PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Disulfide conformation and design at helix N-termini

S. Indu,[1] Senthil T. Kumar,[1] Sudhir Thakurela,[1] Mansi Gupta,[1] Ramachandra M. Bhaskara,[1] C. Ramakrishnan,[1] and Raghavan Varadarajan[1,2]*

[1] Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India
[2] Chemical Biology Unit, Jawaharlal Nehru Center for Advanced Scientific Research, Jakkur, Bangalore 560 004, India

ABSTRACT

To understand structural and thermodynamic features of disulfides within an α-helix, a non-redundant dataset comprising of 5025 polypeptide chains containing 2311 disulfides was examined. Thirty-five examples were found of intrahelical disulfides involving a CXXC motif between the N-Cap and third helical positions. GLY and PRO were the most common amino acids at positions 1 and 2, respectively. The N-Cap residue for disulfide bonded CXXC motifs had average (ϕ,ψ) values of (−112 ± 25.2°, 106 ± 25.4°). To further explore conformational requirements for intrahelical disulfides, CYS pairs were introduced at positions N-Cap-3; 1,4; 7,10 in two helices of an *Escherichia coli* thioredoxin mutant lacking its active site disulfide (nSS Trx). In both helices, disulfides formed spontaneously during purification only at positions N-Cap-3. Mutant stabilities were characterized by chemical denaturation studies (in both oxidized and reduced states) and differential scanning calorimetry (oxidized state only). All oxidized as well as reduced mutants were destabilized relative to nSS Trx. All mutants were redox active, but showed decreased activity relative to wild-type thioredoxin. Such engineered disulfides can be used to probe helix start sites in proteins of unknown structure and to introduce redox activity into proteins. Conversely, a protein with CYS residues at positions N-Cap and 3 of an α-helix is likely to have redox activity.

Proteins 2010; 78:1228–1242.

Key words: intrahelical disulfides; CXXC motifs; MODIP; torsion angle; chemical denaturation; thermostability; redox activity.

INTRODUCTION

Disulfides are the most common covalent interactions within a protein molecule involved in cross-linking of sequentially distant residues. By bringing about entropic destabilization of the unfolded state, disulfides can confer stability on the protein.[1,2] Disulfide engineering is thus an important method of enhancing protein stability. Although the removal of a naturally-occurring disulfide in a protein molecule results in destabilization, the introduction of a non-naturally occurring disulfide bridge does not typically result in stabilization of the protein molecule.[3–5] This might be a consequence of the energetic penalty required for rearrangement of residues surrounding the cysteines to accommodate the disulfide, as well as due to loss of favorable interactions involving the original wild-type (WT) residues.[2] Alternatively, the conformational characteristics of the disulfide formed might be energetically unfavorable. A study of cross-strand disulfides in antiparallel β-sheets had previously shown that while disulfides formed at the non-hydrogen bonded registered pairs are stabilizing, those in the hydrogen bonded positions are destabilizing.[5,6] Hence, an understanding of disulfides in the context of protein structures could provide valuable information for protein engineering. Engineered disulfides can also act as conformational constraints and to validate protein models.[7] Hence, it is important to understand the conformational determinants of disulfides formed within and between different types of secondary structures.

Intrahelical disulfides are a characteristic feature of active sites of several oxidoreductases.[8,9] In addition to playing an important functional role, they also stabilize the protein.[3] The importance of the CXXC (CYS-X-X-CYS tetrapeptide) motif at the N-termini of helices in catalysis of redox reactions is well known.[10–14] This motif has been aptly described as a rheostat in the active site of oxidoreductases.[9] The motif can be

Additional Supporting Information may be found in the online version of this article.
Abbreviations: CD, circular dichroism; CGH10, sodium citrate 10 m*M*, glycine 10 m*M*, HEPES 10 m*M*; DSC, differential scanning calorimetry; DTT, dithiothreitol; GdmCl, guanidinium chloride; MODIP, modeling of disulfide bridges in proteins; PDB, protein data bank; Trx, thioredoxin; UV, ultra-violet; WT, wild type.

S. Indu and Senthil T. Kumar contributed equally to this work.
Grant sponsors: Council of Scientific and Industrial Research, Government of India; Department of Biotechnology, Government of India.
*Correspondence to: R. Varadarajan, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. E-mail: varadar@mbu.iisc.ernet.in.
Received 10 July 2009; Revised 26 September 2009; Accepted 13 October 2009
Published online 20 October 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22641

highly reducing as in thioredoxin or highly oxidizing as in DsbA.[15–17] Altering the two XX residues alters the reduction potential of the disulfide. The redox potential variation and the change in activity of several oxidoreductases have been studied following mutation of the XX residues.[8,15,18–22] The CXXC motif in oxidoreductases has also been shown to play a role in conferring substrate specificity.[23] The CXXC motif is a characteristic feature of the thioredoxin fold. Minor variations in the fold structure and active site leads to a diverse collection of proteins with varying functions.[24] Homology searches for oxidoreductases carrying CXXC motifs has revealed the presence of proteins with CXXC derived sequences like CXXS, CXXT, TXXC, and SXXC. These CXXC derived sequences also occur at the N-termini of helices and are believed to confer redox activity.[25]

A synthetic helical peptide containing CAAC at its N-terminus was shown to form a disulfide with a redox potential of −230 mV and disulfide formation was shown to increase helicity in this peptide.[26] It is not known if introduction of CXXC motifs in protein helices that lack this motif can confer redox function. Other potential disulfide forming motifs such as CXC and CC have been introduced at or near helices to alter activity and engineer redox switches, respectively.[27,28] However, both CXC and CC motifs do not occur in disulfide bonded forms in naturally-occurring protein helices.

In the present work, we have carried out a comprehensive analysis of the stereochemical characteristics of disulfides occurring within α-helices. From the analysis of intrahelical disulfides in the Protein Data Bank (PDB), experimental studies on thioredoxin mutants and modeling studies, we attempt to understand requirements for disulfide formation with respect to positions and conformations of cysteines within an α-helix.

All naturally-occurring intrahelical disulfides occurred at N-termini of helices with the N-terminal CYS being the N-Cap residue (See methods section for N-Cap definition). Further, all naturally-occurring intrahelical disulfides have two residues between the cysteines. These observations prompted us to engineer CXXC motifs at N-termini as well as the interior of helices of *Escherichia coli* thioredoxin to ascertain the ability of different regions of helices to accommodate disulfide bridges. Spontaneous disulfide formation was observed only for mutants where the N-terminal CYS was the N-Cap residue. The effects of engineered helical disulfides on the stability and activity of the protein were determined.

MATERIALS AND METHODS

Dataset generation

Using structures with resolution better than 2.0 Å and a sequence identity cutoff of 40%, a non-redundant protein dataset of 5025 polypeptides with 2311 disulfides was generated for analysis. This protein dataset was used to identify helical segments.

Different criteria (such as hydrogen bonding patterns with and without torsional angle criteria) have been previously used to assign secondary structures in PDB records.[29,30] In the present study, to ensure internal consistency, we adopted a backbone torsion angle based method as previously mentioned by Ramakrishnan and Srinivasan[31] for identification of helices. The values of torsion angle ranges used for identifying helical residues were −120° to −10°, and −120° to 20° for ϕ and ψ, respectively. Stretches of consecutive, helical amino acids at least four residues long were identified as helices. The residues with non-helical (ϕ, ψ) values immediately preceding the N-terminus of a helix and succeeding the C-terminus of a helix are referred to as the N-Cap and C-Cap residues, respectively.[32,33] A helical segment is defined as extending from the N-Cap residue to C-cap residue. Thus, a helical segment consists of an N-Cap residue, a stretch of amino acids with helical (ϕ, ψ) values that is at least four residues long, and a C-Cap residue. The numbering of all helical segments begins at 1 from the residue immediately following the N-Cap residue. A dataset of 29,597 helical segments was obtained from our non-redundant protein dataset.

Modeling of disulfides

The program MODIP was used to model disulfides in helical segments.[34] The non-glycine residue pairs that satisfy the MODIP criteria for C^{α}–C^{α} (3.8 Å−7.0 Å) and C^{β}–C^{β} (3.4 Å−4.5 Å) distances are chosen for geometrical fixation of sulfur atoms. Predicted disulfides are assigned a grade ranging from A–C based on the S-S bond length and dihedral angles of the modeled cystines. The various criteria that the software MODIP uses have been described previously.[34,35]

In silico mutation and energy minimization studies

As an alternative to MODIP, the program Swiss-PDB Viewer[36] was used to model CYS mutations *in silico* following which energy minimizations were carried out in vacuum using the l-bfgs minimizer and OPLS-AA/L force field in the GROMACS 3.3.3 package.[37]

Experimental materials

Enzymes and reagents for mutagenesis were obtained from New England BioLabs. Q-Sepharose resin and NAP-10 columns were purchased from GE Healthcare. Dithiothreitol (DTT) and Guanidinium chloride (GdmCl) (MB grade) were purchased from USB chemicals. All other reagents used in protein purification and characterization were obtained from Sigma.

Mutagenesis

The WT thioredoxin (Trx) gene was cloned into the pET20b vector as described previously.[38] Both active site cysteines in WT Trx were mutated to serines. This latter thioredoxin construct is henceforth referred to as nSS Trx. All other mutants were made in the nSS Trx background. CXXC motifs were introduced in helices of nSS Trx by site directed mutagenesis. There were three kinds of mutants generated; N-CapC-3C, 1C-4C, and 7C-10C where the numbers denote the position of the indicated residue in the helix. Residue numbering in helices is as defined in the dataset generation section. All mutants were generated by site directed mutagenesis following the Stratagene QuikChange protocol as described previously.[39] The mutants generated in this work are listed in Table I.

