Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 15, Issue Number 6, (1998)

Different Types of Interactions Involving Cysteine Sulfhydryl Group in Proteins

http://www.albany.edu/chemistry/sarma/jbsd.html

Debnath Pal and Pinak Chakrabarti*
Department of Biochemistry,
Bose Institute,
P-1/12, CIT Scheme VIIM,
Calcutta 700 054, India

Abstract

Various types of interactions involving the sulfhydryl group of free cysteine residues have been analyzed using known protein structures. In a hydrogen bond the -SH group is more amenable to donating its proton to a carbonyl group, rather than acting as a proton acceptor. It rarely interacts with a carboxylate group, and is a poor ligand to bind an anionic substrate. It is quite prone to make contacts that are definitely non-hydrogen bond type. In the S···C=O interaction the S atom is placed on the face of an amide group (mostly from the main-chain, but there are cases from the side-chain also) close to the C atom. Cases of S···N interaction, where the S atom is on top of the N atom of another residue (both main-, as well as side-chains, including the guanidinium group) are also observed. A considerable number of Cys residues have aromatic residues as neighbors, and here too, the preferred mode of interaction is along the face. The intra-residue S···C=O interaction constrains the main-chain and side-chain torsion angles (ψ and χ_1), whereas the inter-residue interactions are non-local and stabilize the tertiary structure. The S···C=O interaction may have a role in lowering the pK_a values of the Cys residues in enzyme active sites.

Introduction

The structure and function of a protein depends on the interplay of various types of interactions (hydrogen bonding, electrostatic and van der Waals), which, even if they are weak individually, may assume a significant role collectively. (1,2) Interestingly enough, in all these nonbonded interactions, proton (attached to O, N or C) appears to be the common feature making the atoms involved have a fairly rigid stereochemistry. Only recently another type of interaction has been identified which is directional, but does not need any proton to achieve the regular arrangement of the atoms. (3) This involves the S atom in the cysteine side-chain (bound to metal ions) which is found to be positioned over a peptide plane, close to the carbonyl carbon atom (which, in the majority of the cases, belong to the same Cys residue). This results in a delocalisation of electronic charge from the S atom (and by extension, from the cation) all the way to the carbonyl oxygen atom. Besides providing stability, this may favor a lower oxidation state for the metal center, a role that has so far been ascribed to hydrogen bonding only. (4) A putative evidence for the stability that accrues from the Cys ligation is provided by the site-directed mutagenesis experiment on *Azotobacter vinelandii* ferredoxin I - conversion of the ligand Cys20 in the wild type protein to Ser results in a rearrangement of the structure so that Cys24 (and not Ser20) becomes a ligand to the metal cluster. (5) However, there are cases where a Cys has been replaced by Ser without jeopardizing the ligand-binding capability of the site, as in rubredoxin (6) and high potential iron-sulfur protein, (7) but interestingly, these mutations have resulted in the downshift of the redox potential by as much as 200 mV, suggesting that the electron delocalisation due to the S···C=O interaction, makes the metal center more reducible.

*Author to whom correspondence should be addressed. Fax: 091-33-334-3886; E-mail: pinak@boseinst.ernet.in

Given this background it was quite natural to ask if this interaction is also prevalent in free Cys residues (with a partial negative charge residing on the sulfur atom), (8) especially since some of these occur in enzyme active sites where, during the reaction, the -SH group gets deprotonated to the thiolate anion (as in the metal center). However, as the sulfhydryl group can, in principle, act both as a hydrogen bond donor, as well as an acceptor, (9) any attempt to identify a new interaction that does not include proton must first show that the -SH proton is properly engaged in a hydrogen bond interaction. Consequently, although there have been studies of hydrogen bonds involving S atoms in proteins, (9-12) we have also included a similar analysis as part of our attempt to elucidate the various types of interactions a Cys sulfhydryl group can participate in, in particular if these are found on top of the amide group (both main-chain, as well as side-chain), and by analogy, above other planar moieties (guanidinium, aromatic) that are found in protein structures. Moreover, we pay attention to the side-chain conformation and the secondary structural states of the residues in the light of the interactions the S atom is engaged in.

Methods

Atomic coordinates were used from the Brookhaven Protein Data Bank (PDB) (13) for structures which are distinct (14) and for which X-ray diffraction data are available at a resolution better than or equal to 2.0Å. For homo-oligomeric proteins data for only one subunit had been used. All atoms (apart from C_α and C_β of the same residue) at a distance of 3.8Å from the Cys S_γ atom were identified (no crystal symmetry-related contact was included). To take into account the larger radius of the sulfur atom the cut-off distance was slightly longer than 3.5Å that is normally used to investigate hydrogen bond interactions in protein structures. Only the Cys residues with the side-chain not disordered or oxidized, not involved in any disulfide (or other covalent) linkage or bound to cations were considered. If the sulfur atom or its partner had an occupancy less than 1.0, or thermal *B*-factor greater than 30Å^2, the interaction was excluded.

Unlike the case of metal-bound Cys residues, (3) for which the closest contact between the S atom and the face of the peptide group involved the C atom of the latter, a number of free Cys S also showed the shortest distance from the N atom (both main-chain as well as side-chain). As a result, we distinguished two types of interactions, S···C (used synonymously as S···C=O) and S···N. In the context of an amide group, an interaction is classified as S···C or S···N depending on which distance is shorter, (even though both may be within the outer limit); moreover, the angle θ (Figure 1) involving the particular atom should be less than 40° (interaction with the face of the peptide plane).

In the event of the S atom being within 3.8Å of both the atoms of a carbonyl group, if it is eligible to be taken as a S···C interaction (shorter distance and $\theta < 40°$) it is not considered to have a hydrogen bond interaction with the oxygen atom. Whenever a >NH group is found within the limiting distance, the θ value is used to see if it is a S···N interaction ($\theta < 40°$) or hydrogen bond (otherwise). When more than one atom of a chemical entity (like an aromatic ring) is within the limiting dis-

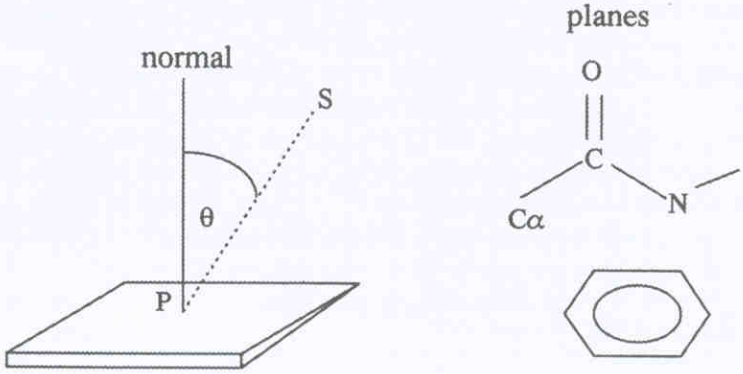


Figure 1: The polar angle θ specifying the direction of S···P relative to the normal to the plane (amide, aromatic or guanidinium) passing through the partner atom, P.

tance, the atom with the shortest distance is assumed to be the partner atom. For an aromatic partner, the S atom is assumed to be on the face of the ring if $\theta < 45°$, and on the edge if it is $> 45°$.

The secondary structural features of the residues were determined using the program DSSP (15), and the molecular plots were made using MOLSCRIPT. (16) Semi-empirical energy calculations (MNDO/3) were performed using the program INSIGHT (Biosym/MSI, San Diego, 1995).

