

An electrophile-nucleophile interaction in metalloprotein structures

PINAK CHAKRABARTI AND DEBNATH PAL

Division of Physical Chemistry, National Chemical Laboratory, Pune 411008, India

(RECEIVED October 15, 1996; ACCEPTED January 22, 1997)

Abstract

A new attractive interaction in metalloprotein structures, between the thiolate anion of a metal-bound cysteine (acting as a nucleophile) and a carbonyl carbon of a peptide group (an electrophile), has been identified. From 82 cases extracted from 23 metalloprotein structures, the interacting S and C atoms are found to be at a distance of $3.2 (\pm 2) \text{ \AA}$, such that the angle $S \cdots C - O$ is $109^\circ (\pm 15^\circ)$. Usually, the interacting atoms are from the same Cys residue, and to allow the S to interact with the carbonyl group the side-chain and the main-chain torsion angles deviate from those found in cysteines not bound by metals. There is a good correlation between the $S \cdots C$ distance and the angular deviation of the $S \cdots C$ vector from the normal to the peptide plane. Various data points may be envisaged to represent "snapshots" along the reaction coordinate for the intra-residue attack of Cys S on the CO group.

Keywords: conformation; cysteine; metalloprotein; stability; sulfur–carbonyl interaction

In spite of the progress made over the last few decades in theory and experiment, the protein folding problem still remains elusive (Honig & Yang, 1995). Although the major forces (hydrophobic, van der Waals, electrostatic and hydrogen bonding) contributing to the stability have been known for a long time, weaker ones like the $C - H \cdots O$ (Derewenda et al., 1995), OH or $NH \cdots$ aromatic (Burley & Petsko, 1986; Perutz, 1993), aromatic \cdots aromatic (Burley & Petsko, 1985), $CH \cdots$ aromatic (Chakrabarti & Samanta, 1995), and $S \cdots$ aromatic (Reid et al., 1985) are only now being identified. These interactions are directional in the sense that each group has a reasonably rigid geometry in which it can engage its partner, and can thus direct the structural organization of the protein molecule and its recognition of the substrate. In this paper we characterize one such interaction that could have important bearing on the structure and function of metalloproteins.

Results from small molecule crystallography have been used not only to understand molecular architecture and structure–function relationship (Dunitz, 1979); they have also been used by Bürgi and Dunitz to derive information on reaction pathways (Bürgi et al., 1973, 1974a, 1974b, 1974c; Dunitz, 1979; Bürgi & Dunitz, 1983). A striking example is the nucleophilic addition of an amino or a hydroxyl group to an electrophilic carbonyl group. One can then ask the question if this energetically favorable (Bürgi et al., 1974c) and thus stabilizing interaction can be observed in protein structures. The cation-bound sulfhydryl group, which really exists as the thiolate anion, is likely to have nucleophilic characteristics as

has been shown in some biological systems (Myers et al., 1993; Wilker & Lippard, 1995), and our analyses of cysteine-containing metalloproteins show that this can indeed interact with a carbonyl group behaving like an electrophile.

Results and discussion

Geometrical parameters (Fig. 1) involving the S and C atoms that satisfy our distance criterion are given in Tables 1 and 2, the former listing the cases where the two atoms are from the same Cys residue, the latter those where the atoms emanate from different residues. The greater number of entries in Table 1 as compared to Table 2 (78:4) suggests that the interaction is predominantly an intra-residue phenomenon, although the existence of cases in Table 2 points to its more general nature.

The sum of van der Waals radii (Bondi, 1964) of single-bonded S and sp^2 -hybridized C atoms is 3.6 \AA . Distances shorter than this (the average $S \cdots C$ length is $3.2 (\pm 2) \text{ \AA}$) point to the attractive nature of the interaction. The $S \cdots C - O$ angle, $109^\circ (\pm 15^\circ)$, between the nucleophile and the CO group is similar to what Bürgi and Dunitz observed (Bürgi et al., 1973, 1974b; Bürgi & Dunitz, 1983). Unfortunately, another hallmark of the electrophile-nucleophile interaction, namely, the deviation of the C atom from the plane defined by CA, N, and O (Fig. 1) toward the nucleophile (Bürgi et al., 1973) cannot be seen in protein structures, as it is small and in any case the refinement process that applies the constraint of planarity on the peptide group would artificially conceal any such displacement. However, in the following sections we put forward some more evidence in support of the attraction between the S and C atoms.

Reprint requests to: Pinak Chakrabarti, Division of Physical Chemistry, National Chemical Laboratory, Pune 411008, India; e-mail: pinak@ncl.ernet.in.

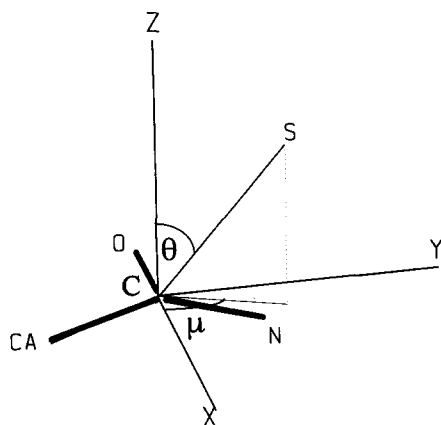


Fig. 1. Definition of the spherical polar coordinate system used to analyze the sulfur position (S) around the peptide CO group. Atom C is assigned as the origin, the x -axis is along the O—C direction, the y -axis is on the peptide plane and going toward the N position; the z -axis is along the normal to the peptide group. θ is the angle made by the $C \cdots S$ vector with the z -axis; μ is the angle between the x -axis and the projection of the $C \cdots S$ direction on the xy plane ($\mu = 90^\circ$ along the y -axis and -90° along the reverse direction).

Cysteine side-chain conformation

The torsion angle, χ_1 of a Cys residue can take up three values, 60° (g^-), -60° (g^+) and 180° (t) (Janin et al., 1978; McGregor et al., 1987). The χ_1 distribution of all Cys ligands is shown in Figure 2, and resembles the one obtained in an earlier analysis on ligand cysteines (Chakrabarti, 1989). Interestingly, all the residues with S and C atoms in close contact have values around $+60^\circ$ and 180° (Table 1). In these orientations the S atom is above or below the peptide group (Fig. 3). At the g^+ conformation, the two concerned atoms are farthest from each other (see the curve in Fig. 2) and there is no intra-residue interaction. However, a few of these S atoms are in close proximity to a carbonyl C atom belonging to a different residue (Table 2), and exhibit a similar geometry of interaction (Fig. 4). In general, in protein residues the g^- state is the least stable because of steric factors, and of the remaining two, g^+ appears to be slightly more favored (Janin et al., 1978; McGregor et al., 1989). In our sample, the population of g^+ has been drastically reduced (only 17% in Fig. 2) in favor of the g^- and t states, where the intra-residue $S \cdots C$ interaction is feasible. Another factor that underlines the attractive nature of the interaction is the shift of the g^- peak to a value larger than 60° , a change that minimizes the $S \cdots C$ distance (Fig. 2).