Expression and purification of nSS Trx mutants

All mutants were transformed into *E. coli* BL21 (DE3) cells and expressed under control of the T7 promoter at 37°C. Cells were harvested by centrifugation, subjected to chloroform shock, and purified by anion exchange chromatography as described previously.[38] The purified proteins were >95% pure as assayed by SDS-PAGE. Protein concentrations were estimated by absorbance measurements at 280 nm in a Jasco V-530 spectrophotometer. The extinction coefficients calculated from protein sequences[40] for nSS Trx and its disulfide containing mutants are 13,980M^{-1} cm^{-1} and 14,105M^{-1} cm^{-1}, respectively

Iodoacetamide labeling of free CYS residues followed by ESI-MS

Thirty micromolar protein was treated with 100 m*M* iodoacetamide in the presence of 100 m*M* Tris, 3*M* GdmCl, pH 8.0 for 4 min at room temperature.[41] The labeling reaction was quenched with 5% formic acid. All samples were desalted using NAP-10 columns into 0.1% formic acid for mass spectrometry. The desalted samples were analyzed by direct injection on a Bruker Daltonics Esquire 3000 plus electrospray ionization mass spectrometer in positive ion mode to determine whether the CYS residues were in the oxidized or reduced state and whether disulfides were intramolecular or intermolecular. Protein samples without the iodoacetamide treatment served as the control.

Table I
Description of nSS Trx Mutants

WT Trx sequence[a]	CX_1X_2C mutant[b]	Position of CX_1X_2C in the helix[c]
NIDQ (59–62)	CIPC	N-Cap-3
NIDQ (59–62)	CIDC	N-Cap-3
NIDQ (59–62)	CGPC	N-Cap-3
SKGQ (95–98)	CKGC	N-Cap-3
SKGQ (95–98)	CGPC	N-Cap-3
IDQN (60–63)	CDQC	1–4
KGQL (96–99)	CGQC	1–4
FLDA (102–105)	CLDC	7–10

[a]Sequence before mutation in nSS Trx. The numbers in parentheses are the residue numbers of the two residues mutated to cysteines.
[b]The identities of residues X_1,X_2 between the introduced cysteines are shown.
[c]N-Cap is the first residue of the relevant helix as defined in PDB records. The residue immediately C-terminal to N-Cap in the helix is residue 1.

In vitro oxidation of protein

To facilitate the oxidation of cysteines in certain mutants, each of the proteins (80–100 μ*M*) was incubated with 5 m*M* 1,10-phenanthroline monohydrate and 1.5 m*M* copper sulfate in 100 m*M* Tris buffer (pH 8.0) at 20°C for 24 h.[42] Following oxidation, 10 mL of the protein was dialyzed against 1 L of 10 m*M* Sodium citrate, 10 m*M* Glycine, 10 m*M* HEPES (CGH 10) buffer, pH 7.0 at 4°C with 4 buffer changes at 4 h intervals.

Circular dichroism

The instrument was calibrated with D-10-camphor sulfonic acid as described.[43] Far UV CD spectra were collected for all the mutants on a Jasco715 spectropolarimeter flushed with nitrogen gas at 25°C. A cuvette of path length 1 mm was used and four scans were averaged to obtain the spectrum for each protein. CD spectra from 200–250 nm were acquired using 50 nm min^{-1} scan rate, 0.5 s response time, 1 nm bandwidth, and 2 nm data pitch. The proteins were in CGH 10 buffer, pH 7.0. For proteins in the reduced state, the buffer included 1 m*M* DTT in addition to the components mentioned above. The concentration of each protein was 10 μ*M*. Spectra were corrected for buffer contributions by subtraction of corresponding buffer spectra collected under identical conditions.

Chemical denaturation studies

Stability studies were carried out by monitoring the CD signal at 220 nm in a Jasco J-715 CD spectropolarimeter for 10 μ*M* protein in different concentrations of GdmCl in the buffer CGH10 at 25°C. For denaturation profiles of reduced proteins, the samples included 1 m*M* DTT in addition to the components mentioned above. Denaturant concentrations were ascertained by refractive index measurements using an Abbe refractometer (Techno Instruments and Chemicals). Prior to measurement of CD signals, the samples were incubated with denaturant for 3 h at 25°C. The data obtained was fit to a two state denaturation profile to obtain the various thermodynamic parameters as described previously.[44]

Thermal denaturation studies

Differential scanning calorimetry (DSC) was performed using a VP-DSC instrument (Microcal) to ascertain the thermal stability of the proteins. A scan rate of 60°C h^{-1} was used. The concentration of each protein used was 25 μ*M* in CGH10 buffer, pH 7.0. Two scans were performed in succession to determine the reversibility of the transitions. The data was fit to a two state transition using Origin software to obtain thermodynamic parameters as described previously.[38,45,46]

Insulin reduction assay[47]

All experiments were carried out at 25°C in a Jasco V-530 spectrophotometer using a 1 cm path length cuvette. Each assay mixture contained 0.1*M* phosphate buffer (pH 7.0), 2 m*M* EDTA, 0.13 m*M* porcine insulin, 0.33 m*M* DTT, and 10 μ*M* of the protein to be studied. Following DTT addition, the formation of insoluble insulin B-chain aggregates was monitored by measurement of the intensity of scattered light at 650 nm as a function of time. As control experiments, the activity of WT Trx, nSS Trx, MBP230C 30C, and Trx20C 73C were monitored.

RESULTS AND DISCUSSION

Location of intrahelical disulfides and CXXC motifs

Using the helix definitions described in the methods section and disulfide identification by using SSBOND records in the PDB files, the non-redundant dataset was examined for intrahelical disulfides. Thirty-five instances of intrahelical disulfides were found. All 35 naturally-occurring intrahelical disulfides were found between N-Cap and third residue of helices. Naturally-occurring intrahelical disulfides exclusively occur as a CXXC motif. The requirement for a CXXC motif is indicative of the optimal proximity that the CYS side chains can achieve in a helix for the formation of disulfide when the loop length is two. In this situation, both residues are on the same face of the helix, given that an α-helix typically has 3.6 residues per turn.

The occurrence of all intrahelical disulfides in CXXC motifs led us to search for CXXC motifs in the protein dataset. Examination of the non-redundant protein dataset identified 642 CXXC motifs. Of these, 102 motifs occur at the N-termini of the helices such that the N-terminal CYS is the N-Cap residue. As mentioned earlier in this section, 35 of these 102 CXXC motifs are disulfide bonded (Table II[48]) and the remaining 67 are in the reduced state. Of the 35, 24 disulfides occurred in oxidoreductases and one is involved in metal ion transfer. For the remaining 10 instances, the disulfide does not appear to have a role in the function of the protein.

There are 67 instances of CXXC motifs between N-Cap and third residue with both CYS residues in the reduced state. Of these 67, the cysteines in 45 of them are involved in coordination of metal ions. In addition to the 45 metal coordinating proteins, there are four cytochromes where the heme moiety is coordinated by the cysteines. Nine proteins have redox function indicating that their CXXC motifs can also exist in the oxidized state. In the protein PLP synthase from *P. falciparum* (PDB ID: 2abw), the N-terminal CYS was part of the active site of the protein. For the remaining cases, the function of the proteins cannot be attributed to the cysteine oxidation state. A set of 20 proteins (out of the 67 described above) with nonbridged CXXC motifs are tabulated in Table III.[48] These 20 cases were predicted by MODIP as potential disulfide introduction sites. The program MODIP predicts residue locations in a protein that are stereochemically compatible with disulfide formation.[34,35]

The above observations and conclusions were based on specific dihedral angle based criteria to identify helical segments. An alternate method of identifying helical segments is to use the HELIX records in the header of each PDB file. A comparison of the helical segments carrying N-terminal helical CXXC motifs (N-terminal CYS of CXXC is the N-Cap of helix) defined by us with the corresponding helical segments defined using PDB records was done (Table SI, Supporting information). In 76 out of 102 instances, N-Cap defined by us coincides with the first residue of the helix in PDB records. There were 17, 4, and 3 instances found, where the first residue of the helix in the PDB records is the first, second, and third residue of helix, respectively, as defined in this article. In the remaining four cases, none of the residues that are part of the N-terminal helical CXXC motif identified using our torsional angle criteria are part of helices defined in the PDB records. This analysis suggests that defining helical segments through our torsional angle criteria is preferable to using the PDB helix records for the purposes of the present study.

There are 115 instances of CXXC motifs occurring within a helix such that all four residues have a helical conformation. There are 17 cases of a CXXC motif occurring at the C-termini of helices such that the C-terminal CYS is the C-Cap residue. However, none of them showed the presence of an intrahelical disulfide. There are a total of 408 instances of CXXC motifs in non-helical positions. Of these 408 cases, only 14 of them showed the presence of a disulfide. Thus, CXXC motifs in regions other than N-terminal helical positions have a very low tendency to form disulfides. None of the above disulfide containing 14 proteins show redox activity (Table IV[48]). Only one of them (PDB ID: 2cog) shows a variation in activity depending on the oxidation state of the cysteines.

Table II
Proteins with Intrahelical Disulfides from a Non-Redundant Dataset of 5025 Polypeptides