The PDB files (with the number of Cys residues satisfying our screening criteria) are: 1ABK(3), 1ADS(7), 1BAB(3), 1BGC(1), 1BTC(2), 1CAJ(1), 1CMB(1), 1COB(1), 1CPC(2), 1FBA(4), 1FDD(1), 1GKY(1), 1GLT(4), 1GOX(1), 1GPB(8), 1HIV(1), 1HLE(1), 1OFV(1), 1PDA(3), 1PHB(7), 1RRO(1), 1SHA(3), 1TFG(1), 1TRB(3), 1UTG(2), 1YCC(1), 2BOP(3), 2CPL(4), 2CTS(4), 2CYP(1), 2END(1), 2HAD(4), 2HPD(2), 2MHR(2), 2MNR(2), 2PIA(3), 2RN2(3), 3CLA(4), 3COX(1), 3GRS(6), 4ENL(1), 4FXN(3), 4GCR(4), 4INS(2), 5P21(3), 7AAT(5), 8ABP(1), 8ACN(9), 9LDT(5).

Results

(a) *Hydrogen Bond Interactions*. Out of 137 Cys residues used in this analysis 39 residues (28%) have no hydrogen bond partner (Figure 2). 70% of the residues form hydrogen bonds to carbonyl groups (the S···O distance being 3.5(±2)Å, and the C_β-S_γ···O angle, 96(±24)°, Table I). In such interactions the -SH group can be expected to donate protons to the acceptor carbonyl groups. Although the sulfhydryl group was assumed to be an acceptor in one study, (11) another (10) found it to be less of an acceptor. The latter observation is also supported by our analysis. However, it is to be noted that even as a donor its interaction is mostly with the carbonyl group, but rarely with a carboxylate side-chain.

Gregoret and co-wokers (10) did not find any angular preferences for cysteine-S_γ···donor/acceptor pairs, but Ippolito and co-workers (9) found a slight preference for values near 60° of χ_2(HB), which specifies the position of the partner relative to the side-chain (torsion angle C_α-C_β-S_γ···hydrogen bond partner). Our data (Figure 3) show a clear peak at ~90°, and the points are contributed mostly by Cys residues interacting with CO of residues four residues behind. There is also a shallow peak at about -70°. This behavior is quite akin to what was observed for metal-binding by Cys residues, (17) except that the peak at 180° is almost non-existent here. When the partner is a side-chain atom or a water molecule there is no preferential value.

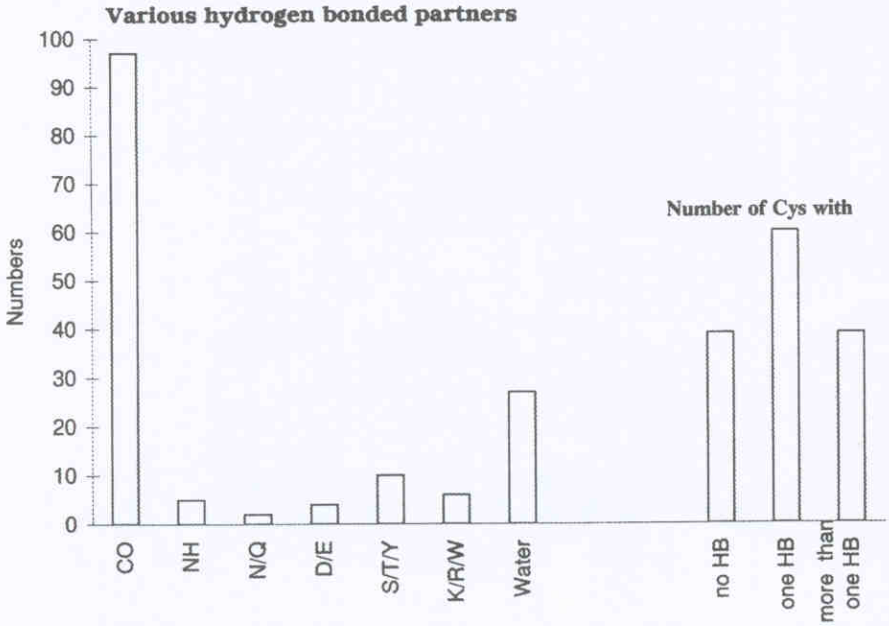


Figure 2: Hydrogen bonding features of 137 Cys residues. The three bars on the right indicate the number of residues with 0, 1 and >1 interactions. Of the total number of hydrogen bond interactions those involving main-chain atoms are represented by the left two bars, and the ones with side-chain atoms (one-letter amino acid codes are used) are given next.

Table I

Geometry (average values, with standard deviations) of various types of interactions involving the cysteine sulfur atom.

Type	Distance (Å)	Angle (°)
S···O=C (hydrogen bonding)	3.5(2)	C_β-S···O, 96(24)
S···C (intra-residue)	3.2(1)	S···C=O, 102(11)
S···C (inter-residue)	3.6(1)	S···C=O, 85(7)
S···N (inter-residue)	3.5(1)	-
S···aromatic (face)	3.6(2)	θ, 19(11)
S···aromatic (edge)*	3.6(1)	θ, 71(13)

*Considering the aromatic proton, the interaction is of the type S···H-C.

(b) *Intra-residue S···C Interactions*. As can be seen from the Newman projection diagrams (Figure 4(a)), when the side-chain conformation is in the *t* (torsion angle, $\chi_1 \approx 180°$) and g^- ($\chi_1 \approx 60°$) states, S is spatially close to CO of the same Cys residue. Most of these show S···C interaction. As the -SH group is hydrogen bonded to an acceptor for most of the Cys residues, it can be surmised that the proton is not available to intervene in this interaction. In the g^+ state ($\chi_1 \approx 300°$), S and C atoms are *trans* to each other, and there can not be any S···C interaction. Although in this conformation S is close to the N atom, the angle θ calculated with N as the origin (Figure 1) is unacceptable (>40°) for 60% of the cases. Consequently, there appears to be no intra-residue S···N interaction comparable to the S···C interaction. This is also borne by the fact that in the g^- state the S atom, in spite of being in between N and the CO group, is attracted more towards the latter.

In order to get an estimate of the intra-residue S···C interaction, energy calculations were performed at the MNDO/3 level for a Cys-containing model peptide structure (Figure 4(b)) at different positions of χ_1 torsion angle (the torsion, C_α-C_β-S_γ-H was kept fixed at 180°, so that the proton always faces away from the main-chain atoms). The same calculations were repeated for another molecule in which the Cys -SH is replaced by a methyl group. For this unnatural amino acid residue where there can not be any S···C interaction, the energy, as expected on steric grounds, (18) is highest for the g^- state. However, for the Cys-containing molecule, the ener-