Table 1. Metal bound cysteine residue showing $S \cdots C$ interaction involving its SG and carbonyl C atoms

Protein name	Code ^a	Residue ^b	Distance (Å)	Angle (°)	Torsion angle (°) ^c			Spherical polar angle (°) ^d	
					$S \cdots C$	$S \cdots C - O$	χ_1	$\chi_2(C)$	ψ
Alcohol dehydrogenase	2OHX	46(A)	3.17	122.5	172.4	-25.7	147.5	144.9	-19.9
		97(A)	3.17	109.4	74.5	26.7	-20.0	25.3	-35.1
		100(A)	3.09	95.3	74.7	25.1	166.3	154.5	-77.5
		111(A)	3.08	114.5	174.6	-26.1	137.6	151.2	-24.9
		174(A)	3.07	109.6	182.4	-28.5	-72.0	23.7	-43.5
Carbonic anhydrase Metallothionein	1DCA 4MT2	199	3.30	112.9	50.1	33.3	-10.8	34.9	-38.9
		5	3.33	113.9	174.1	-28.9	130.0	150.5	-42.5
		13	3.12	122.5	182.6	-28.3	138.0	145.5	-23.6
		15	3.28	133.7	52.6	34.7	-33.8	47.3	-24.6
		19	3.18	95.7	52.8	31.8	174.1	147.8	-78.5
		24	3.27	116.1	196.9	-34.5	131.6	145.2	-33.2
		29	3.45	109.2	200.2	-34.4	123.8	143.4	-51.0
		33	3.50	116.7	43.4	34.2	-8.9	37.9	-44.6
		34	3.42	93.2	45.3	33.8	165.7	142.7	-84.1
		36	3.46	122.0	32.6	36.2	-8.5	45.4	-41.8
		41	3.11	105.2	180.2	-26.9	122.4	152.1	-55.2
		48	3.28	117.3	189.7	-33.8	129.6	146.5	-39.9
		57	2.98	80.8	87.2	20.1	150.0	154.9	-112.7
DNA binding protein (ZIF268)	1ZAA	7(C)	3.14	109.0	178.0	-27.5	121.1	149.0	-51.1
		12(C)	3.11	127.2	183.4	-27.1	143.7	140.9	-14.6
		37(C)	3.11	109.9	168.9	-25.9	127.0	151.3	-45.8
		40(C)	3.25	96.4	70.8	28.4	-1.2	30.9	-77.6
		65(C)	3.14	114.9	164.4	-23.1	139.0	150.6	-32.3
Glucocorticoid receptor	1GLU	68(C)	3.11	100.6	67.7	29.8	-0.7	28.0	-66.8
		440(A)	3.13	117.1	180.8	-29.1	127.8	146.4	-35.8
		443(A)	2.84	67.3	98.0	12.7	19.4	43.1	-127.0
		457(A)	3.44	112.3	55.1	33.6	174.1	144.4	-52.7
		476(A)	3.12	136.2	156.5	-16.3	138.5	140.5	-30.0
Aspartate carbamoyl transferase	8ATC	492(A)	3.26	92.4	203.7	-33.9	80.8	137.3	-86.8
		109(B)	3.21	109.5	178.3	-24.8	122.9	150.0	-48.6
		114(B)	3.60	124.2	205.7	-39.1	142.2	134.9	-36.9
		138(B)	3.26	112.8	190.4	-34.9	123.5	147.6	-43.7
		141(B)	3.49	115.4	49.5	34.6	-22.6	39.3	-45.4

(continued)