PDB ID	Chain ID	N-term CYS	C-term CYS	CX_1X_2C	Function of protein	Scop FOLD[a]
1a8l[b]		146	149	CPYC	Protein disulfide oxidoreductase from *A. pyrococcus*	Thioredoxin like
1a8l[b]		35	38	CQYC	Protein disulfide oxidoreductase from *A. pyrococcus*	Thioredoxin like
1aba[b]		14	17	CGPC	Bacteriophage T4 glutaredoxin	Thioredoxin like
1eej[b]	A	98	101	CGYC	*E. coli* DsbC	Thioredoxin like
1faa[b]	A	46	49	CGPC	Thioredoxin F from spinach chloroplast	Thioredoxin like
1fl2[b]	A	345	348	CPHC	*E. coli* alkylhydroperoxide reductase	FAD/NAD(P)-binding domain
1gai		210	213	CSWC	*A. awamori* glucoamylase-471 with D-gluco-dihydroacarbose	Alpha/Alpha Toroid
1goi	A	328	331	CEEC	*S. marcescens* chitinase B D140N mutant	WW-domain like
1h75[b]	A	11	14	CVQC	Glutaredoxin like protein from *E. coli*	Thioredoxin like
1jfu[b]	A	72	75	CVPC	Thiol disulfide interchange protein from *B. japonicum*	Thioredoxin like
1jr8[b]	A	54	57	CGEC	Sulfhydryl oxidase from *S. cerevisiae*	Four-helical up-and-down bundle
1kng[b]	A	92	95	CVPC	Thioredoxin like protein from *B. japonicum*	Thioredoxin like
1lsh	A	198	201	CPTC	Lipovitellin from silver lamprey	Alpha-alpha superhelix
1oqc	A	62	65	CEEC	FAD dependent sulfhydryl oxidase from *R.. norvegicus*	Four-helical up-and-down bundle
1osd[c]	A	14	17	CSAC	Hypothetical protein from *R. metallidurans*	Ferredoxin-like
1qnr	A	26	29	CYWC	Mannanase from *T. reesei*	TIM beta/alpha-barrel
1r26[b]	A	30	33	CGPC	Thioredoxin from *T. brucei*	Thioredoxin like
1sen[b]	A	66	69	CGAC	Human thioredoxin like protein	Thioredoxin like
1thx[b]		32	35	CGPC	Thioredoxin from *Anabena*	Thioredoxin like
1v58[b]	A	109	112	CPYC	*E. coli* DsbG	Thioredoxin like
1v98[b]	A	62	65	CGPC	Thioredoxin from *T. thermophilus*	No SCOP entry
1vke	A	59	62	CDDC	Carboxymuconolactone decarboxylase from *T. maritima*	AhpD-like
1w4v[b]	A	31	34	CGPC	Human thioredoxin	No SCOP entry
1wou[b]	A	43	46	CPDC	Thioredoxin related protein from human	Thioredoxin like
1z3e[b]	A	10	13	CTSC	Transcription regulator from *B. subtilis*	Thioredoxin like
2a40	A	101	104	CESC	Actin from rabbit in a ternary complex	Ribonuclease H-like motif
2axo[b]	A	54	57	CASC	Hypothetical protein from *A. tumifaciens*	Thioredoxin like
2b1k[b]	A	80	83	CPTC	DsbE from *E.coli*	No SCOP entry
2b94	A	68	71	CAIC	Purine nucleoside phosphorylase homologue from *P. falciparum*	No SCOP entry
2f51[b]	A	35	38	CGPC	*T. vaginalis* thioredoxin	No SCOP entry
2f9s[b]	A	74	77	CEPC	*B. subtilis* thiol-disulfide oxidoreductase resA	Thioredoxin like
2fx5	A	243	246	CSLC	*P. mendocina* lipase	No SCOP entry
2i4a[b]	A	32	35	CGPC	Thioredoxin from *A. aceti*	No SCOP entry
2j23[b]	A	30	33	CGPC	Thioredoxin from *M. sympodialis*	No SCOP entry
2ouw	A	83	86	CSYC	Alkylhydroperoxidase from *R. rubrum*	No SCOP entry

[a]The folds for the protein cited are as mentioned in the SCOP database.[48]
[b]Proteins that either have the thioredoxin fold or function as oxidoreductases. In these cases (24 in number), the disulfide is directly implicated in protein function.
[c]CXXC motif is believed to play a role in metal ion transfer.

Residue identitites and main chain conformational preferences in intrahelical CXXC motifs

Figure 1 represents the distribution of residues occurring between the cysteines in intrahelical CXXC motifs. As is evident from the histograms, the most commonly occurring motif is CGPC. This is due to the presence of a large number of thioredoxin family members, which have this sequence as a characteristic feature. The most preferred residues at positions 2 and 3 of the motif are glycine and proline, both of which occur at frequencies of ~35%. However, other residues also occur between the cysteines.

The backbone torsion angles ϕ, ψ, and ω were examined for all residues in intrahelical CXXC motifs. With the exception of the N-terminal CYS, the other three residues had conformational preferences characteristic of a right-handed α-helix. The three residues following the N-Cap CYS involved in disulfide bonded intrahelical CXXC motifs have mean (ϕ, ψ) values of ($-58 \pm 6.0^\circ$, $-45 \pm 9.5^\circ$), ($-69 \pm 15.5^\circ$, $-27 \pm 15.2^\circ$), and ($-64 \pm 7.2^\circ$, $-40 \pm 6.8^\circ$), respectively. The N-Cap CYS had ϕ and ψ angles corresponding to a non-helical conformation [mean ϕ and ψ values of ($-112 \pm 25.2^\circ$, $106 \pm 25.4^\circ$), respectively] (see Fig. 2). The obligatory requirement for the N-terminal CYS residue to be non-helical explains the occurrence of intrahelical disulfides exclusively at N-termini. The ω angle (dihedral angle involving the amide bond) analysis revealed that there were no cis peptides present in intrahelical disulfide bonded CXXC motifs, despite the frequent occurrence of proline in these sequences.

Table III
Proteins with a MODIP Predicted, Non-Disulfide Bonded Intrahelical CX_1X_2C Motif from a Non-Redundant Dataset of 5025 Polypeptides

PDB ID	Chain ID	N-term CYS	C-term CYS	CX_1X_2C	Function of protein	Fold(SCOP)[a]
1erv[b]		32	35	CGPC	Human thioredoxin mutant	Thioredoxin like
1lu4[b]	A	1036	1039	CPFC	Alkylperoxidase from *M. tuberculosis*	Thioredoxin like
1svm	A	302	305	CLKC	SV40 large T antigen helicase	P-loop containing nucleoside Triphosphate hydrolases
1z6m[b]	A	36	39	CPYC	Protein of unknown function from *E. faecalis*	Thioredoxin like
1z6n[b]	A	66	69	CPDC	Protein of unknown function from *P. aeroginosa*	Thioredoxin like
1z84[c]	A	216	219	CCLC	Galactose-1-phosphate uridyl transferase like protein from *A. thaliana*	HIT-lke
1zma[b]	A	38	41	CPYC	Bacterocin transport accessory protein from *S. pneumoniae*	Thioredoxin like
2a1k[c]	A	87	90	CPYC	ssDNA binding protein core domain from *Enterobacter* phage RB69	No SCOP entry
2ckl[c]	A	39	42	CKTC	Polycomb complex protein BMI1 from *M. musculus*	No SCOP entry
2ckl[c]	B	72	75	CADC	Polycomb complex protein BMI1 from *M. musculus*	No SCOP entry
2fb6	A	74	77	CQDC	Protein of unknown function from *B. thetaiotaomicron*	No SCOP entry
2fwh[b]	A	461	464	CVAC	C-terminal domain of DsbD from *E. coli*	Thioredoxin like
2fyg[c]	A	74	77	CLYC	NSP10 from SARS coronavirus	Coronavirus NSP10-like
2gmy	A	48	51	CAFC	Polycomb comples protein *A. tumifaciens*	AhpD-like
2h30[b]	A	68	71	CPLC	Peptide methionine sulfoxide reductase from *N. gonorrhoeae*	No SCOP entry
2ht9[b]	A	37	40	CSYC	Human glutaredoxin	No SCOP entry
2nll[c]	B	357	360	CQEC	Thyroid hormone receptor from human	Glucocorticoid receptor-like
2o4d	A	48	51	CAYC	Hypothetical protein from *P. aeruginosa*	AhpD-like
2oik[c]	A	11	14	CELC	Histidine triad protein from *M. flagellatus*	No SCOP entry
4mt2[c]		41	44	CSQC	Metallothienin from *R. novewrgicus*	Metallothienin

[a]The folds for the protein cited are as mentioned in the SCOP database.[48]
[b]Proteins that either have the thioredoxin fold or function as oxidoreductases.
[c]Proteins where Zn^{2+} ion is coordinated through the CYS residues in CX_1X_2C.

Side chain conformational preferences of disulfide bonded CXXC motifs

A schematic description of side chain torsion angles in a disulfide bond is indicated in Figure 3(A) and the distribution of torsion angles is shown in Figure 3(B–D). The overall distribution of χss for all categories of disulfides shows two peaks at $+90^\circ$ and -90° [Fig. 3(B)]. In contrast, χss of intrahelical disulfides is distributed exclusively around $+90^\circ$.

The intrahelical disulfide bonded N-terminal CYS shows a distribution for χ_1 and χ_2 very different from that seen for other disulfide bonded cysteines [compare

Table IV
Proteins with Non-Helical Disulfide Bonded CX_1X_2C from a Non-Redundant Dataset of 5025 Polypeptides

PDB ID	Chain ID	N-term CYS	C-term CYS	CX_1X_2C	Function of protein	Fold(SCOP)[a]
2cog	A	335	338	CTGC	Human transferase	No SCOP entry
1d7c	A	121	124	CQGC	*P. chrysosporium* B-type *cytochrome*	Immunoglobulin-like beta-sandwich
1kuf	A	162	165	CDTC	*P. mucrosquamatus* hydrolase	Zincin-like
1qnr	A	172	175	CNGC	*T. reesei* hydrolase	TIM beta/alpha-barrel
2hl7	A	25	28	CPKC	*P. aeruginosa* C-type cytochrome	No SCOP entry
1k5c	A	300	303	CGNC	*C. purpureum* hydrolase	Single-stranded right-handed beta-helix
1mp8	A	456	459	CKNC	Human tyrosine protein kinase	Protein kinase-like (PK-like)
1dl2	A	468	471	CVDC	*S. cerevisiae* hydrolase	Alpha/Alpha Toroid
1m6y	A	46	49	CPGC	*T. maritima* SAM-dependent Methyltransferase Fold	SAM domain-like
1k3i	A	515	518	CGDC	*Fusarium* oxidoreductase	Immunoglobulin-like beta-sandwich
1rki	A	80	83	CDKC	*P. aerophilum* protein of unknown function	THUMP domain
1tib	A	104	107	CSGC	*T. lamiginoosus* carboxylic esterase	α/β hydrolase
1gof	A	515	518	CGDC	*H. rosellus* galactose oxidase	Immunoglobin like β-sanddwich
2apr	A	48	51	CTNC	*R. chinesis* acid protease	Acid protease

[a]The folds for the protein cited are as mentioned in the SCOP database.[48]