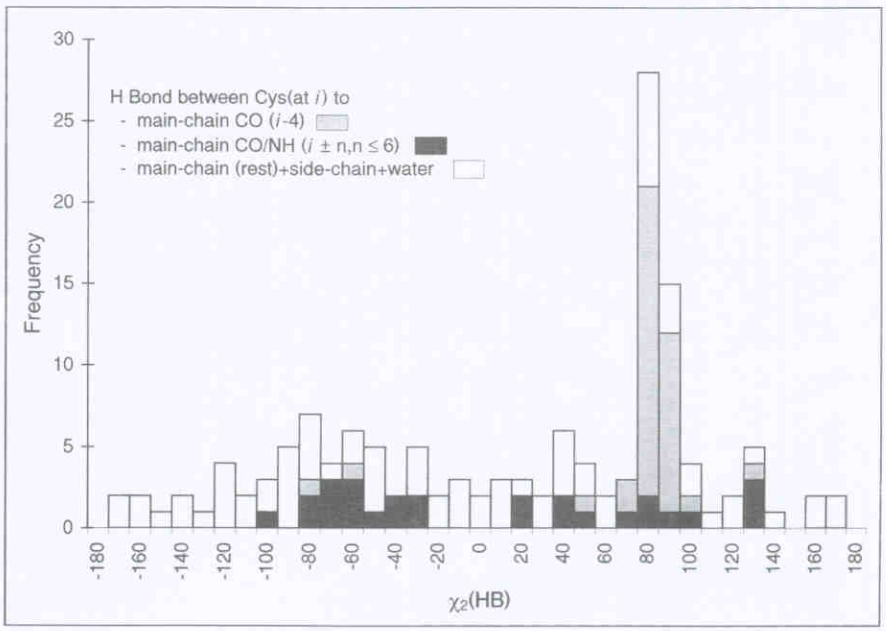
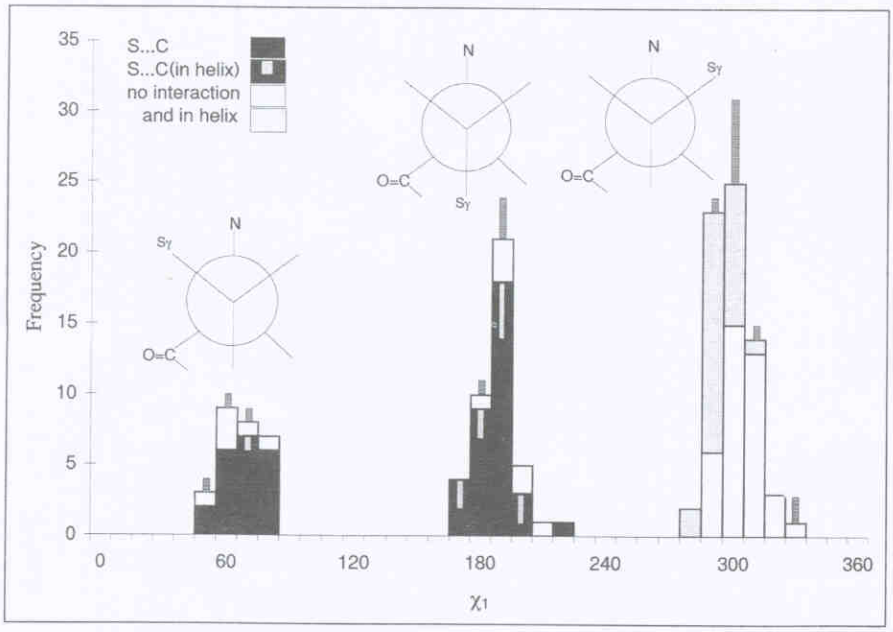
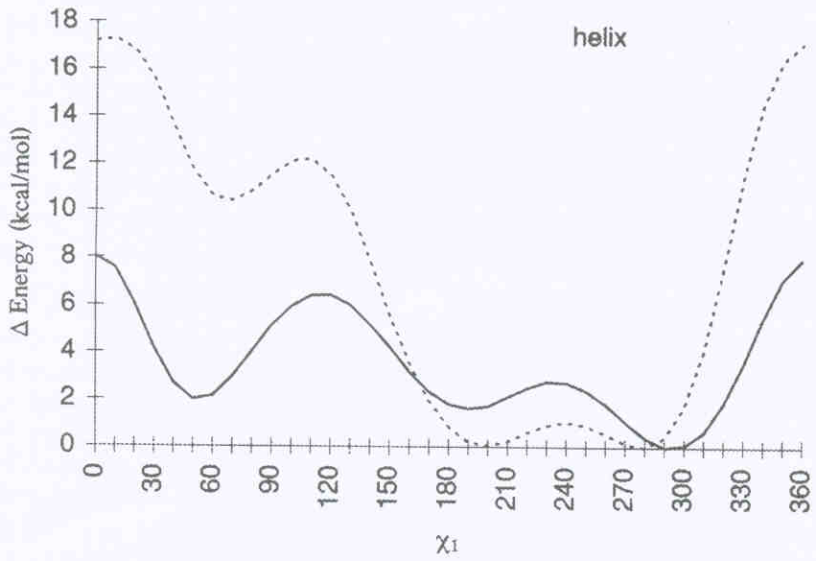


Figure 3: Histogram of the torsion angle, χ_2(HB) (°) (corresponding to C_α-C_β-S_γ···hydrogen bond partner). The codes for different shadings are given in the inset.



4b



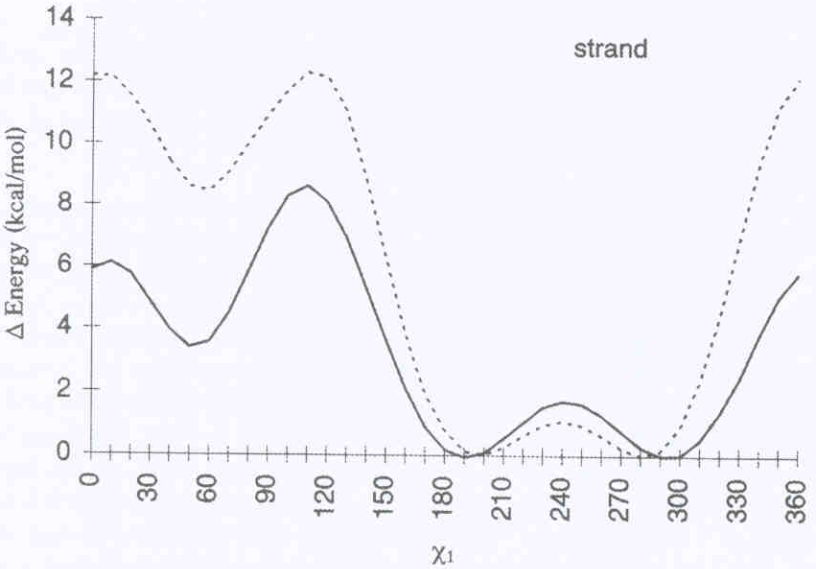


Figure 4: Geometry and energetics of intra-residue S···C interaction. a) Histogram showing the distribution of the torsion angle, χ_1 (⁰) (corresponding to N-C_α-C_β–S_γ). Two types of residues are identified depending on whether or not they show the S···C interaction; helical residues within these two classes are distinguished further. Cases of inter-residue S···C/N interaction are indicated by protrusion above the bars. Newman projection diagrams (looking down the C_β-C_α bond) for the χ_1 angles of 60, 180,and 300⁰ have been shown. (b) Variation of the energy with the Cys χ_1 torsion angle in CH_3-CO-Cys-NH-CH_3 (solid line), having helical (ϕ = -65⁰, ψ = -43⁰) and β-strand (ϕ = -102⁰, ψ = 123⁰) conformations. The energy values plotted are relative to the value of the g^+ state ($E(\chi_1)$ - $E(g^+)$). Energy values for a similar molecule (the Cys -SH replaced by a -CH_3 group) (dashed line) for the same two main-chain conformations are also shown.