Table 1. Continued

Protein name	Code ^a	Residue ^b	Distance (Å) S...C	Angle (°) S...C—O	Torsion angle (°) ^c			Spherical polar angle (°) ^d	
					χ_1	$\chi_2(C)$	ψ	θ	μ
Plastocyanin	7PCY	84	3.16	100.2	166.5	-24.8	115.6	150.5	-70.6
Azurin	2AZA	112(A)	3.05	116.8	169.1	-25.5	128.4	147.3	-41.0
Pseudoazurin	1PAZ	78	3.04	102.5	165.5	-29.9	121.4	152.9	-60.7
Amicyanin	1AAN	92	3.19	109.6	181.1	-30.8	122.7	148.9	-50.9
Cytochrome P450	2CPP	357	3.11	98.4	179.2	-28.9	124.9	156.5	-61.7
Nitrite reductase	1AFN	136(A)	3.36	108.2	187.6	-32.9	123.0	148.8	-55.1
Phthalate dioxygenase reductase	2PIA	272	3.21	126.4	80.6	26.0	-37.5	37.4	-19.0
		277	3.15	109.4	86.1	22.2	-22.7	25.9	-39.9
		280	3.12	88.5	68.7	27.5	16.7	32.4	-92.8
		308	3.27	121.6	67.6	30.7	-23.5	34.8	-31.7
Ascorbate oxidase	1AOZ	507(A)	3.08	112.6	174.6	-28.8	134.1	150.9	-35.8
Rubredoxin	8RXN	6	3.13	104.4	178.3	-30.0	119.7	150.5	-61.0
		9	3.07	102.3	75.1	29.1	-8.8	24.0	-58.6
		39	3.14	107.0	179.3	-31.2	127.6	151.4	-47.3
		42	3.14	102.0	76.7	27.0	-12.2	23.6	-54.0
High potential Fe-S protein	IISU	25(A)	3.14	116.4	177.4	-27.4	139.6	149.5	-27.7
		40(A)	3.11	113.7	171.6	-25.2	135.9	151.1	-33.9
		55(A)	3.13	109.3	91.7	20.1	161.9	151.6	-46.9
Ferredoxin I	IFRR	38(A)	3.09	119.8	74.2	27.1	-33.2	33.9	-28.8
		43(A)	3.12	118.9	82.5	17.3	190.1	147.7	-21.8
		46(A)	2.98	86.0	68.0	29.9	23.0	30.6	-98.0
		76(A)	3.17	114.6	69.2	27.4	-15.1	29.9	-33.1
Ferredoxin II	1FDX	8	3.49	101.4	188.1	-34.1	128.4	153.9	-54.1
		11	3.24	126.5	80.9	17.1	-46.8	39.3	-8.9
		14	3.36	123.4	64.5	37.4	-42.0	38.8	-24.4
		18	3.35	110.2	174.2	-24.6	119.6	148.0	-53.2
		35	3.06	108.0	176.7	-31.8	129.2	155.7	-41.7
		38	2.80	90.2	85.7	18.4	-1.0	29.5	-89.5
		41	3.27	143.4	47.6	32.4	-36.4	54.3	-10.8
Aconitase	8ACN	45	3.18	123.4	211.8	-39.5	122.2	138.9	-39.2
		358	3.41	133.5	56.8	30.8	-26.3	45.5	-21.7
Endonuclease	1ABK	424	3.30	80.8	189.2	-30.3	-35.1	42.5	-104.2
		187	3.37	76.7	197.6	-33.4	56.4	114.1	-104.8
		197	3.06	103.6	171.5	-25.0	122.8	155.2	-57.8
Trimethylamine dehydrogenase	2TMD	203	3.05	111.2	170.9	-24.5	133.6	153.6	-37.3
		345(A)	3.12	105.2	172.2	-26.8	129.1	155.3	-48.2
		348(A)	3.21	117.2	68.1	29.0	-22.2	32.5	-30.5
		351(A)	3.42	143.9	46.8	35.2	-48.2	56.6	-10.0
		364(A)	3.13	115.2	174.2	-26.9	129.3	150.8	-39.9
Nitrogenase Iron protein	1NIP	97(A)	3.00	141.0	173.5	-20.7	154.3	131.1	-9.2
		132(A)	2.85	90.7	56.8	29.4	190.8	160.2	-85.9
Nitrogenase Mo-Fe protein	1MIN	62(A)	3.06	93.0	71.1	29.2	170.8	155.1	-82.3
		275(A)	2.88	92.3	180.8	-29.6	102.8	154.3	-83.3
		70(B)	3.06	98.3	74.1	20.6	159.6	146.7	-76.9
		153(B)	3.46	130.7	23.2	46.6	-40.5	51.2	-30.5

^aFile identifiers in PDB.^bThe letter in parenthesis identifies the subunit to which the cysteine belongs when the protein is made up of different type of subunits.^c $\chi_1 = \text{N}-\text{CA}-\text{CB}-\text{SG}$, $\chi_2(C) = \text{CA}-\text{CB}-\text{SG}\cdots\text{C}$.^dAs defined in Figure 1.

Directionality vs. the strength of interaction

The interaction between the nucleophile and electrophile will be stronger if the sulfur lone-pair orbital approaches the CO group from a nearly perpendicular direction (Bürgi et al., 1973). In other words, the S...C distance should be shorter as θ (Fig. 1) decreases, as observed (Fig. 5A). Given the relative inaccuracy of the

protein structure, the near linearity of the distribution of the points is noteworthy.

For maximum interaction (along the tetrahedral direction), the absolute value of μ should be as small as possible. The steric constraints in the intra-residue interaction force the μ angles in Table 1 to assume negative values. Shorn of those constraints, the μ angles in Table 2 lie in the positive region also.

Table 2. Metal bound cysteine whose SG atom interacts with the carbonyl C atom of another residue^a

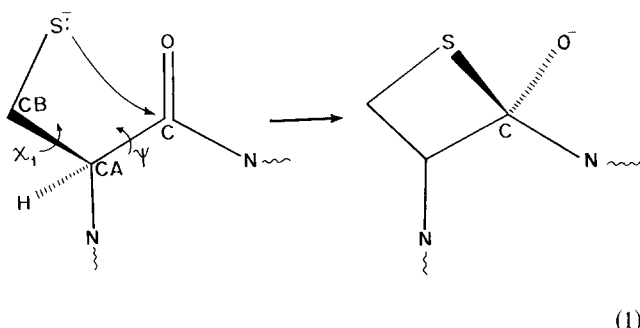
Protein name	Code ^b	Residues providing		Distance (Å)	Angle (°)	Torsion angle (°)		Spherical polar angle (°)	
		SG	CO			χ_1	$\chi_2(C)$	θ	μ
Metallothionein	4MT2	37	C36	3.60	83.6	-65.1	21.6	39.5	100.1
		60	C59	3.54	94.0	-53.4	14.6	38.3	83.2
Glucocorticoid receptor	1GLU	495(A)	N480(A)	3.53	108.1	-88.5	-110.7	160.9	-7.5
High potential Fe-S protein	1ISU	22(A)	G53(A)	3.44	81.0	-58.4	64.6	170.38	157.6

^aFor all the definitions see Table 1 footnotes with the modification that C is from the residue whose CO group is interacting.

^bThe cysteines that are neither involved in intra- nor inter-residue interaction are (code followed by residue number): 2OHX(103(A)), 4MT2(7,21,26,44,50,59), 1GLU(460(A),482(A)), 8ACN(421), 1MIN(95(B),154(A)). All these noninteracting residues have ψ_1 around -60° .

Mapping a reaction coordinate

The attack by the thiolate anion on the carbonyl group of a Cys residue would lead to the formation of a thietane derivative:



Because of various bonded and non-bonded constraints in a protein milieu, a covalent S—C bond cannot possibly be formed. Nevertheless, each entry in Table 1 can be regarded as a “frozen” intermediate at different points along the reaction pathway at an incipient stage. The geometrical changes taking place whenever there is a