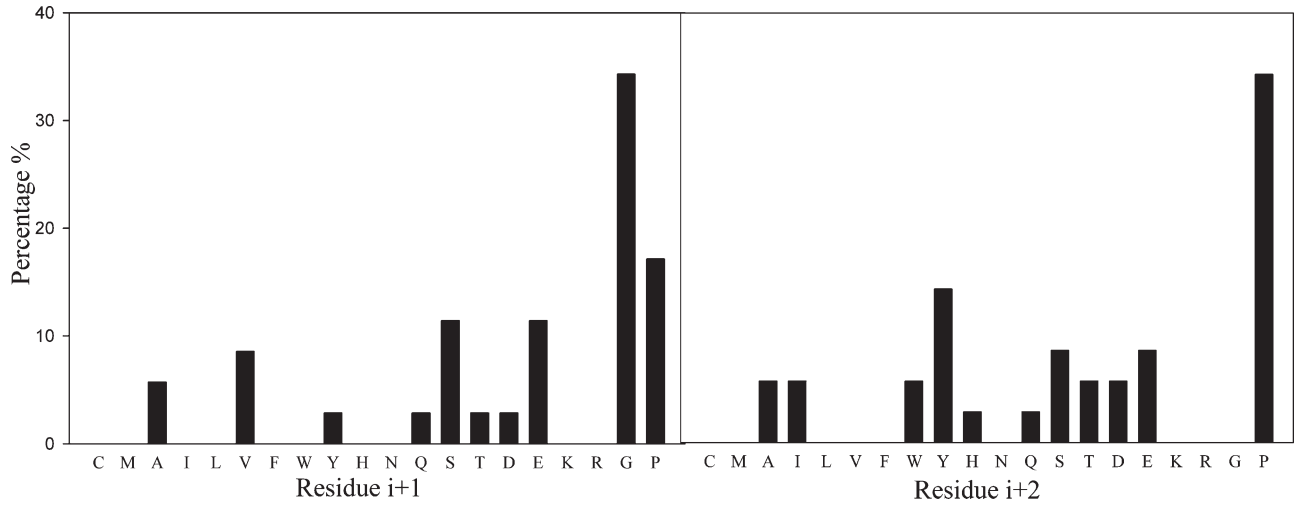

Figure 1

Residue preferences in the loop between the CYS residues in the intrahelical disulfide bonded CXXC motifs ('i' is the N-terminal cysteine).

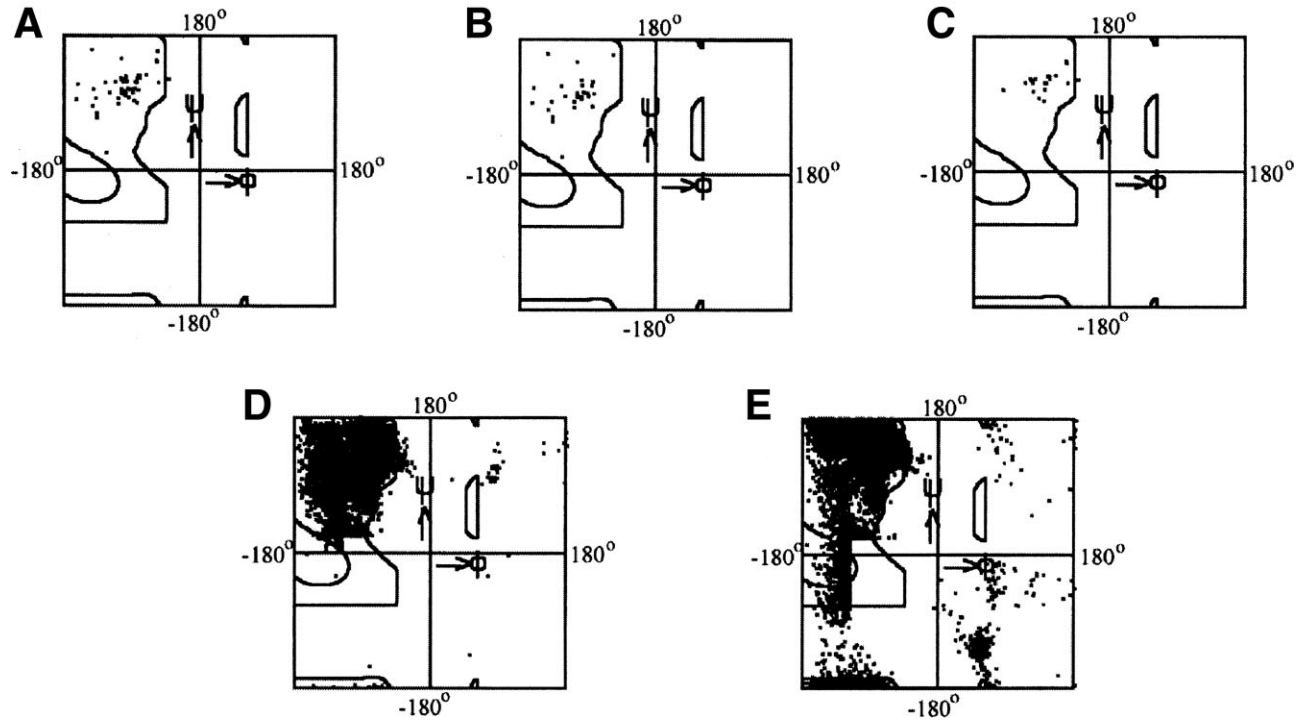

Figure 2

Ramachandran plots for the helix N-Cap residues. **A**: (ϕ, ψ) Values for N-Cap CYS in all intrahelical CXXC motifs predicted by MODIP to form disulfides. **B**: Subset of A which actually form disulfides. **C**: Subset of A which do not form disulfides. **D**: All N-Cap residues in dataset predicted by MODIP to be a potential site for an intrahelical disulfide. **E**: (ϕ, ψ) Values for N-Cap residues predicted by MODIP not to form a disulfide.

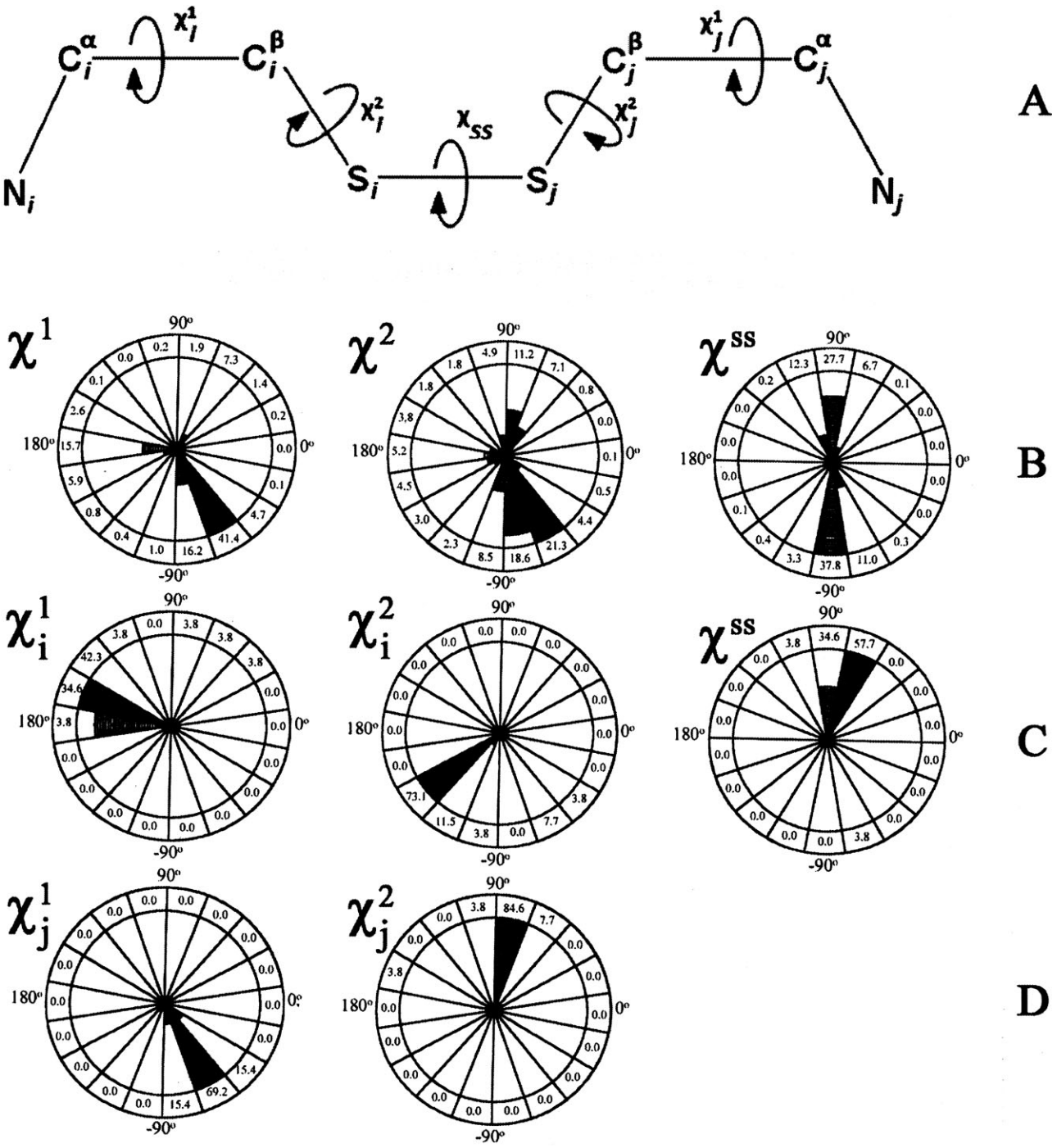

Figure 3

Disulfide torsion angles and their distributions. **A**: Torsion angle definitions. **B–D**: Circular histograms showing the percentage distributions for CYS side chain torsion angles for (B) all naturally-occurring disulfides. C,D: Naturally-occurring intrahelical disulfides. The percentage values for various angle ranges are indicated on the circumference of the histograms.

Fig. 3(C) with 3(B)]. In addition, there are marked differences between χ_i^1 and χ_j^1 as well as χ_i^2 and χ_i^2 where i and j are the N and C terminal CYS residues involved in the disulfide bond [compare Fig. 3(C) with 3(D)].