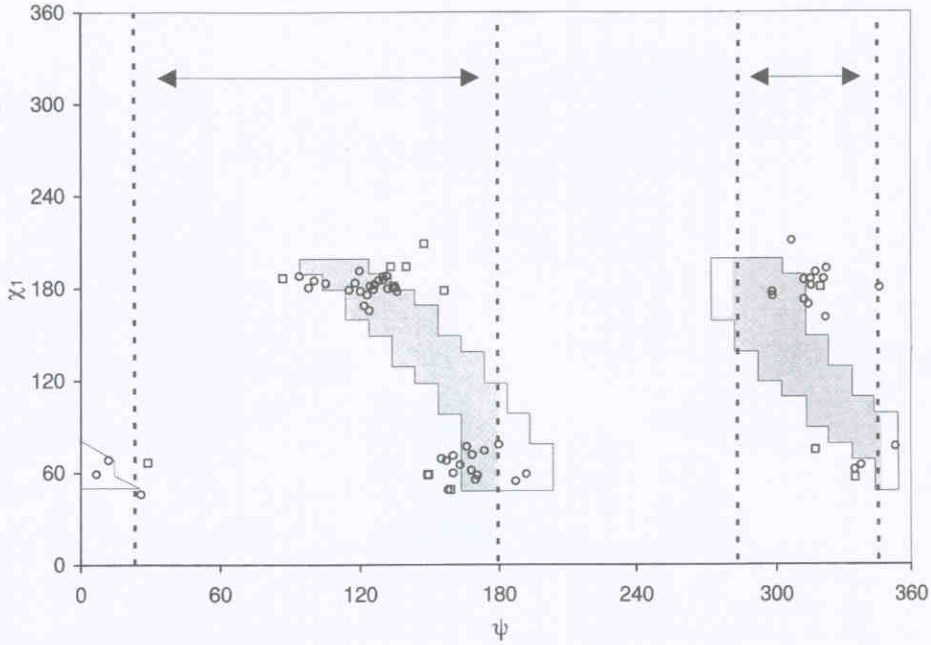


Figure 4: (c) χ_1 (⁰) vs ψ (⁰) plot for residues having χ_1 around 60 and 180⁰ (there can not be any intra residue S···C interaction when $\chi_1 \approx 300^0$). Circles (and squares) represent cases having (and not having) intra-residue S···C interaction. Regions in the plot where the S···C interaction is feasible (the distance S···C < 3.4Å, the angle S···C=O in the range 90 to 130⁰, and $\theta \leq 40^0$; calcuations were done at grid points separated by 10⁰) are delineated. Two sets of broken vertical lines indicate the ψ ranges that are allowed in the Ramachandran plot.

gy for this state is lowered by more than 5 kcal/mol. Both the molecules, built with two sets of ϕ,ψ values, behave similarly, and this can be taken to represent the stabilization energy from the S···C interaction. It is to be noted that for Cys residues bound to metal ions, the g^- state is populated more abundantly than for any other residue, (3,17) pointing to the favorable contribution from the S···C interaction. For free Cys residues, however, g^+ is the most occupied state (Figure 4(a)), but a significant number of such residues are located in helices, where, as already documented, (10,19) there is the possibility of a hydrogen bond between the -SH group and a main-chain oxygen atom in the previous turn of the helix.

It is the combination of the χ_1 and the ψ values that determine the position of the S atom relative to the CO group. (3) Only in a limited range of these two angles will the S atom be close to the top of the C atom. Boundaries surrounding two such regions (S···C < 3.4Å, ∠ S···C=O in the range 90° to 130°, and $\theta \leq 40°$) are drawn in Figure 4(c). Within these boundaries the region that also has allowed values of ψ (in the Ramachandran plot) is shaded, and should contain the points showing the S···C interaction, but exclude the points without this interaction. This is broadly what is observed - some discrepancies arise due to the fact that the boundary has been drawn using a model having standard dimensions, which many of the X-ray structures may lack.

The sulfur atoms showing the S···C interaction (Figure 5(a)) have a rather rigid geometry given by the S···C distance and the S···C=O angle (Table I). Moreover, the torsion angle χ_2(C) that specifies the position of the carbon atom relative to the side-chain of the Cys residue is ±(20 to 40°) (Figure 6).

(c) *Inter-residue S···C and S···N Interactions*. Table II and Figure 5 provide the examples of S···C and S···N interactions (distances are to be found in Table I) where the two atoms belong to different residues. Because of the inaccuracy associated with the X-ray coordinates it may not always be possible to say with certainty which of the two bonded atoms, C or N, in a peptide group is really in contact with sulfur. Although a comparative analysis of the stabilization due to the S···C and S···N interactions may be contentious, it needs to be pointed out that S interacting with N from the top of an amide group is known in literature. (20) Most of the Cys residues are involved in an interaction with a residue remote in sequence. When the partner is a helical residue it is usually located in the first or the last turn of the helix. In three such cases the Cys is also from a helix, so that the interaction stabilizes helix-helix contact. Thus a Cys residue, through the intra-helical (*i* to *i*-4 carbonyl group) hydrogen bonding, can not only sta-

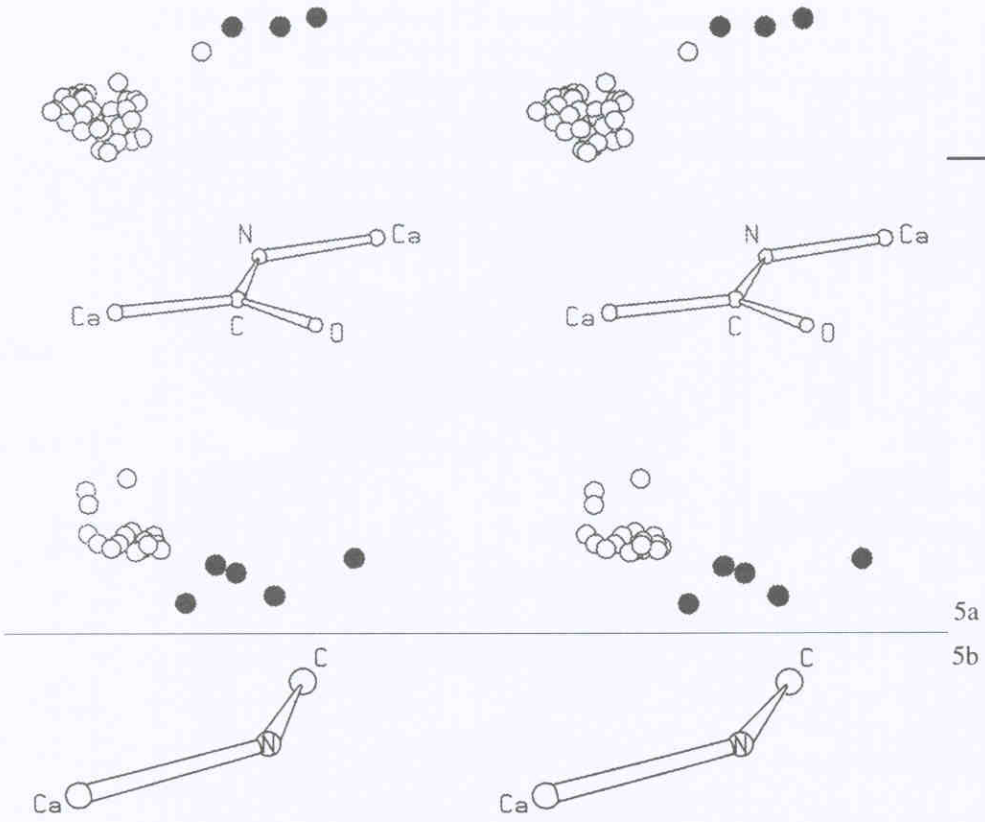


Figure 5: (a) Stereoview displaying the disposition of the sulfur atoms (showing intra- as well as inter residue S···C interactions denoted by open and filled circles, respectively) on the face of the peptide plane. (b) Disposition of S atoms showing inter residue S···N interaction (details given in Table II). Although labeled as C_α and C the two atoms bonded to N have different identities when the N atom belongs to a side-chain.