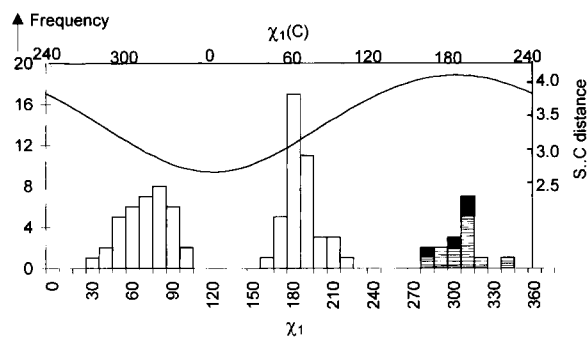


Fig. 2. Distribution of χ_1 torsion angle ($^\circ$) of all metal-bound Cys residues. Open and closed bars are for the cases with intra- (Table 1) and inter-residue (Table 2) interactions; hatched bar represents the examples (given in Table 2 footnote) showing no interaction. To facilitate visualization, another torsion angle, $\chi_1(C)$, ($C-CA-CB-SG$), has been marked against the upper abscissa—the two angles are related by $\chi_1 = \text{mod } 360(\chi_1(C) + 120)$. The curve shows the variation of the $S \cdots C$ distance (Å) (right side ordinate) as the torsion is changed through 360° using a model of a Cys-containing peptide fragment constructed using the standard dimensions and a main-chain ψ torsion angle of 120°

short contact between the γ position of the side-chain and the CO group, as is the case here, can be delineated by a plot of the $CB-SG \cdots C$ angle against the $S \cdots C$ distance (Fig. 5B). As the former increases from around 40° , the latter shortens from around 3.6 \AA . Besides χ_1 one can also use $\chi_2(C)$ (Table 1) to monitor the conformational changes. This angle exists in the range $0 \pm 35^\circ$ depending on whether χ_1 is in the g^- or t state, respectively. Moreover, the absolute value of $\chi_2(C)$ decreases with the shortening of the $S \cdots C$ distance.

Stability of the interaction

The stabilizing interaction between a nucleophile and an electrophile results from the overlap between the highest occupied molecular orbital (HOMO) of the former (usually a lone pair of electrons) and the lowest unoccupied molecular orbital (LUMO) of the latter (Liotta et al., 1984; Fujimoto, 1987). Ab initio calculation (Madura & Jorgensen, 1986) indicates that the addition of hydroxide ion to formaldehyde is exothermic by $\Delta E = 35 \text{ kcal/mol}$. Other theoretical studies on addition reactions (Bürgi et al., 1974a, 1974c) have produced energy surfaces with no barrier to the tetrahedral species. Even if there is an energy barrier between the reactants and the product, it can be surmised that at a separation distance below the “critical” distance (which is $\sim 2.8 \text{ \AA}$ for $O \cdots C=O$) (Menger, 1985, 1987) the system is on the product side and thus can provide increasing stability. For the $S \cdots C=O$ interaction the distance would be $\sim 3.2 \text{ \AA}$ (taking into account the difference of the van der Waals radii of S and O), a value around which most of our sample points are distributed. Arguments based on the electrostatics also substantiate the favorable nature of the atomic arrangement. For the $O \cdots C=O$ interaction, Maccallum et al. (1995) have shown that the Coulombic interaction between the oppositely charged non-bonded atoms, even when they are at a distance greater than the sum of their van der Waals radii, can have a magnitude as much as 80% of a hydrogen bond. In the present case, the non-bonded atoms are not only at a much shorter distance, they also orient for the optimal overlap of their respective molecular orbitals, and the gain in energy could be comparable to a hydrogen bond.

Cysteine main-chain torsion, ψ

The intra-residue interaction between S and C atoms places restriction on the values ψ can assume (see the left side of equation 1); only a limited number of (ψ, χ_1) combinations can achieve the

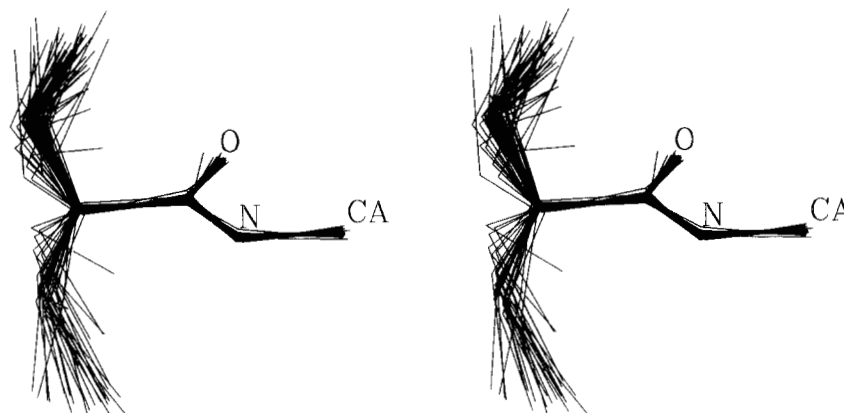


Fig. 3. The superposition (using the atoms belonging to the peptide group), displayed in stereo, of all Cys residues given in Table 1.

geometric requirement of the interaction. As can be seen in Figure 6, the points are clustered in two main groups: for χ_1 in the g^- state, ψ is in the range of -50 to 20° ; whereas the t state of χ_1 places ψ between 110 and 150° . Besides, there is a minor group with χ_1 in the g^- state and $\psi = 150$ – 200° . This is strikingly different from the general distribution. The χ_1 vs. ψ plot for all residues with a γ atom (Chakrabarti & Pal, to be published) shows this region to be populated equally with the other two, and moreover, ψ values are centered about 150° . Additionally, there is no dearth of points having the fourth combination (χ_1 in the t state and $\psi \approx -50^\circ$), which is practically non-existent in Figure 6. Thus, relative to the normal distribution, the (ψ, χ_1) pairs of Cys ligands have a lower propensity for two regions, one of which also shows a shift towards larger ψ values. This could be because the optimum interaction between the nucleophile and the electrophile requires the $S \cdots C-O$ angle to be in the range of 100 – 110° (Bürgi et al., 1973). The regions along the ψ axis that permit this magnitude of the angle to be maintained are shown against the two curves in Figure 6. The virtual absence of points centered at $(-50, 180)$ can be attributed to the fact that the $S \cdots C-O$ angle can be in the optimum range when ψ values are between -100 and -50° , a zone that is sparsely occupied in the plot of Ramachandran and coworkers (1963). Although the $S \cdots C-O$ angle gives a simplistic representation of the relative orientation of the two interacting groups, the stereochemistry can be best described using the

spherical polar angles, θ and μ . The points in the region of $(160, 60)$ tend to have a larger absolute value of μ (Table 1), which results in a larger deviation of the S atom from the tetrahedral direction (109° , relative to the carbonyl group) and a position closer to the carbonyl oxygen atom (carrying a partial negative charge). This may explain why this region is less dense.