Lack of conformational change upon reduction of naturally-occurring intrahelical disulfides

To identify intrahelical disulfide containing proteins with structures available in both the oxidized and

Table V
Differences in Torsion Angles of CX_1X_2C Residues in the Oxidized and Reduced Forms

PDB ID Oxidized/ Reduced	Residue No.[a]	Residue	Difference in torsion angle[b] $\Delta\phi$	$\Delta\psi$	$\Delta\chi_1$
1eej/1tjd	98	CYS	−19.1	−6.4	−19.2
1eej/1tjd	99	GLY	4.6	−3.2	–
1eej/1tjd	100	TYR	5.4	−8.4	−0.6
1eej/1tjd	101	CYS	3.3	−2.5	15.5
1fvk/1a2l	30	CYS	−12.2	−11.6	153.7
1fvk/1a2l	31	PRO	6.7	−1.2	−1.1
1fvk/1a2l	32	HIS	−11.7	18.4	−21.9
1fvk/1a2l	33	CYS	−8	−4.1	34.3
1w4v/1w89	31	CYS	12.4	−3.1	21.7
1w4v/1w89	32	GLY	−20.3	12.4	–
1w4v/1w89	33	PRO	−0.3	−10	2.2
1w4v/1w89	34	CYS	4.9	3.5	2.2
2fwg/2fwh	461	CYS	−4.4	−2.4	9.4
2fwg/2fwh	462	VAL	−2	−3.9	92.6
2fwg/2fwh	463	ALA	−3.9	5	–
2fwg/2fwh	464	CYS	−1.5	0	−0.4
1eru/1erv	32	CYS	−11.6	−14.6	5.7
1eru/1erv	33	GLY	−19	10.4	–
1eru/1erv	34	PRO	5.8	−7.4	−11.0
1eru/1erv	35	CYS	−0.3	5.9	16.4

[a]Residue numbering for the CX_1X_2C motifs of the two proteins in each pair is the same. In all cases, the chain ID is "A."
[b]Δ(Torsion angle) = $(\text{Torsion angle})_{\text{oxidized}} - (\text{Torsion angle})_{\text{reduced}}$.

reduced state, each of the proteins from Tables II and III were queried against the PDB database for structures with sequence identities ≥95% and differing in cysteine oxidation state. Five such proteins belonging to the thioredoxin fold were identified. The C^{α} RMSD values between the members of each pair of structures were less than 1 Å. A comparison of main chain torsion angles, (ϕ, ψ) for residues in the helix that carries the CXXC motif revealed very small differences between the two helices indicating that the N-terminal intrahelical disulfide can be accommodated without introducing any drastic changes in the structure (Table V).

MODIP modeled intrahelical disulfides

MODIP is a program that identifies sites for disulfide introduction based on stereochemical criteria.[34,35] A database of 29,597 helical segments was created from the non-redundant dataset as described in the methods section.[31] MODIP was used to identify potential sites for introduction of disulfide forming CYS residues in the helical segments (Table VI). Depending on the geometry of the modeled disulfide, the predicted sites are graded as A, B, and C by MODIP in descending order of stereochemical quality. Consistent with observations of naturally-occurring intrahelical disulfides, all sites identified by MODIP also required the N-terminal residue of the predicted site to have a non-helical conformation. This residue was the N-Cap of the helical segment. The number of residues (loop length) between the two potential cysteines varied from 0–4 with the maximum number of sites identified for a loop length of two. Although there are no A grade disulfides when loop length is zero, there are predicted disulfides with A grade for other loop lengths of 1–4. The above findings are consistent with those for naturally-occurring disulfides. Disulfide formation is not possible in the interior or towards the C-terminus of a helix because the C^{β}–C^{β} distances between residues are incompatible with disulfide formation. The average C^{β}–C^{β} distance for residues i and i+3 within a helix is 5.87 ± 0.55 Å. The favored C^{β}–C^{β} distance for disulfide formation is in the range of 3.4 Å–4.6 Å.[34,35] These results indicate that only the residues at the N-Cap positions of helices have the required stereochemistry to allow intrahelical disulfide formation.

If appropriate backbone torsion angles were the necessary and sufficient conditions for intrahelical disulfide formation, then the non-helical N-Cap of every helix may be a potential site for disulfide introduction depending on the extent of backbone conformational rearrangement that is possible. Since MODIP requires the position of the C^{β} atom for modeling cystines, only non-glycine residues are considered for disulfide introduction. Of the 20,157 helical segments with non-glycine residues at N-Cap and third position, MODIP predicted a disulfide linkage between the N-Cap and third residues of the helical segments for 8779 cases. Helical segments where MODIP failed to predict a disulfide between residues N-Cap and 3 are hereafter referred to as non-MODIP predicted cases. The average (ϕ,ψ) values for the N-Cap residue for MODIP predicted and non-MODIP predicted cases are (−95 ± 31.6°, 118 ± 38.6°) and (−108 + 47°, 154 ± 59.3°), respectively. A comparison of the ϕ–ψ plot for the N-Cap modeled CYS reveals a greater spread of values for non-MODIP predicted cases relative to MODIP predicted cases (see Fig. 2). The cases that were not predicted could be owing to MODIP's modeling of a disulfide on a rigid protein molecule. Hence, energy minimizations were carried out for the *in silico* introduced

Table VI
MODIP Predicted Intrahelical Disulfide Sites in Non-Redundant Dataset

Residue position within helix[a]			Grade[b]		
N-terminal residue	C-terminal residue	No. of Examples	A	B	C
N-Cap	1	16	0	0	16
N-Cap	2	1523	234	168	1121
N-Cap	3	9232	2398	2218	4616
N-Cap	4	162	33	44	85
N-Cap	5	7	2	0	5

[a]The helix is numbered such that residue 1 is the first residue with helical ϕ, ψ and N-Cap is immediately N-terminal to residue 1.
[b]Stereochemical grade of modeled disulfide assigned by MODIP.[34,35]

disulfides for a small subset of the MODIP predicted and non-MODIP predicted cases.

Energy minimization for structures with introduced intrahelical disulfides

Using GROMACS 3.3.3, energy minimizations were carried out in vaccum for six cases of non-MODIP predicted, three cases of MODIP predicted disulfide-bonded, helical segments, and five nSS Trx protein derivatives in which intrahelical disulfides had been introduced *in silico*. As mentioned earlier in this article, nSS Trx is a derivative of *E. coli* thioredoxin in which both CYS in the active site are mutated to SER. The nSS Trx derivatives chosen for the energy minimization studies have also been studied experimentally following mutagenesis of nSS Trx (described in a later section). The coordinates of the starting structures for energy minimization were obtained by replacing the coordinates of the appropriate residues in each PDB file with those of cysteines. The coordinates for the modeled cysteines in proteins were obtained using MODIP for MODIP predicted cases and through the SWISS-PDB mutagenesis tool for non-MODIP predicted cases. In the case of nSS Trx derivatives, the sulfur atoms of CYS 32 and CYS 35 (2trx) were replaced by oxygen atoms in the coordinates file before mutagenesis by SWISS-PDB and energy minimization. Energy minimizations were also carried out for the five pairs of proteins listed in Table V. Prior to minimization of the reduced structure in each pair, a disulfide was introduced corresponding to the oxidized protein structure by the *in silico* mutagenesis tool of SWISS-PDB.

Following minimization, the bond angles at sulfur atoms and the disulfide bond length converged at values that are normally observed in naturally occurring disulfides for all the cases. However, the torsion angles across the disulfide bridge for the energy minimized structures carrying disulfides introduced either with MODIP or SWISS-PDB (MODIP predicted cases, non-MODIP predicted cases, nSS Trx mutants, and four out of five reduced protein structures from Table V) showed large deviations from those observed for naturally-occurring intrahelical disulfides (Table SII, Supporting information). However, the oxidized protein structures (from Table V) on energy minimization converged to structures with torsion angles across disulfide bonds similar to those observed for naturally-occurring intrahelical disulfides. The nomenclature for the disulfide parameters is indicated in Figure 3. The energy minimization studies did not reveal significant differences between MODIP and non-MODIP predicted cases except that the α_R and positive ϕ regions of the plot are incompatible with intrahelical disulfide formation. These studies in conjunction with the main chain dihedral angle analysis described above suggest that disulfides can potentially be formed at virtually all exposed α-helix termini except possibly for those with positive ϕ values. However, one caveat from these studies is that energy minimized predicted structures have very different side chain dihedral angle values from naturally-occurring disulfides. This suggests that the potential energy functions used for disulfide modeling need improvement.

CXXC motifs as an indicator of redox activity/metal binding activity

Most proteins carrying an intrahelical disulfide are either redox active or bind metal ions through the intrahelical CXXC motif. Hence, the presence of an intrahelical CXXC motif indicates potential redox activity or metal binding. In the reduced state, there are several instances of these N-terminal intrahelical CXXC motifs binding metal ions. There are no instances of oxidized CXXC motifs at helical C-termini or inside helices where both CYS are in helical conformations. These intrahelical and C-terminal helical motifs, however, often coordinate metal ions and cofactors such as heme. Non-helical, non-disulfide bridged CXXC motifs also have been found to coordinate metal ions and cofactors in several cases. The secondary structure was predicted for proteins with intrahelical CXXC and non-helical, oxidized CXXC sequences using PSIPRED.[49] PSIPRED predictions with at least two out of the four residues of CXXC sequence being helical were considered as a prediction for an intrahelical CXXC motif. All 14 non-helical CXXC containing proteins (Table IV) were correctly predicted to be non-helical. However, only 79 out of 102 known N-terminal intrahelical CXXC motifs were correctly predicted to be helical. This is consistent with a known prediction accuracy of about 75% for PSIPRED.[49,50] Hence, PSIPRED secondary structure prediction can be used to decide if a CXXC sequence is helical and consequently involved in redox activity or metal binding. Further, if a CXXC motif is predicted to be helical and is known to be disulfide bridged, then the probability of the N-terminal CYS being the N-Cap of the helix is high.