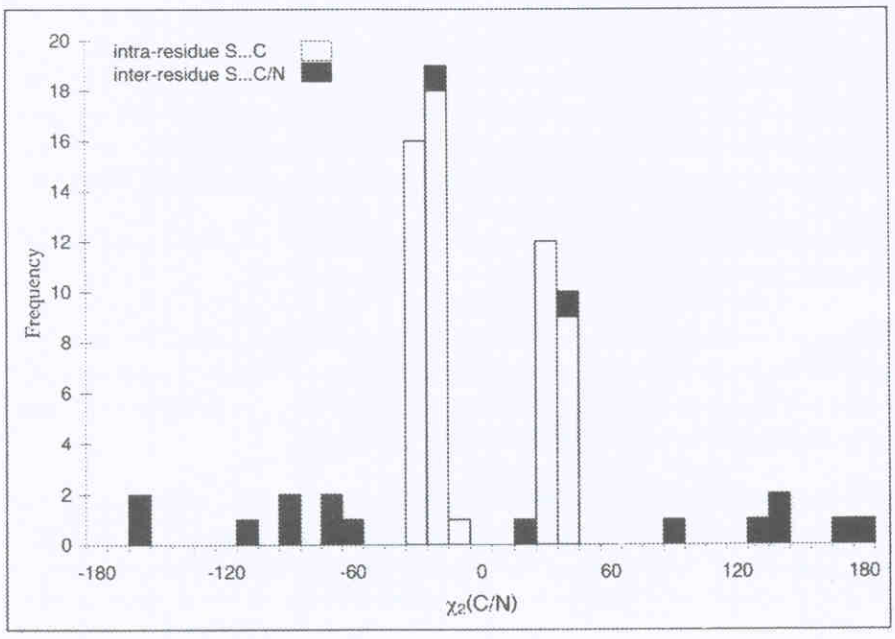


Figure 6: Histogram of the torsion angle χ_2(C/N) (°) (corresponding to C_α-C_β-S_γ···C/N).

bilize the helical structure, but additionally, can anchor it to another helix through the S···C/N interaction. An extensive tertiary contacts involving the S···C interactions is found in aconitase (21) (Figure 7) - interestingly, the two consecutive residues at the end of a 3_{10}-helix that partake of S···C interactions are Gly. Indeed, there is a surfeit of Gly residues whose main-chain C/N atoms interact with Cys residues (Table II). The quaternary structure, in addition to the tertiary structure, can also gain in stability through this interaction, as exemplified by lactate dehydrogenase, (22) where a Cys S atom from one subunit associates with the C atom from another subunit.

Table II
Inter-residue S···C/N interaction*

Protein Name	PDB code	Cysteine Residue	Cysteine Structure	Partner Atom	Partner Residue	Partner Structure
(A) Main-chain partner						
Hemoglobin	1BAB	105(A)	H(9/17)	N	M33	H(12/15)
Ferredoxin	1FDD	11	T	N	D95	T
Glycogen phosphorylase B	1GPB	495	T	C	W491	I
Cytochrome P450-CAM	1PHB	85	S	C	G93	H(4/6)
		148	E(2/3)	N	F150	H(1/4)
Cyclophilin A	2CPL	161	E(6/8)	N	K49	T
Myohemerythrin	2MHR	35	H(17/19)	C	N43	H(3/24)
Glutathione reductase	3GRS	333	G(4/4)	C	A328	E(3/3)
		440	H(1/13)	N	G437	S
Aconitase	8ACN	178	E(2/4)	C	G171	G(4/7)
		257	H(8/10)	C	G172	G(5/7)
				C	S268	E(2/3)
Lactate dehydrogenase	9LDT	187(A)	E(1/2)	N	G177(A)	H(12/15)
				C⊥	T306(B)	C
(B) Side-chain partner						
Cytochrome P450-CAM	1PHB	85	S	NE2	Q317	E(1/3)
Tyrosine kinase transforming protein	1SHA	42	E(2/9)	NH2	R32	E(4/5)
Bovine papilloma virus	2BOP	327	E(1/7)	NE	R386	H(4/10)
Cytochrome P450	2HPD	62	T	CG	N395	S

*Atom labels, according to the PDB convention, are given. The partner is represented by one-letter amino acid code, followed by the residue number (and the subunit designation). The secondary structure is as defined by the program, DSSP:(15) H, α-helix; G, 3_{10}-helix; P, π-helix; E, extended strand; B, residue in isolated β-bridge; S, bend; T, H-bonded turn; C, non-regular structure. H and E are followed by (m/n), where n corresponds to the number of residues in the given secondary structural element and m is the position of the residue from the N-terminal end of the structure.

⊥Although symmetry-related (crystallographic, as well as non-crystallographic) molecules were not considered in this analysis, this particular case has been included here to show how a Cys residue can facilitate the association between two subunits.

The amide group interacting with the S atom can come from a side-chain also (Table II). The interaction with the Arg side-chain provides an interesting case. Although it is quite rare to find Cys and Arg side-chains in a close encounter (possibly because they seek different protein environments), when they do come close to each other, it is quite likely that the Cys S is on the face of the guanidinium group (there are two examples of hydrogen bond formation between them in Figure 2).

There does not appear to be much restriction on the χ_2(C/N) torsion angle (Figure 6), unlike the case of the intra-residue S···C interaction.

(d) *Secondary Structural Features*. Figure 8 shows the secondary structural elements in which the Cys residues are located. None of the structures has a very high proportion of residues showing the intra-residue S···C interaction (the percentage being slightly higher than 50 for the strand structures).

28 out of 43 helical Cys residues (at position *i*) form hydrogen bond with the carbonyl group (at *i*-4) (such residues have χ_1 around 300°, Figure 4(a)); in 5 cases Cys is located in the first turn of the helix, so that a similar intra-helical hydrogen bond interaction is not possible. In the remaining 10 cases, the sulfur atom has S···C

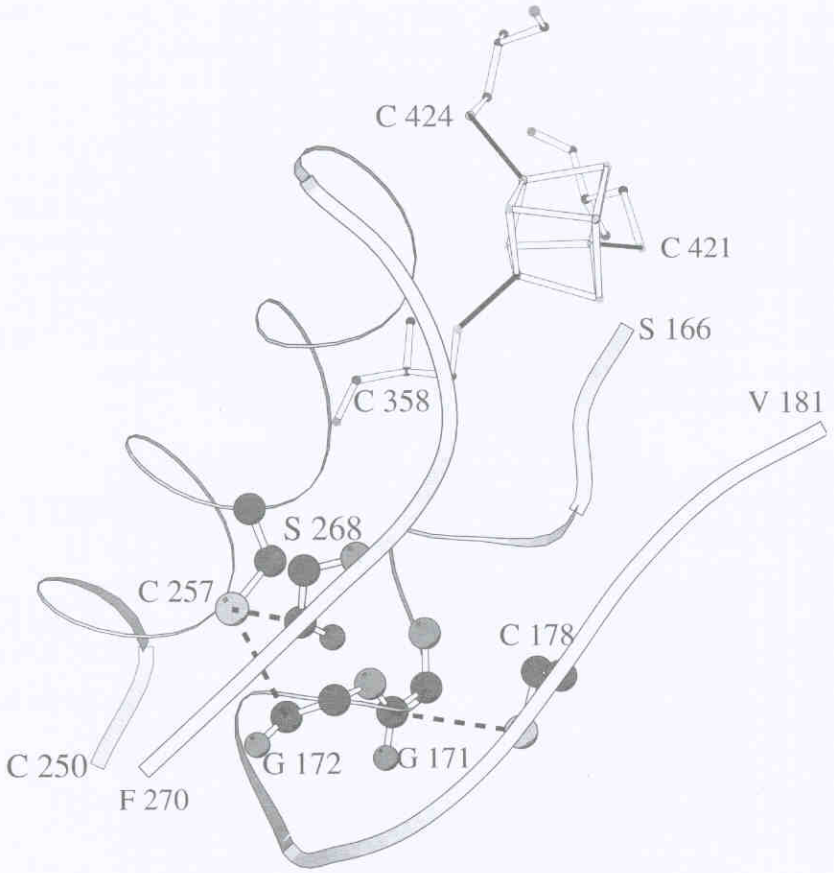


Figure 7: Intermolecular S···C interactions (dashed lines), taken from aconitase (Table II), involving the sulfur atoms of C178 and C257, and the carbonyl carbon atoms of G171, G172 and S268. The secondary structural elements containing these residues, and the iron-sulfur cluster (with its associated cysteine ligands) at the active site are also shown.

interaction, does not participate in any hydrogen bond at all, and the χ_1 angle is mostly around 180°. Thus in two out of three conformational states the Cys side-chain is involved in stabilizing intra-helical interactions (S···C or hydrogen bonding), and may explain why among all the secondary structural elements the maximum number of Cys residues occur in helices (Figure 8).