In Figure 2 it has been shown that only those χ_1 values are allowed that provide shorter contact between S and C atoms. Now for a given χ_1 , only those ψ values are occupied that can maintain the optimum $S \cdots C-O$ and μ angles, giving credence to the nucleophilic and electrophilic characters of the two interacting groups.

S \cdots C=O interaction vs. hydrogen bonding

One might argue that the geometry observed for the $S \cdots C=O$ interaction is not the result of any attractive force existing between the two groups, but is really imposed by the steric requirement of the previously documented (Adman et al., 1975) $NH \cdots S$ hydrogen bond between the cysteine sulfur atom and the peptide NH group two residues following it. To investigate this we have presented in Table 3 the hydrogen bond partners (within a rather long distance limit of 3.8 \AA) of all the Cys residues discussed in Tables 1 and 2. Out of a total of 82 Cys residues twelve partake in no hydrogen bond at all, and a further twelve have only side-chain

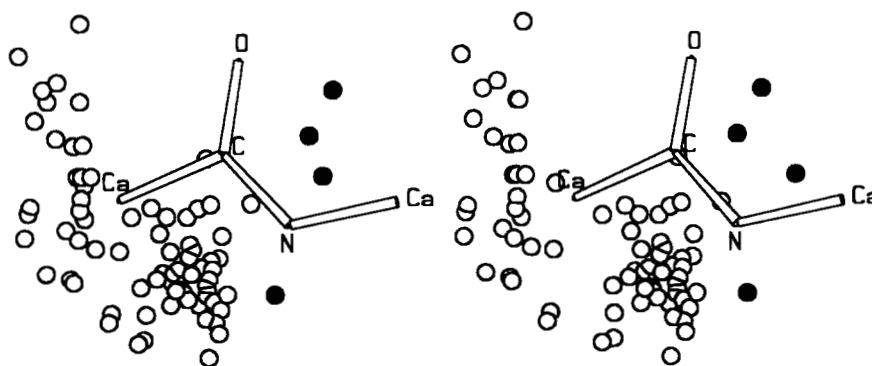


Fig. 4. Disposition of the sulfur atoms (Table 1 examples are represented by open circles, Table 2 by filled circles) with respect to the interacting CO group, shown in stereo.

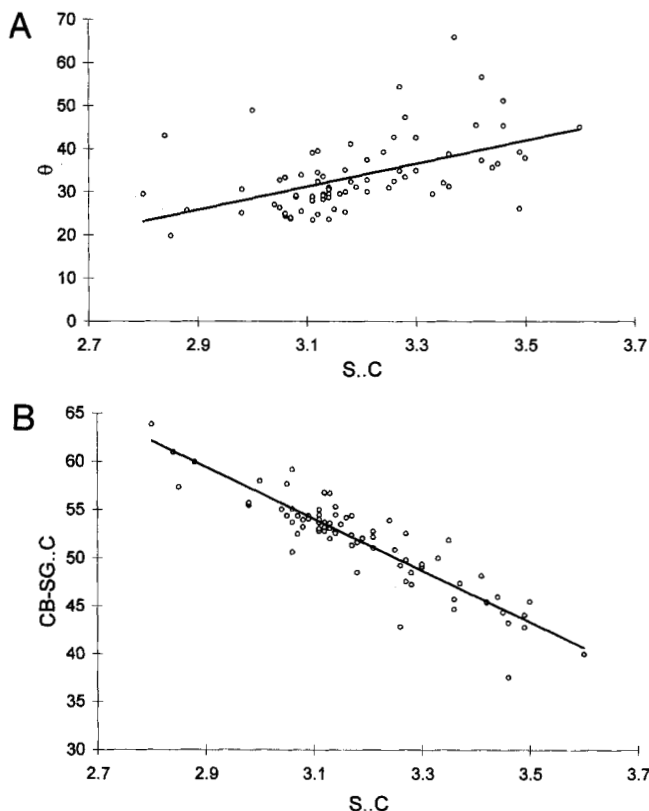


Fig. 5. Plot of S...C distance; Å vs. (A) θ (if the angle is greater than 90° in Table 1, it is made acute by subtracting from 180°) and (B) CB-SG...C angles ($^\circ$). In A the best-fitted straight line has an equation $y = 26.9x - 52.1$ and a correlation coefficient of 0.51; in B they are $y = -26.8x + 137.0$ and -0.90 , respectively.

atoms or water molecules as partners. Of the remaining 58 residues (at position i), the proton donor is a peptide NH group (at position $i + 2$) in 45 cases, which tend to have other hydrogen bond interactions as well. Also, not all these interactions have the optimum geometry, for example, 5 out of the 45 examples mentioned above deviate from a linear X-H...S angle orientation by more than 50° . Thus, even when it is present there is no common hydrogen bonding pattern encompassing all the Cys residues which, nevertheless, maintain a very consistent S...C=O geometry, thus pointing to the importance and the stability of this interaction vis-a-vis hydrogen bonding.

Implications

Various stabilizing interactions in protein structures, like hydrogen bonding or salt bridges, involve the participation of protons. Here we have identified an interaction that does not involve any protons, but requires the placement of a S atom (from a metal bound Cys residue) onto the top of the carbonyl C atom of the peptide plane. In the majority of the cases the S and C atoms are from the same Cys residue. Attractive interaction between different residues is believed to hold the protein structure together (Karlin et al., 1994). Inasmuch as most of the nucleophile-electrophile pairs used in this analysis are from the same Cys residue it can be surmised that atoms belonging to the same residue, as long as they are brought to each other in the right orientation in the folded state, can con-

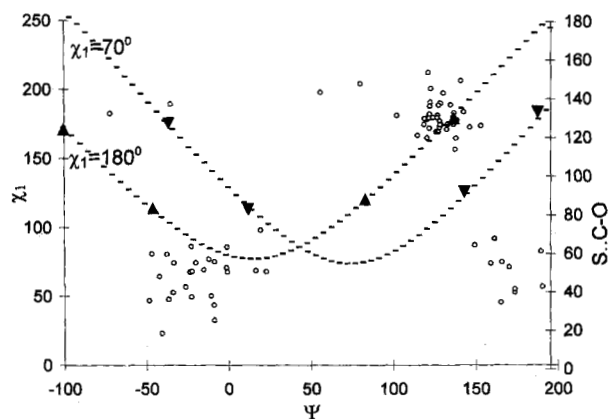


Fig. 6. Joint distribution of the torsion angles ($^\circ$) χ_1 and ψ . The two curves represent the variation of the S...C-O angle ($^\circ$) (right side ordinate) with ψ , keeping χ_1 fixed at 70° (to represent the g^- state for which the peak shows a shift toward 70° in Fig. 2) and 180° in a model peptide moiety. Over the ψ -space the S...C-O angle lies in the range 90 – 125° in two regions which are approximately indicated by the two pairs of markers. The triangles point in the direction of the four zones in the plot that have been discussed in the text.