Engineered intrahelical disulfides in thioredoxin

To experimentally validate the above analysis, pairs of cysteines were introduced at multiple locations in α-helices of *E. coli* thioredoxin. Thioredoxin is a 108 amino acid long oxidoreductase.[51] It is a monomeric, highly stable protein containing a core β-sheet flanked by helices. There are four α-helices and a distorted 3_{10} helix. The α-helices designated as α1–α4 (as defined in the PDB-ID 2trx) are from residues 11–17, 32–49, 59–63, and 95–107, respectively. Each of the residues at N-termini have non-helical (ϕ, ψ) values of (-89.1°, -6.8°), (-90.2°, 108.7°), (-79.1°, 118.0°), and (-85.2°, 162.9°), respectively, for residues 11, 32, 59, and 95. The residues

Table VII
Protein Mass (Da) Determined by ESI-MS Before and After Iodoacetamide Labeling for the Confirmation of Disulfide Bond Formation

Protein	Expected mass[a]	Observed mass	Observed mass[b] (Iodoacetamide treated)
nSS Trx	11643.3	11642.3	11643.0
CIDC (59–62)	11605.3	11605.3	11606.5
CIPC (59–62)	11587.4	11588.7	11587.4
CGPC (59–62)	11531.2	11531.5	11534.7
CKGC (95–98)	11632.4	11633.5	11633.8
CGPC (95–98)	11601.3	11601.3	11603.7
CDQC (60–63)	11620.3	11622.0	11736.0[c]
CGQC (96–99)	11606.2	11606.8	11722.1[c]
CDQC (60–63)[d]	11620.3	11619.9	11619.7
CGQC (96–99)[d]	11606.2	11606.8	11606.9

[a]Mass of oxidized protein.
[b]Mass observed after denaturation and incubation with iodoacetamide.
[c]An observed mass increase of 116 Da (with respect to the oxidized protein mass).
[d]Disulfide bond formation was observed after copper-phenanthroline oxidation.

11, 32, 59, and 95 are the N-Cap residues of the helices in thioredoxin. Of the four helices described above, it is already known that α1 is disordered and a disulfide occurs at the N-terminus of α2 in WT Trx. We therefore, attempted to introduce disulfides at the N-terminus of helices α3 and α4, respectively. Since analysis of PDB structures showed that glycine and proline are the preferred amino acids at positions 2 and 3 of the CXXC motif, additional mutants were made incorporating these residues. Finally, to study if disulfides could be formed at non N-Cap helical positions, CYS pairs were also introduced at positions 1,4 in α3 and α4 as well as positions 7,10 in α4. This could theoretically occur if there was a change in the helix start site in the mutant protein. Since our aim was to check if disulfide formation at engineered CXXC sites could be used to probe helix start sites in proteins of unknown structure, it was important to include these positions. Thioredoxin contains two active site cysteines at positions 32 and 35. These were mutated to serines to yield nSS Trx and additional intrahelical pairs of CYS residues were introduced into this mutant background. Following expression and purification from *E. coli*, the redox status of the introduced CYS residues was examined by iodoacetamide labeling. Iodoacetamide will react only with reduced CYS residues, and the resultant acetylated product will result in a mass increase of 57 Da per acetylation. The results are summarized in Table VII. Iodoacetamide labeling studies confirmed that while N-Cap-3 disulfides formed spontaneously, the 1-4 disulfides did not form until subjected to more oxidizing conditions in the presence of 1,10-phenanthroline monohydrate.[42] This shows that the N-Cap-3 positions of a helix are poised to form the disulfide as observed in naturally-occurring intrahelical disulfides. The absence of spontaneous disulfide formation in the 1–4 positions further reiterates the importance of the N-terminal CYS residue to be non-helical for disulfide formation. The mutant with the disulfide in the 7–10 position formed a intermolecular disulfide leading to dimer formation and is described in more detail elsewhere[46] and this mutant was not used further in the present work. Mass spectrometry and SDS-PAGE in the absence of a reducing agent confirmed that engineered cysteines were not involved in the formation of intermolecular disulfides with the exception of the 7–10 mutant (Trx F102C A105C) described above.

Far UV CD spectroscopy of intrahelical disulfide mutants

Purified mutants were characterized by far UV-CD spectroscopy in the wavelength range of 200–250 nm (see Fig. 4). All N-Cap-3 mutants and the 1–4 mutant CDQC (60–63) displayed spectra similar to nSS Trx in both oxidized and reduced states indicating that there were no gross structural changes in the structure of the protein.

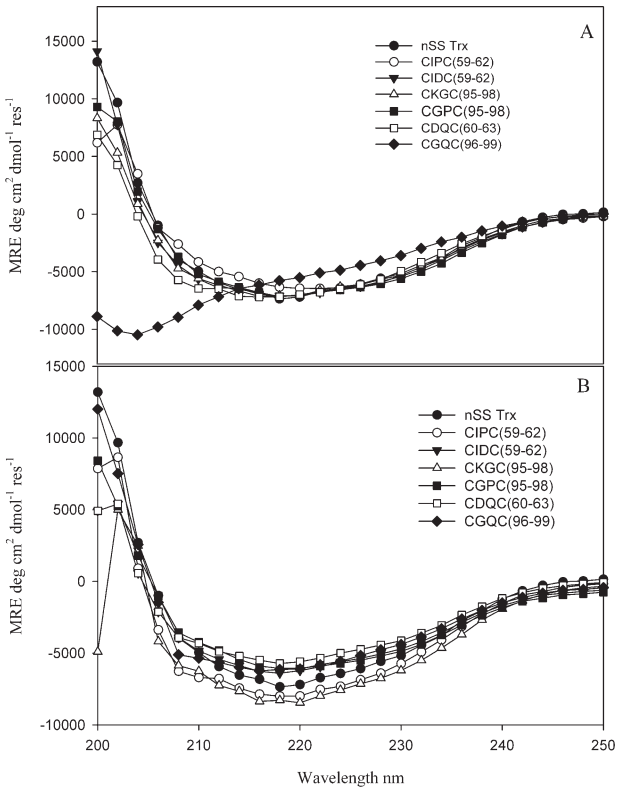

Figure 4
CD spectra of 10 μ*M* nSS Trx and its intrahelical disulfide mutants in (**A**) oxidized and (**B**) reduced states. All measurements were collected with 10 μ*M* protein in CGH 10 buffer at pH 7.0, 298 K. The spectra for the reduced state were collected in the presence of 0.5 m*M* DTT. All proteins have similar secondary structure except for oxidized CQGC (96–99), which appears to be unfolded.

Table VIII
The Stability of nSS Trx and Its Intrahelical Disulfide Mutants Under Oxidized and Reduced Conditions as Measured by Isothermal Denaturation Studies at 25°C, pH 7.0, and DSC Studies

Protein	ΔG° (25°C)[a] (kcal mol^{-1})	C_m (M)[b]	$\Delta\Delta G^\circ$ (25°C)[c] (kcal mol^{-1})	ΔC_m (M)	T_m (°C)[d]	ΔT_m (°C)	$\Delta H^\circ(T_m)$ (kcal mol^{-1})[e]
nSS Trx	5.5 ± 0.3	1.6	—	—	77.4	—	117.0
CIDC (59–62)	4.1 ± 0.1	1.2	−1.4	−0.4	62.0	−15.4	52.9
CIPC (59–62)	3.5 ± 0.1	1.0	−2.0	−0.6	65.6	−11.8	67.1
CGPC (59–62)[f]	—	—	—	—	—	—	—
CKGC (95–98)	3.8 ± 0.1	1.1	−1.5	−0.4	66.1	−11.3	67.4
CGPC (95–98)	4.3 ± 0.1	1.2	−1.2	−0.4	66.2	−11.2	62.0
CDQC (60–63)[g]	1.7 ± 0.1	0.5	−3.8	−1.1	—	—	—
CGQC (96–99)[h]	—	—	—	—	—	—	—
CIDC (59–62)red[i]	4.6 ± 0.1	1.3	−0.9	−0.3	—	—	—
CIPC (59–62)red[i]	3.3 ± 0.4	1.0	−2.2	−0.6	—	—	—
CKGC (95–98)red[i]	4.3 ± 0.9	1.2	−1.2	−0.4	—	—	—
CGPC (95–98)red[i]	3.4 ± 0.1	1.0	−2.1	−0.6	—	—	—
CDQC (60–63)red[i]	3.8 ± 0.1	1.1	−1.7	−0.5	—	—	—
CGQC (96–99)red[i]	2.9 ± 0.1	0.8	−2.6	−0.8	—	—	—

[a]All isothermal melts were fitted with the same m value of –3.4 kcal mol^{-1}.
[b,d,e]Approximate errors are 0.05M for C_m, 1°C for T_m, and 5% for $\Delta H^\circ(T_m)$.
[c]$\Delta\Delta G^\circ = \Delta G^\circ$(mutant)$-\Delta G^\circ$(nSS Trx) where ΔG° is free energy of unfolding.
[f]Protein CGPC (59–62) was unstable and could not be purified.
[g]CDQC (60–63) could not be characterized by DSC since it precipitated during thermal melt.
[h]CGQC (96–99) underwent drastic structural change on formation of disulfide bond.
[i]All melts in the reduced state were carried out in the presence of 20-fold molar excess of protein.

However, the oxidized 1–4 mutant CGQC (96–99) had a drastically altered CD spectrum from that of nSS Trx. The reduction of the disulfide resulted in a spectrum similar to that of nSS Trx indicating that forced intrahelical disulfide formation at a non N-terminal position can disrupt the protein structure. These observations emphasize the importance of the N-terminal position with non-helical (ϕ, ψ) values for intrahelical disulfide formation.

Effect of engineered intrahelical disulfides on protein stability

To examine the effect of the engineered disulfides on protein stability, nSS Trx, and all mutants were characterized by both isothermal chemical denaturation with GdmCl and DSC. All chemical denaturation melts of proteins were fit using the m-value for nSS Trx chemical denaturation (m-value = -3.5 kcal mol^{-1} M^{-1}). Chemical denaturation was reversible for all oxidized mutants with the exception of CGQC (96–99). All mutants except CDQC (60–63) showed reversible thermal melt profiles on DSC. The thermal melt of CDQC (60–63) in the oxidized state could not be studied owing to aggregation of the protein. It was not possible to analyze DSC scans for any of the reduced proteins, carried out in the presence of 1 mM DTT. This could be owing to some irreversible changes such as aggregation, non-2 state melts or proline isomerisations. No stability studies were carried out with the CGQC (96–99) mutant in the oxidized state since it has undergone a drastic structural change.