(e) *Active Site Residues*. A Cys residue plays a crucial role in the active site of cysteine proteases and many other enzymes. A prerequisite for these residues seems to be a reduced pK_a value, for which the helix dipole model has been used to provide an explanation. (23) According to this, the ionization of these residues, many of which occur at helical N-terminus, is facilitated because the resulting negative

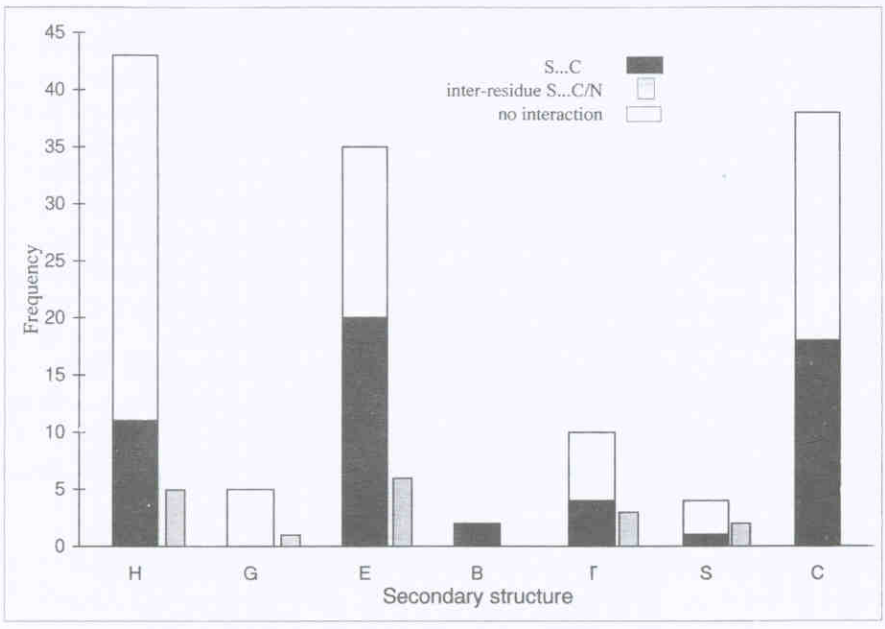


Figure 8: Distribution of Cys residues among the various secondary structural elements(15) (H stands for α-helix; G, 3_{10}-helix; E, extended strand; B, residue in isolated β-bridge; T, H-bonded turn; S, bend; C, non-regular structure). Thin bars (subsets of the adjacent wide bars) contain the examples of inter residue S···C/N interaction.

charge can then interact favorably with the positive pole of the helix-dipole. Moreover, an adjacent proton donor can also stabilize the thiolate anion by hydrogen bond formation. Thus both these factors, individually or in unison, can lower the pK_a of a Cys residue. However, a look at some representative Cys residues occurring in the active site (Table III) shows that quite a few of them are not located at helical N-terminus, or lack any protein atom within a hydrogen bond distance. In various structures the nearest base is of different type and at different distances. Many of these Cys residues, however, show intra-residue S···C interaction. Interestingly, although papain, (24) does not show this interaction in the free state, a closely related structure, actinidin in complex with the substrate, (25) has the S atom in an interacting position. The delocalization of the negative charge from the S atom to the carbonyl group, brought about by the S···C=O interaction, can be an effective element in reducing the pK_a value of the Cys residue.

Table III

Active site cysteines, their side-chain conformation, χ_1(°), and interactions*

Protein			Cysteine			Interactions		Nearest base-atom[3]
Name	Resolution (Å)	PDB code	Residue no.	Structure[1]	χ_1	H. bond	S···C (θ)[2]	
Papain	1.7	1PIP	25	H (1/18)	37	-	No (64)	ND1-H159(4.0)
Actinidin	1.86	1AEC	25	H (1/18)	63	N-A163	Yes (39)	ND1-H162(3.9)
Glyceraldehyde-3-phosphate dehydrogenase	1.8	2GD1	149	H (1/17)	53	O(w)	No (53)	NE2-H176(4.1)
Thymidylate synthase	2.1	3TMS	146	S	-72	-	No	NH2-R166(4.0)
Thioredoxin reductase	2.0	1TRB	135	C	-177	OG-S138, N-S138, O(w)	Yes (35)	-
Thioredoxin	1.68	2TRX	32	C	165	N-C35, O-I75	Yes (30)	-
			35	H (3/16)	-62	N-P76	No	-
DNA methyl transferase	2.5	1HMY	81	C	171	-	Yes (27)	NE2-H127(4.5)
Tyrosine phosphatase	2.2	1PNT	12	E (7/7)	69	N-N15, OG-S19, N-S19	Yes (39)	NH2-R18(4.7)

*Structures with a slightly lower resolution have also been included. Cys residues are identified by their residue number; partner atoms are specified by their labels followed by one-letter amino acid code and residue number; O(w) stands for the water-oxygen atom.

[1]For the definition of the secondary structure see Table II footnote.

[2]The presence (or absence) of the intra-residue S···C interaction is indicated by Yes (or No), and (when the side-chain conformation is in the g^- or t state) the corresponding value of θ (°) is given in parentheses.

[3]Atoms within 5Å are considered; distances (Å) are in parentheses.

(f) *S···aromatic Interactions*. 28 Cys residues were found adjacent to an aromatic ring (Phe, 15 cases; Tyr, 11; His, 1; Trp, 1), the average of the shortest S···C contact lengths being 3.6Å (Table I). Figure 9 shows the distribution of the angle, θ (Figure 1) for this interaction. If a value of 45° is used to demarcate the cases interacting with the face ($\theta < 45°$) from those interacting with the edges ($\theta > 45°$), a greater number (19) is found in the former category than the latter (9). Moreover, the observed occurrence along the face is much more than what is expected from a consideration of the area available for the S atom to interact at a given θ value (Figure 9). Our observation is in contrast to the result obtained by Reid and coworkers, (26) who found, using all S-containing residues and a distance of separation ≤ 6Å from the ring centroid, that the S atoms express an affinity towards the edge of the aromatic rings and avoid the region above the ring in the vicinity of the π-electrons. Even if we define θ (Figure 1) using the ring center as the origin, the character of our plot (Figure 9) is not changed significantly, suggesting that for the Cys residues used in this analysis the S atom has a preference to interact with an aromatic face. This conclusion is along the line of an earlier analysis (27,28) and a more recent one involving Met residues. (29)

The S···aromatic interaction appears to be a long range one; the numbers of residues separating the two interacting groups are: 2 (in 2 cases), 4 (3), 5 (1), 6 (1),

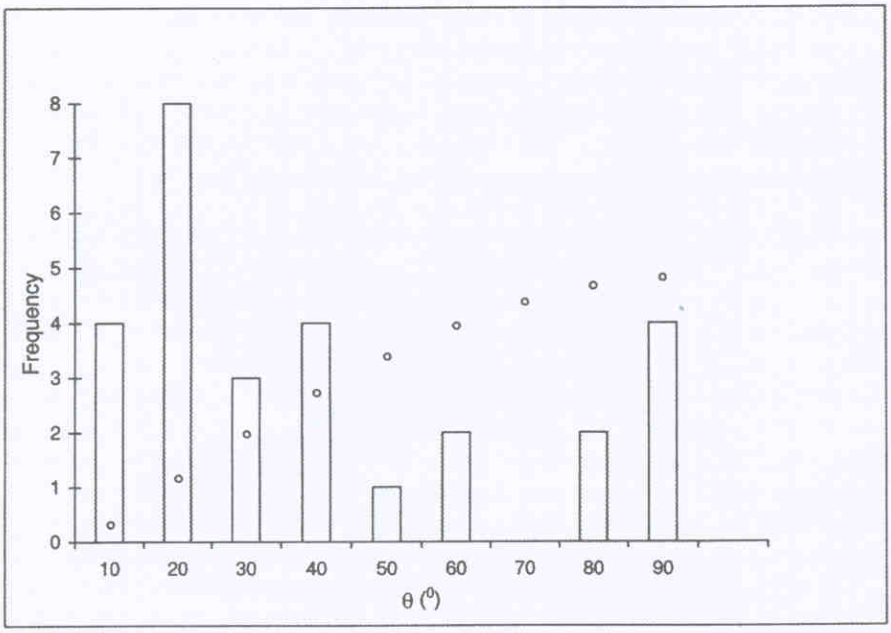


Figure 9: Histogram of the angle θ (°) when the Cys S atom is in contact with one or more atoms of an aromatic ring. The points show the values expected at various θ angles (approximated by sinθ) if the total number of examples were distributed randomly.