tribute to the overall stability. In this light it is instructive to look at the structure of ferredoxin (Adman et al., 1976). Although the regular secondary structures are supposed to be the stable building blocks of a protein, ferredoxin is almost devoid of such structures. In this connection it is noteworthy that the structure has two Fe_4S_4 clusters, and all eight ligands are in a conformation (four g^- and four t) that can sustain nucleophile-electrophile interaction. Similarly, in alcohol dehydrogenase the structural zinc site has four ligands, all Cys residues at positions 97, 100, 103, and 111; all but one of them show S...C=O interaction (Table 1) pointing to the usefulness of the metal coordination by Cys residues.

In our discussion so far, although we have considered metal-bound Cys residues, the metal atoms have been confined to the background. However, the S...C=O interaction could be of utmost importance to the function of metalloproteins because of possible charge delocalization from the metal center to the carbonyl group through the intervening ligand S atom. This then, in addition to hydrogen bonding (Adman et al., 1975), can modulate the redox potential of the metal center. Moreover the overlap between the metal and Cys S orbitals (Solomon et al., 1983; Chakrabarti, 1989; Han et al., 1991), and then S and peptide orbitals as discussed here provides an elegant way of electron transfer from the metal center. For example, in plastocyanin (Collyer et al., 1990), of the two paths of electron to and from the Cu atom, one was identified to involve Tyr 83 (Colman et al., 1978; Sykes, 1991). As shown in Figure 7, the atoms Cu, S, and C (the last two belonging to the ligand Cys 84) are close to linear, and the distances involving various atoms (especially C and N) in the Cys 84-Asp 85 peptide group and the Tyr ring center are much shorter than the Cu...Tyr length. Essentially, this is a "through space" (rather than "through-bond") pathway, and yet there is sequential overlap between various orbitals along the way to direct the electron transfer. Although a considerable number of free (i.e., non-ligand) Cys in various proteins also exhibit this interaction (52%, as compared to 87% found here) (Pal & Chakrabarti, to be published), it is in metalloproteins that it is likely to have the most profound implications for the protein function.

Table 3. Hydrogen bond interactions involving the thiolate SG stom^a

Code	Residue	Residue hydrogen-bonded, atom name(X) (distance X...S(Å), angle X—H...S(°))	Code	Residue	Residue hydrogen-bonded, atom name(X) (distance X...S(Å), angle X—H...S(°))
2OHX	46	48N (3.44,163), 67NE2 (3.65,132)	2PIA	272	274N (3.33,163), 275N (3.46,121), 276N (3.39,162), 276OG1 (3.47), 277N (3.30,149)
	97	99N (3.34,153), 100N (3.27,141), 113NZ (3.46)		277	271OG (3.16), 272N (3.72,168), 279N (3.51,163), 279OG (3.33)
	100	102N (3.60,127), 103N (3.36,164), 112N (3.54,173)		280	308N (3.53,155)
	111	97N (3.44,165), 113N (3.41,171)	1AOZ	507	509N (3.46,168), 512ND1 (3.67,135)
	174	67NE2 (3.55,133)	8RXN	6	8N (3.54,147), 9N (3.55,168)
1DCA	199	94NE2 (3.74,154), 96NE2 (3.62,140), 314Ow (3.71)		9	11N (3.42,161)
4MT2	5	7N (3.49,157), 25N (3.33,176), 124Ow (3.46)		39	41N (3.57,144), 42N (3.62,155)
	13	15N (3.19,145)		42	44N (3.50,152)
	19	31NZ (3.54)	1ISU	25	27N (3.42,162), 59N (3.54,162)
	24	26N (3.65,162)		40	42N (3.39,144)
	29	17N (3.02,135), 71Ow (3.18)		55	57N (3.48,151)
	33	86Ow (2.99), 96Ow (3.20)	1FRR	38	40N (3.44,161), 42N (3.51,159)
	34	37N (3.71,120)		43	37OG (3.45), 45N (3.33,164), 45OG (3.51)
	36	120Ow (3.05)		46	76N (3.54,163)
	41	101Ow (3.16), 137Ow (3.41)		76	41N (3.62,131)
	48	50N (3.59,148)	1FDX	8	10N (3.63,144), 28N (3.37,164), 13N (3.59,156)
	57	60N (3.15,144), 106Ow (3.23)		11	2N (3.46,160), 37N (3.34,156)
1ZAA	7	9N (3.71,161)		35	40N (3.36,142)
	12	14N (3.43,154), 25NE2 (3.57,137) 29NE2 (3.67,160), 421Ow (2.86)		38	47N (3.69,161), 49N (3.60,160)
	37	39N (3.44,142), 40N (3.52,165), 53NE2 (3.54,123)		45	258ND2 (3.31,159)
	40	42N (3.73,154), 53NE2 (3.52,142), 57NE2 (3.47,145)	8ACN	358	79Ow (3.74), 320Ow (3.20)
	65	67N (3.37,137), 68N (3.32,165), 81NE2 (3.54,121)		424	147NE (3.78,153), 147NH2 (3.61,162)
	68	70N (3.57,150), 81NE2 (3.38,147), 85NE2 (3.52,143), 415Ow (3.30)	1ABK	187	199N (3.56,153)
1GLU	440	442N (3.54,157), 443N (3.61,125), 457N (3.65,132)		197	205N (3.35,153), 206N (3.38,159)
	457	459OG (2.56)		203	347N (3.37,159)
	476	478N (3.30,136), 480N (3.51,121)	2TMD	345	350N (3.68,167), 365OG1 (3.05)
	492	495N (3.45,136)		348	366N (3.53,124), 367N (3.25,160)
8ATC	109	111N (3.53,149)		364	99N (3.70,121)
	138	140N (3.69,143), 141N (3.70,158), 178Ow (3.06)	1NIP	97	134N (3.36,145)
7PCY	84	37ND1 (3.57,145), 38N (3.51,157)		132	64N (3.50,129), 65N (3.49,158)
2AZA	112	47N (3.52,171), 114N (3.50,161), 117ND1 (3.60,128)	1MIN	62	278N (3.42,148), 278OG (2.96), 91OH (3.17)
1PAZ	78	41N (3.61,162), 81N (3.76,152), 81ND1 (3.54,131)		70	188OG (3.49)
1AAN	92	53ND1 (3.79,143), 54N (3.66,165) 95ND1 (3.38,135)	1GLU	495	482N (3.57,171)
2CPP	357	359N (3.25,126)			
1AFN	136	96N (3.63,167)			