Formation of engineered intrahelical disulfides led to destabilization of the protein with respect to nSS Trx in all cases. The reduction in stability was confirmed by both, chemical and thermal denaturation studies (Table VIII and Figs. 5 and 6) in the oxidized state. All oxidized mutants had lower values of C_m and T_m relative to nSS Trx. Both chemical and thermal melts were analyzed assuming two-state denaturation. In all cases, the transitions fit well to a two-state model. A more rigorous demonstration would involve monitoring chemical denaturation with an independent probe such as tryptophan fluorescence as was done for WT Trx in the oxidized state.[52] In the case of reduced thioredoxin, the baselines obtained by monitoring fluorescence in a chemical melt have large slopes. Hence, denaturation of reduced thioredoxin is typically monitored by CD as is done in the present work for nSS Trx. Thus, we have not conclusively shown that denaturation of nSS Trx and its derivatives is two-state. However, this does not alter any of the principal conclusions derived from the data in Table VIII, namely that the engineered intrahelical disulfides in nSS Trx are all destabilizing. The extent of destabilization depends on the identity of the residues in the CXXC motif. In addition, forcible formation of a disulfide at non N-terminal helical positions is highly destabilizing.

An analysis of residue preferences at positions 2 and 3 of the CXXC motif had suggested that glycine and proline are the preferred substitutions at these positions. However, incorporation of these preferred amino acids also failed to improve protein stability. In the case of CXXC mutants in the 59–62 region, introduction of

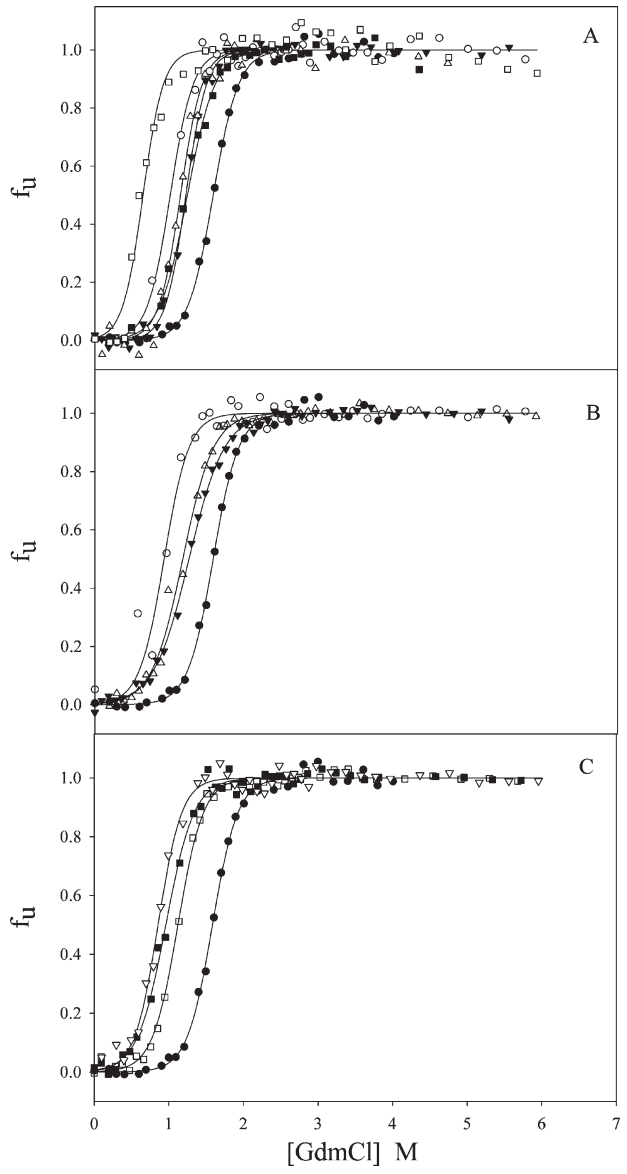

Figure 5

Isothermal chemical denaturation studies on nSS Trx (●) and intrahelical disulfide mutants of nSS Trx [CIPC (59–62)(○), CIDC (59–62)(▼), CKGC (95–98)(Δ), CGPC (95–98)(■), CDQC (60–63)(□), and CGQC (96–99)(▽)] showing the fraction unfolded (f_u) as a function of denaturant concentration for (**A**) oxidized and (**B**,**C**) reduced proteins at pH 7.0, 298 K. Data were fitted to a two state model with globally fit m values of -3.4 kcal mol^{-1} M^{-1} for both oxidized and reduced proteins. Unfolding of the oxidized and reduced proteins was monitored by CD spectroscopy at 298 K at 222 nm. All proteins have reduced stability relative to nSS Trx in both oxidized and reduced states.

glycine, proline at positions 60–61 or proline alone at position 61 resulted in significant destabilization of the protein relative to nSS Trx. In contrast, for the 95–98 mutants, the CGPC mutant was less destabilized than the CKGC mutant, which retained the WT residues at positions 1,2 of the CXXC motif (Table VIII). A comparison of mutant stabilities in the oxidized and reduced states shows that all N-Cap-3 mutants had similar stabilities in both the reduced and oxidized states. In one case CGPC (95–98), the mutant was marginally more stable in the oxidized state ($\Delta C_m = 0.2M$).

In the case of the two 1–4 mutants studied here, both are significantly stabilized in the reduced state relative to the oxidized state. The CDQC (60–63) mutant is conformationally deformed in the oxidized state. The CDQC (60–63) reduced mutant is stabilized by about 2.1 kcal mol^{-1} with respect to the oxidized state. The destabilizing nature of these disulfides could be owing to the non N-terminal nature of the CXXC motif.

For WT Trx, reduction of the disulfide leads to a large decrease in stability ($\Delta\Delta G^\circ = 2.4$ kcal mol^{-1} at 298 K).[44] Consistent with this, nSS Trx also has lower stability than oxidized WT Trx, but similar stability to reduced Trx. However, the data in Table VII clearly demonstrate that introduction of disulfides at N-termini of other helices in thioredoxin does not lead to similar improvements in protein stability. The two contributing factors to the above observation are discussed below.

First, formation of disulfide involves a degree of protein conformational rearrangement, which may have an energetic cost. This is especially true for disulfides at buried positions. Second, the residues that were mutated to cysteines may have been involved in various favorable interactions. In the present case, for example the side chains of N59, S95, and Q98 all form hydrogen bonds to the amide nitrogen atoms of residues D61, Q98, and S95, respectively, in the WT Trx protein. All of these hydrogen bonds would be lost upon mutation of residues N59, S95, and Q98 to CYS.

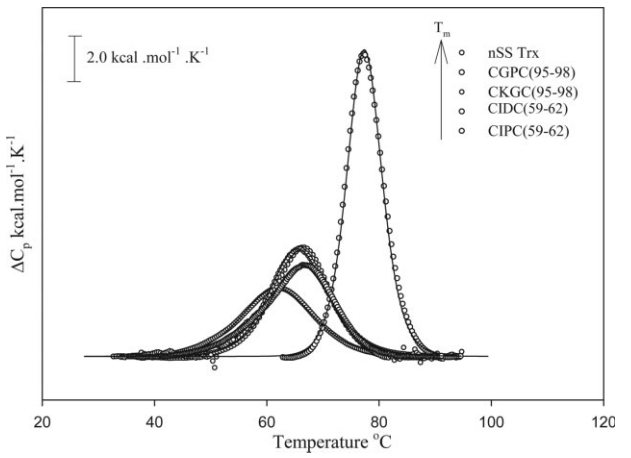

Figure 6

DSC studies on nSS Trx and its intrahelical disulfide mutants in the oxidized state at pH 7.0. DSC scans of baseline subtracted excess heat capacity as a function of temperature for the oxidized proteins show that all mutants are thermally less stable than nSS Trx. Protein identities are indicated in increasing order of T_m in the inset. Raw data are shown as open circles and the fitted data are shown as lines in all cases.

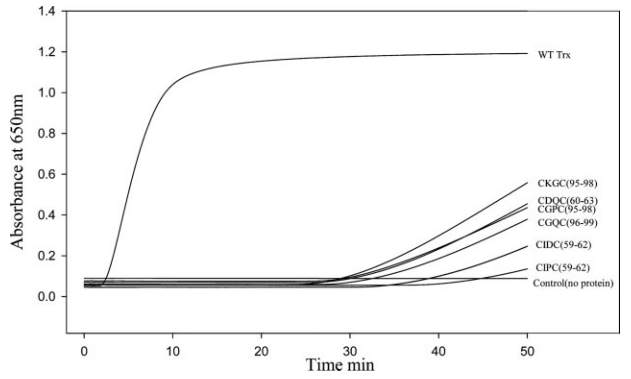

Figure 7

Thioredoxin-catalyzed reduction of insulin. Insulin aggregation following reduction was monitored by the increase in light scattering at 650 nm. Assay conditions were 0.1*M* phosphate buffer, 2 m*M* EDTA, 0.13 m*M* porcine insulin, 0.33 m*M* DTT, and 10 μ*M* protein. Protein identitites are adjacent to each trace. Incubation mixture without protein served as control.

Activity of thioredoxin with engineered CXXC motifs

Intrahelical CXXC motifs are a characteristic feature of several oxidoreductases and are responsible for their redox activity. Hence, it was of interest to examine if the thioredoxin mutants carrying engineered intrahelical CXXC motifs can display thioredoxin like redox activity. Thioredoxin catalyzes the reduction of insulin by DTT. Upon reduction, the insulin B chain aggregates and this can be monitored by light scattering at 650 nm. All the CXXC mutants (N-Cap-3 and 1–4 disulfide mutants) show varying degrees of activity (see Fig. 7). However, all mutants have activities lower than WT Trx. This reduced level of activity is expected since the active site helix of thioredoxin is evolutionarily optimized for redox activity. Further, recent studies have shown that the conserved cis-proline loop in thioredoxin plays an important role in its substrate specificity and activity.[53] Surprisingly, there was no clear correlation between the position of the CXXC motif in the helix with respect to the level of activity. Expectedly, nSS Trx that lacks cysteine residues showed no activity. Control proteins MBP 230C 30C and Trx 20C 73C contain two CYS residues such that the cysteine C^{α} atoms are separated by 35.3 Å and 27.3 Å apart in the structures of WT MBP and WT Trx, respectively. They also showed no activity (Fig. S1, Supporting information). This shows that merely having two free cysteine residues in a protein does not confer redox activity.

CONCLUSIONS AND FUTURE DIRECTIONS

The present study clearly demonstrates that intrahelical disulfides can only occur at the N-terminus of an α-helix and that the N-terminal CYS residue must adopt a non-helical backbone conformation. Experimental studies with *E. coli* thioredoxin showed that disulfides could be introduced at the N-termini of two different helices. The engineered disulfides did not increase protein stability, probably because the favorable interactions involving the WT residues are lost upon substitution with cysteines. However, the mutants showed redox activity in an insulin reduction assay. It is, therefore, possible to engineer such intrahelical disulfides and such engineered intrahelical disulfides can confer redox activity on a protein. In cases where the three dimensional structure is unknown, such engineered disulfides might also be used to experimentally determine the location of helix start sites. CYS can potentially be replaced by selenocysteine. This suggests the possibility of using engineered diselenide derivatives to obtain phase information by anomalous scattering. In contrast to the commonly used selenomethionine derivatives, such diselenides may be better ordered.

ACKNOWLEDGMENTS

C.R. is a senior Scientist of the Indian National Science Academy.

REFERENCES

1. Harrison PM, Sternberg MJ. Analysis and classification of disulphide connectivity in proteins. The entropic effect of cross-linkage. J Mol Biol 1994;244:448–463.
2. Creighton TE. Disulphide bonds and protein stability. Bioessays 1988;8:57–63.
3. Ladbury JE, Kishore N, Hellinga HW, Wynn R, Sturtevant JM. Thermodynamic effects of reduction of the active-site disulfide of Escherichia coli thioredoxin explored by differential scanning calorimetry. Biochemistry 1994;33:3688–3692.
4. Pace CN, Grimsley GR, Thomson JA, Barnett BJ. Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. J Biol Chem 1988;263:11820–11825.
5. Chakraborty K, Thakurela S, Prajapati RS, Indu S, Ali PS, Ramakrishnan C, Varadarajan R. Protein stabilization by introduction of cross-strand disulfides. Biochemistry 2005;44:14638–14646.
6. Haworth NL, Feng LL, Wouters MA. High torsional energy disulfides: relationship between cross-strand disulfides and right-handed staples. J Bioinform Comput Biol 2006;4:155–168.
7. Ye X, O'Neil PK, Foster AN, Gajda MJ, Kosinski J, Kurowski MA, Bujnicki JM, Friedman AM, Bailey-Kellogg C. Probabilistic cross-link analysis and experiment planning for high-throughput elucidation of protein structure. Protein Sci 2004;13:3298–3313.
8. Chivers PT, Laboissiere MC, Raines RT. The CXXC motif: imperatives for the formation of native disulfide bonds in the cell. EMBO J 1996;15:2659–2667.
9. Chivers PT, Prehoda KE, Raines RT. The CXXC motif: a rheostat in the active site. Biochemistry 1997;36:4061–4066.
10. Buchanan BB, Schurmann P, Decottignies P, Lozano RM. Thioredoxin: a multifunctional regulatory protein with a bright future in technology and medicine. Arch Biochem Biophys 1994;314:257–260.
11. Holmgren A. Thioredoxin. Annu Rev Biochem 1985;54:237–271.
12. Laboissiere MC, Sturley SL, Raines RT. The essential function of protein-disulfide isomerase is to unscramble non-native disulfide bonds. J Biol Chem 1995;270:28006–28009.

13. Walker KW, Lyles MM, Gilbert HF. Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. Biochemistry 1996;35:1972–1980.
14. Ziegler DM. Role of reversible oxidation-reduction of enzyme thiols-disulfides in metabolic regulation. Annu Rev Biochem 1985;54: 305–329.
15. Grauschopf U, Winther JR, Korber P, Zander T, Dallinger P, Bardwell JC. Why is DsbA such an oxidizing disulfide catalyst? Cell 1995;83:947–955.
16. Moore EC, Reichard P, Thelander L. Enzymatic synthesis of deoxyribonucleotides. Part V. Purification and properties of thioredoxin reductase from E. Coli. J Biol Chem 1964;239:3445–3452.
17. Wunderlich M, Glockshuber R. Redox properties of protein disulfide isomerase (DsbA) from Escherichia coli. Protein Sci 1993;2: 717–726.
18. Krause G, Lundstrom J, Barea JL, Pueyo de la Cuesta C, Holmgren A. Mimicking the active site of protein disulfide-isomerase by substitution of proline 34 in Escherichia coli thioredoxin. J Biol Chem 1991;266:9494–9500.
19. Joelson T, Sjoberg BM, Eklund H. Modifications of the active center of T4 thioredoxin by site-directed mutagenesis. J Biol Chem 1990; 265:3183–3188.
20. Lewin A, Crow A, Hodson CT, Hederstedt L, Le Brun NE. Effects of substitutions in the CXXC active-site motif of the extracytoplasmic thioredoxin ResA. Biochem J 2008;414:81–91.
21. Huber-Wunderlich M, Glockshuber R. A single dipeptide sequence modulates the redox properties of a whole enzyme family. Fold Des 1998;3:161–171.
22. Mossner E, Huber-Wunderlich M, Glockshuber R. Characterization of Escherichia coli thioredoxin variants mimicking the active-sites of other thiol/disulfide oxidoreductases. Protein Sci 1998;7:1233–1244.
23. Quan S, Schneider I, Pan J, Von Hacht A, Bardwell JC. The CXXC motif is more than a redox rheostat. J Biol Chem 2007;282:28823–28833.
24. Pan JL, Bardwell JC. The origami of thioredoxin-like folds. Protein Sci 2006;15:2217–2227.
25. Fomenko DE, Gladyshev VN. Identity and functions of CxxC-derived motifs. Biochemistry 2003;42:11214–11225.
26. Iqbalsyah TM, Moutevelis E, Warwicker J, Errington N, Doig AJ. The CXXC motif at the N terminus of an alpha-helical peptide. Protein Sci 2006;15:1945–1950.
27. Woycechowsky KJ, Raines RT. The CXC motif: a functional mimic of protein disulfide isomerase. Biochemistry 2003;42:5387–5394.
28. Park C, Raines RT. Adjacent cysteine residues as a redox switch. Protein Eng 2001;14:939–942.
29. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. Proteins 1995;23:566–579.
30. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
31. Ramakrishnan C, Srinivasan N. Glycyl residues in proteins and peptides: an analysis. Current Sci 1990;59:851–861.
32. Doig AJ, MacArthur MW, Stapley BJ, Thornton JM. Structures of N-termini of helices in proteins. Protein Sci 1997;6:147–155.
33. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of alpha helices. Science 1988;240:1648–1652.
34. Sowdhamini R, Srinivasan N, Shoichet B, Santi DV, Ramakrishnan C, Balaram P. Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis. Protein Eng 1989;3:95–103.
35. Dani VS, Ramakrishnan C, Varadarajan R. MODIP revisited: re-evaluation and refinement of an automated procedure for modeling of disulfide bonds in proteins. Protein Eng 2003;16:187–193.
36. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.
37. Lindahl EHB, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 2001;7:306–317.
38. Chakrabarti A, Srivastava S, Swaminathan CP, Surolia A, Varadarajan R. Thermodynamics of replacing an alpha-helical Pro residue in the P40S mutant of Escherichia coli thioredoxin. Protein Sci 1999;8: 2455–2459.
39. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 1989;77:51–59.
40. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci 1995;4:2411–2423.
41. Creighton TE. Counting integral numbers of amino acid residues per polypeptide chain. Nature 1980;284:487–489.
42. Careaga CL, Falke JJ. Thermal motions of surface alpha-helices in the D-galactose chemosensory receptor. Detection by disulfide trapping. J Mol Biol 1992;226:1219–1235.
43. Chen GC, Yang JT. Two-point calibration of circular dichrometer with D-10-camphorsulfonic acid. Anal Lett 1977;10:1195–1207.
44. Kelley RF, Shalongo W, Jagannadham MV, Stellwagen E. Equilibrium and kinetic measurements of the conformational transition of reduced thioredoxin. Biochemistry 1987;26:1406–1411.
45. Prajapati RS, Lingaraju GM, Bacchawat K, Surolia A, Varadarajan R. Thermodynamic effects of replacements of Pro residues in helix interiors of maltose-binding protein. Proteins 2003;53: 863–871.
46. Das M, Kobayashi M, Yamada Y, Sreeramulu S, Ramakrishnan C, Wakatsuki S, Kato R, Varadarajan R. Design of disulfide-linked thioredoxin dimers and multimers through analysis of crystal contacts. J Mol Biol 2007;372:1278–1292.
47. Holmgren A. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. J Biol Chem 1979; 254:9627–9632.
48. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
49. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
50. Bajaj K, Madhusudhan MS, Adkar BV, Chakrabarti P, Ramakrishnan C, Sali A, Varadarajan R. Stereochemical criteria for prediction of the effects of proline mutations on protein stability. PLoS Comput Biol 2007;3:e241.
51. Katti SK, LeMaster DM, Eklund H. Crystal structure of thioredoxin from Escherichia coli at 1.68 A resolution. J Mol Biol 1990;212: 167–184.
52. Kelley RF, Stellwagen E. Conformational transitions of thioredoxin in guanidine hydrochloride. Biochemistry 1984;23:5095–5102.
53. Ren G, Stephan D, Xu Z, Zheng Y, Tang D, Harrison RS, Kurz M, Jarrott R, Shouldice SR, Hiniker A, Martin JL, Heras B, Bardwell JC. Properties of the thioredoxin fold superfamily are modulated by a single amino acid residue. J Biol Chem 2009; 284:10150–10159.