>6 (21). Two cases comprising of a Cys residue following an aromatic group by 4 residues may constitute an interesting helical motif. In both the structures (Figure 10) the sulfhydryl moiety interacts with the face of the ring, as well as forms a hydrogen bond with the carbonyl group of the aromatic residue. In order to achieve these interactions the two side-chains take up almost identical conformations in the two structures.

Glycolate oxidase (30) (PDB code, 1GOX) provides the only example where a S atom (Cys343) is on top of a ring N atom (NE2) of a His residue (353). Likewise, in flavodoxin (31) (1OFV), S of Cys54 is on top of CD1 (and also quite close to the ring N atom) of Trp66. Although limited in number (which can be expected as the concerned residues are relatively less abundant), these fall in the pattern of the S···N interaction (Table II), and may be useful in the binding of heterocyclic rings in protein structures. (32)

Discussion

In this paper we have attempted to take a holistic view of the various types of interactions involving the Cys sulfur atom. Earlier studies (9,11-12) on hydrogen bond interactions pertaining to Cys residues did not distinguish between the metal-bound

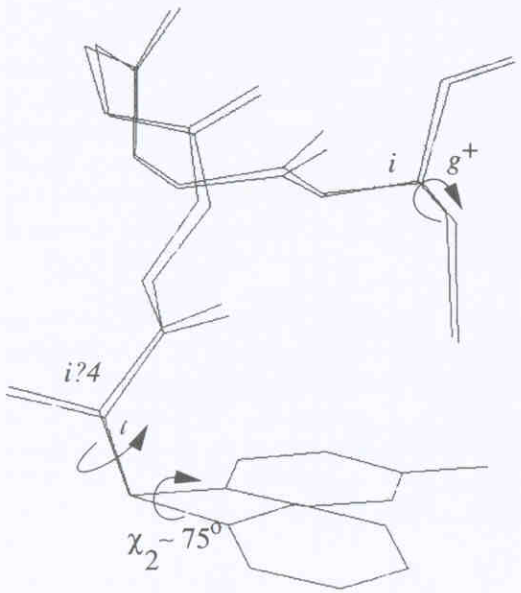


Figure 10: View of the superimposition (using the backbone atom coordinates only) of the 5-residue helical stretch, -Phe-Leu-Leu-Lys-Cys18- in granulocyte colony-stimulating factor (PDB code, 1BGC) and -Tyr-Ala-Ser-Ile-Cys177(A)- in fructose-1,6 biphosphate aldolase (1FBA). The side-chain atoms of the 3 central residues are omitted. The helix in the former structure is 29 residues long and Cys occupies the 15th position; the corresponding numbers for the latter structure are 20 and 18, respectively. The side chain conformations, χ_1 (Cys), and χ_1 and χ_2 (aromatic) angles are: -73, 176 and 74° in the former, and -71, 142 and 74° in the latter.

Cys (whose -S⁻ atom can only be an acceptor) and the free ones (with a -SH group and both donor and acceptor capability). Our study, concerned only with those belonging to the latter type, has revealed the hydrogen bond characteristics of the -SH group. In addition to forming hydrogen bonds, the S atom of Cys residues can also be placed on the face of an amide plane, on top of the C (or N) atom. Some of these residues are, in addition, also close to the face of an aromatic ring.

(a) *Hydrogen Bond Features*. As most of the interactions of the sulfhydryl group is with a carbonyl oxygen atom (an acceptor) (Figure 2), its role in a hydrogen bond interaction can be assumed to be a proton donor, rather than an acceptor. It shows very little inclination to associate with acidic side-chains. One reason for this could be the tendency of the acidic groups to be exposed to protein surface and the Cys residues to be buried inside, giving the residues very little chance to face each other. However, it is equally true that from a consideration of the charge, electronegativity and polarizability of the constituent atoms the -SH group is rather a 'soft' donor, whereas the carboxylate group is a 'hard' acceptor, resulting in a mismatch in the hyrogen bonding capability of the two groups. (33) As a result, the -COO⁻···HS- interaction is not strong enough to enable a thiol group to displace the water molecule(s) from the solvated carboxylate group and establish a direct contact. In this connection, it is instructive to recollect that no Cys residue was found in the binding sites of sulfate and phosphate anions in protein structures. (34) Again, a replacement of Ser130, a ligand in the sulfate-binding protein, with isostructural Cys caused a 3200-fold decrease in the sulfate-binding activity. (35) Although it has been attributed to steric effect, our results also suggest that the use of a thiol group to bind anions may not be favorable energetically.

(b) *S···C/N Interaction*. Most of the Cys residues with the side-chain conformation in the g^- and t states (Figure 4(a)) have the S atom sitting nearly on top of the peptide C atom of the same residue such that the angle θ (Figure 1) is less that 40°, and the angle S···C=O, 102° (Table I). This S···C interaction, which can provide stability comparable to a hydrogen bond, restricts the rotation of χ_1 and ψ torsion angles (Figure 4(c)).

The S atom can similarly interact with the face of another amide group (main-chain, as well as side-chain), by being on top of C or N atom (Figure 5). A modest number of cases with sulfur atom interacting with the face of a guanidinium side-chain or a heterocyclic ring (near the ring-N atom) has also been observed.

(c) *Tertiary Interactions*. The inter-residue S···C and S···N interactions (Table II), where the interacting atoms are located in different secondary structural elements, can stabilize the tertiary structure (Figure 7). One of the elements is usually a helix (α or 3_{10}), whereas in a few cases both the residues are helical. Thus, although the interactions between the side-chains are assumed to be the determining factor for helix-helix packing, (36) here is a case of side-chain - main-chain interaction stabilizing the association between helices.The residue providing the main-chain C/N atom is predominantly Gly. Even though Gly is not a very common residue in helix, (37) due to its minimum bulk its position in a helix can help another Cys side-chain come near the face of its peptide group, thereby holding the two secondary structural blocks in close proximity.

(d) *Functional Relevance*. The helix dipole model has been used to explain the low pK_a values of some active site Cys residues. (23,38) However, it is found that not all such residues are near the N-terminal ends of helices (Table III), and for a few, there is no proton acceptor nearby. This, coupled with the fact that in most of these cases there is S···C interaction, suggests that the delocalisation of the excess charge carried by the ionized sulfhydryl group over the peptide moiety may provide an alternate method of reducing the pK_a value.

The S···N nonbonding contact has been observed in some model compounds

designed to represent the acyl-thiolprotease intermediates during the reaction of cysteine proteases. (20) As such, the preference shown by the S atom to interact with the face of the peptide group can be exploited to design enzyme inhibitors and drugs.

(e) *S···aromatic Interactions*. 20% of Cys residues have an aromatic neighbor (which is usually quite distant in sequence), and there is a distinct tendency for the S atom to interact with the face of the aromatic residue (Figure 9). A question arises concerning the nature of this interaction with the π-electron cloud of the aromatic ring - whether it is of the type S-H···π or there is a direct contact (S···π) involving the S atom. In absence of any knowledge on the position of the H atoms one can only conjecture as to nature of the exact bonding. On one hand, as 64% of these Cys residues are also in close contact with carbonyl oxygen atoms, it may be assumed that the -SH proton is already occupied as hydrogen bond donor, leaving the S atom to interact directly with the aromatic system. The S···π interaction is observed involving the Met side-chain (29) and also in small molecule structures (39) (where there is no proton attached to the S atom). On the other hand however, with the -SH group the S-H···π interaction may be more relevant; looking at Figure 10 if one places the sulfhydryl proton so as to make a hydrogen bond with the CO group of the aromatic residue in the preceding turn, the proton will be very nearly on top of the aromatic ring.

The S···aromatic interaction, in conjuction with S-H···O=C hydrogen bonding, can be used for designing helical structures. In two independent structures, a Phe/Tyr-X-X-X-Cys sequence has been found in helical conformation (Figure 10).

(f) *Protein Stability*. As the S atom can participate in a variety of interactions, a Cys residue in protein interior can provide considerable stability to the structure, and its substitution by any other residue, even if it does not cause any change in the structure, may be destabilizing. Indeed, in a recent mutational study on tryptophan synthase, it has been shown that when Ser, Ala, Val or Gly was substituted for each of the three buried Cys residues the resultant mutants were less stable than the wild type. (40)

Conclusions

In proteins the Cys side-chain finds a milieu so as to satisfy two important requirements of the S atom - hydrogen bonding, mainly to a carbonyl oxygen atom, and a position on top of a peptide plane (of the same or different residue). Moreover, there is also a marked tendency of the S atom to interact with the face of an aromatic ring. Proton is invariably an important player in most of the directional interactions between various groups in proteins. However, S does not always need a proton to mediate its association with another group. These interactions provide stability to the secondary and the tertiary structures, and can affect the pK_a of the Cys residues, and are, therefore, equally important for protein function. Various interactions involving the sulfur atom endows the Cys residues with unique features, making them, even when they serve no apparent functional role, rather uninterachangeable by any other residue in the sequence.

Acknowledgments

We thank Dr. Gautam Basu for encouragement, the Council of Scientific and Industrial Research for a fellowship (to DP), the Department of Science and Technology for a grant (to PC), and the Department of Biotechnology for the National Facility for Biomolecular Modeling.

References and Footnotes

1. S. K. Burley and G. A. Petsko, *Adv. Prot. Chem. 39*, 125-189 (1988).
2. T. Lazaridis, G. Archontis and M. Karplus, *Adv. Prot. Chem. 47*, 231-306 (1995).

3. P. Chakrabarti and D. Pal, *Protein Science 6*, 851-859 (1997).
4. E. Adman, K. D. Watenpaugh and L. H Jensen, *Proc. Natl. Acad. Sci. USA. 72*, 4854-4858 (1975).
5. B. Shen, D. R. Jollie, T. C. Diller, C. D. Stout, P. J. Stephens and B. K. Burgess, *Proc. Natl. Acad. Sci. USA. 92*, 10064-10068 (1995).
6. J. Meyer, J. Gaillard and M. Lutz, *Biochem. Biophys. Res. Comm. 212*, 827-833 (1995).
7. E. Babini, I. Bertini, M. Borsari, F. Capozzi, A. Dikiy, L. D. Eltis and C. Luchinat, *J. Am. Chem. Soc.118*, 75-80 (1996).
8. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc. 117*, 5179-5197 (1995).
9. J. A. Ippolito, R. S. Alexander and D. W. Christianson, *J. Mol. Biol. 215*, 457-471 (1990).
10. L. M. Georget, S. D. Rader, R. J. Fletterick and F. E. Cohen, *Proteins: Struct. Func. Genet. 9*, 99-107 (1991).
11. D. F. Stickle, L. G. Presta, K. A. Dill and G. D. Rose, *J. Mol. Biol. 226*, 1143-1159 (1992).
12. I. K. McDonald and J. M. Thornton, *J. Mol. Biol. 238*, 777-793 (1994).
13. F. C. Bernstein, T. F. Koetzle, G. J. B Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol. 112*, 535-542 (1977).
14. U. Hobohm and C. Sander, *Protein Science 3*, 522-524 (1994).
15. W. Kabsch and C. Sander, *Biopolymers 22*, 2577-2637 (1983).
16. P. Kraulis, *J. Appl. Crystallogr. 24*, 946-950 (1991).
17. P. Chakrabarti, *Biochemistry 28*, 6081-6085 (1989).
18. J. Janin, S. Wodak, M. Levitt and B. Maigret, *J. Mol. Biol. 125*, 357-386 (1978).
19. T. M. Gray and B. W. Matthews, *J. Mol. Biol. 175*, 75-81 (1984).
20. K. I. Varughese, A. C. Storer and P. R. Carey, *J. Amer. Chem. Soc. 106*, 8252-8257 (1984).
21. H. Lauble, M. C. Kennedy, H. Beinert and C. D Stout, *Biochemistry 31*, 2735-2748 (1992).
22. C. R. Dunn, H. M. Wilks, D. J. Halsall, T. Atkinson, A. R. Clarke, H. Muirhead and J. J Holbrook, *Trans. R. Soc. London 332*, 177 (1991).
23. W. G. J. Hol, *Prog. Biophys. Molec. Biol. 45*, 149-195 (1985).
24. A. Yamamoto, K. Tomoo, M. Doi, H. Ohisi, M. Inoue, T. Ishida, D. Yamamoto, S. Tsuboi, H. Okamoto and Y. Okada, *Biochemistry 31*, 11305-11309 (1992).
25. K. I. Varughese, Y. Su, D. Cormwell, S. Hasnain and N. H. Xuong, *Biochemistry 31*, 5172-5176 (1992).
26. K. S. C. Reid, P. F. Lindley and J. M. Thornton, *FEBS Letters 190*, 209-213 (1985).
27. R. S. Morgan, C.E. Tatsch, R. H. Gushard, J. M. McAdon and P. K. Warme, *Int. J. Peptide Protein Res. 11*, 209-217 (1978).
28. R. S. Morgan and J. M. McAdon, *Int. J. Peptide. Protein. Res. 15*, 177-180 (1980).
29. T. M. Kingler and D. L. Brutlag, *Protein Science 3*, 1847-1857 (1994).
30. Y. Lindqvist, *J. Mol. Biol. 209*, 151-166 (1989).
31. M. L. Ludwig, C. L. Drennan, D. Hoover, A. L. Metzger, K. A. Pattridge and C. H. Weber, (to be published).
32. P. Chakrabarti and U. Samanta, *J. Mol. Biol. 251*, 9-14 (1995).
33. C. B. Aakeröy and K. R. Seddon, *Chem. Soc. Reviews 22*, 397-407 (1993).
34. P. Chakrabarti, *J. Mol. Biol. 234*, 463-482 (1993).
35. J. J. He and F. A. Quiocho, *Science 251*, 1479-1481 (1991).
36. C. Chothia, M. Levitt and D. Richardson, *J. Mol. Biol. 145*, 215-250 (1981).
37. J. S. Richardson and D. C. Richardson, *Science 240*, 1648-1652 (1988).
38. T. Kortemme and T. E. Creighton, *J. Mol. Biol. 253*, 799-812 (1995).
39. T. Kimura and N. Furukawa, *Tetrahedron Letters 36*, 1079-1080 (1995).
40. K. Hiraga and K. Yutani, *Protein Engng. 9*, 425-431 (1996).

Date Received: December 5, 1997

Communicated by the Editor Ramaswamy H. Sarma