^aCys residues are presented in the sequence they occur in Tables 1 and 2 (but the subunit name has been removed). For the hydrogen bond donor NH group (peptide, histidine), the hydrogen has been placed at the stereochemically expected position. For the other XH groups, the proton cannot be fixed unambiguously; for these there is no entry for the parameter X—H...S angle. The label Ow stands for a water molecule. The criteria used for identifying hydrogen bonds: the distance X...S less than 3.8 Å, and the angle X—H...S (in cases where the H position can be calculated) greater than 120°. The cut-off distance for distinguishing a hydrogen bond is more lenient than what has been discussed in Methods. Six carbonyl oxygen and two carboxylate oxygen atoms, found within this limit, have been omitted from the table as they cannot donate any proton to the sulfur atom.

The apparent pK_a of the active site thiol in many enzymes are lowered considerably (Jeng et al., 1995; Zhang et al., 1995). This reduction is usually attributed to the stabilization of the thiolate by the surrounding residues. Our findings suggest that a CO group, if suitably juxtaposed near the S atom, can help to delocalize the negative charge, and thus lower the pK_a of the thiol group.

The χ_1 preferences of residues are known to depend on the secondary structure they are embedded in (McGregor et al., 1987). Here we have shown that, irrespective of the secondary structure, the direct interaction between the main-chain and side-chain atoms of a residue can modify the conformational preferences of χ_1 , as well as of ψ . The Ramachandran plot is the embodiment of various

nonbonded interactions in a polypeptide chain. However, for metal-bound Cys, the ψ values appear to be controlled by an additional potential that confines the S...C—O angle to be in the range of 90–125°.

Finally, as the interaction is likely to be quite common in metalloprotein structures, the existing force-field parameters and the crystallographic refinement programs should take cognizance of the fact that the S and C atoms can constitute a cohesive interaction that may result in pulling the C atom out of the plane of its bonded neighbors.

In summary, we have identified a new interaction in protein structures that manifests in having S and carbonyl C atoms in close

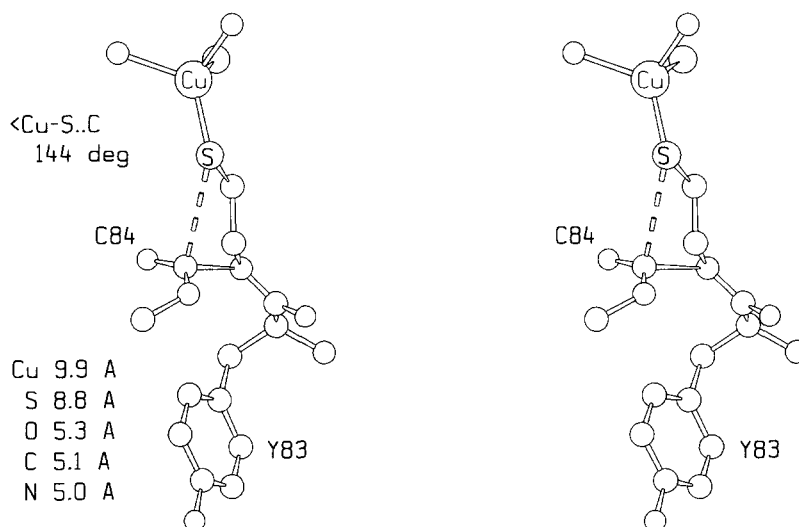


Fig. 7. Stereoview showing the disposition of the ligand Cys 84 and Tyr 83 relative to Cu and its three other bonded atoms, in plastocyanin. The dashed line represents the $\text{S}\dots\text{C}$ interaction. Distances involving the centroid of the aromatic ring and a few relevant atoms, as also a bond angle are marked. The angle between the planes of the Cys 84–Asp 85 peptide group and Tyr 83 aromatic ring is 88° .

contact, mostly within the same residue, thereby affecting the distribution of the side-chain as well as the main-chain conformations. The interaction between an electrophile and a nucleophile drives many chemical reactions. Nature can also use this type of interaction to impart stability to the protein structure.

Methods

Brookhaven Protein Data Bank (PDB) (Bernstein et al., 1977), version April 1994, was searched for Cys residues coordinated to any metal ion in refined crystal structures. When more than one structure of the same protein was available, the one with the best resolution and/or R-factor was used, irrespective of the organism from which the protein was derived. For $[\text{Fe}_4\text{S}_4]$ ferredoxin, the structure with two clusters was used as it offered the maximum number of ligands. For multimeric proteins (or when the crystal asymmetric unit had more than one independent molecule), only one subunit was considered.

Structures were displayed centered at the metal site on a graphics workstation, and all the carbonyl C atoms within 3.6 \AA from the Cys S (or SG, as the atom is labeled in PDB) were recorded. The reason for choosing this cut-off distance was the earlier observation (Ippolito et al., 1990) that hydrogen-bonded neighbors are found at a distance of $3.5 (\pm 1) \text{ \AA}$ from the S atom. Besides carbonyl C atoms and hydrogen-bonded partners (usually a main-chain NH or water) (Adman et al., 1975), we rarely encountered any other atom type at such close distances. Other relevant geometric parameters, such as the virtual bond angles, $\text{SG}\dots\text{C}-\text{O}$ and $\text{CB}-\text{SG}\dots\text{C}$, torsion angles, χ_1 ($\text{N}-\text{CA}-\text{CB}-\text{SG}$ as an indicator of the side-chain conformation), and $\chi_2(\text{C})$ ($\text{CA}-\text{CB}-\text{SG}\dots\text{C}$, which defines the position of the interacting CO group relative to the Cys residue) were measured. To view the interaction from the perspective of the CO group, the relative position of S was expressed in terms of the spherical polar angles (θ and μ) defined in Figure 1. The plots given in Figures 1, 4, and 7 have been created using SCHAKAL (Keller, 1992).

Acknowledgments

It is a pleasure to express our gratitude to Prof. J.D. Dunitz for his critical comments on the manuscript, both scientific as well as stylistic. Dr. Vinay Kumar was kind enough to bring to our notice a review article. This work was made possible by a grant from the Department of Science and Technology (India). D.P. was supported by a fellowship from the Council of Scientific and Industrial Research (India).

References

- Adman E, Watenpaugh KD, Jensen LH. 1975. $\text{NH}\dots\text{S}$ hydrogen bonds in *Peptococcus aerogenes* ferredoxin, *Clostridium pasteurianum* rubredoxin, and *Chromatium* high potential iron protein. *Proc Natl Acad Sci USA* 72:4854–4858.
- Adman ET, Sieker LC, Jensen LH. 1976. Structure of *Peptococcus aerogenes* ferredoxin. *J Biol Chem* 251:3801–3806.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The protein data bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535–542.
- Bondi A. 1964. Van der Waals volumes and radii. *J Phys Chem* 68:441–451.
- Bürgi HB, Dunitz JD. 1983. From crystal statics to chemical dynamics. *Acc Chem Res* 16:153–161.
- Bürgi HB, Dunitz JD, Lehn JM, Wipff G. 1974a. Stereochemistry of reaction paths at carbonyl centers. *Tetrahedron* 30:1563–1572.
- Bürgi HB, Dunitz JD, Shefter E. 1973. Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J Am Chem Soc* 95:5065–5067.
- Bürgi HB, Dunitz JD, Shefter E. 1974b. Chemical reaction paths. IV. Aspects of $\text{O}\dots\text{C}=\text{O}$ interactions in crystals. *Acta Crystallogr B* 30:1517–1527.
- Bürgi HB, Lehn JM, Wipff G. 1974c. An ab initio study of nucleophilic addition to a carbonyl group. *J Am Chem Soc* 96:1956–1957.
- Burley SK, Petsko GA. 1985. Aromatic–aromatic interaction: A mechanism of protein structure stabilization. *Science* 229:23–28.
- Burley SK, Petsko GA. 1986. Amino–aromatic interactions in proteins. *FEBS Lett* 203:139–143.
- Chakrabarti P. 1989. Geometry of interaction of metal ions with sulfur-containing ligands in protein structures. *Biochemistry* 28:6081–6085.
- Chakrabarti P, Samanta U. 1995. CH/π interaction in the packing of the adenine ring in protein structures. *J Mol Biol* 251:9–14.
- Collyer CA, Guss JM, Sugimura Y, Yoshizaki F, Freeman HC. 1990. The crystal structure of plastocyanin from a green algae, *Enteromorpha prolifera*. *J Mol Biol* 211:617–632.
- Colman PM, Freeman HC, Guss JM, Murata M, Norris VA, Ramshaw JAM, Venkatappa MP. 1978. X-ray crystal structure analysis of plastocyanin at 2.7 \AA resolution. *Nature(Lond)* 272: 319–324.

- Derewenda Z, Lee L, Derewenda U. 1995. The occurrence of C—H...O hydrogen bonds in proteins. *J Mol Biol* 252:248–262.
- Dunitz JD. 1979. *X-ray analysis and the structure of organic molecules*. Ithaca, New York: Cornell University Press.
- Fujimoto H. 1987. Paired interacting orbitals: A way of looking at chemical interactions. *Acc Chem Res* 20:448–453.
- Han J, Adman ET, Beppu T, Codd R, Freeman HC, Huq L, Loehr TM, Sanders-Loehr J. 1991. Resonance Raman spectrum of plastocyanin and pseudoazurin: Evidence for conserved cysteine ligand conformations in cupredoxins (blue copper proteins). *Biochemistry* 30:10904–10913.
- Honig B, Yang AS. 1995. Free energy balance in protein folding. *Adv Prot Chem* 46:27–58.
- Ippolito JA, Alexander RS, Christianson DW. 1990. Hydrogen bond stereochemistry in protein structure and function. *J Mol Biol* 215:457–471.
- Janin J, Wodak S, Levitt M, Maigret B. 1978. Conformation of amino acid side-chains in proteins. *J Mol Biol* 125:357–386.
- Jeng M, Holmgren A, Dyson HJ. 1995. Proton sharing between cysteine thiols in *Escherichia coli* thioredoxin: Implications for the mechanism of protein disulfide reduction. *Biochemistry* 34:10101–10105.
- Karlin S, Zuker M, Brocchieri L. 1994. Measuring residue associations in protein structures. Possible implications for protein folding. *J Mol Biol* 239:227–248.
- Keller E. 1992. SCHAKAL 92. A computer program for the graphic representation of molecular and crystallographic models. Crystallography Institute, University of Freiburg, Germany.
- Liotta CL, Burgess EM, Eberhardt WH. 1984. Trajectory analysis. I. Theoretical model for nucleophilic attack at π -systems. The stabilizing and destabilizing orbital terms. *J Am Chem Soc* 106:4849–4852.
- Maccallum PH, Poet R, Milner-White EJ. 1995. Coulombic interactions between partially charged main-chain atoms not hydrogen-bonded to each other influence the conformation of α -helices and antiparallel β -sheet. A new method for analysing the forces between hydrogen-bonding groups in proteins includes all the coulombic interactions. *J Mol Biol* 248:361–373.
- Madura JD, Jorgensen WL. 1986. Ab initio and Monte Carlo calculations for a nucleophilic addition reaction in the gas phase and in aqueous solution. *J Am Chem Soc* 108:2517–2527.
- McGregor MJ, Islam SA, Sternberg MJE. 1987. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J Mol Biol* 198:295–310.
- Menger FM. 1985. On the source of intramolecular and enzymatic reactivity. *Acc Chem Res* 18:128–134.
- Menger FM. 1987. Nucleophilicity and distance. *Adv Chem Series (Am Chem Soc)* 215:209–218.
- Myers LC, Terranova MP, Ferentz AE, Wagner G, Verdine GL. 1993. Repair of DNA methylphosphotriesters through a metalloactivated cysteine nucleophile. *Science* 261:1164–1167.
- Perutz MF. 1993. The role of aromatic rings as hydrogen-bond acceptors in molecular recognition. *Phil Trans Roy Soc Ser A* 345:105–112.
- Ramachandran GN, Ramakrishnan C, Sasisekharan V. 1963. Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7:95–99.
- Reid KSC, Lindley PF, Thornton JM. 1985. