study of protease families. Bioinformatics 2004;20:18.
42. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol 2004;14:208–216.
43. Holbourn KP, Acharya KR, Perbal B. The CCN family of proteins: structure-function relationships. Trends Biochem Sci 2008;33: 461–473.
44. Collings AF, Critchley C. Artificial photosynthesis: from basic biology to industrial application. In: Anthony F. Collings et al., editors. Weinheim, Germany, Wiley; 2007. pp 280–282.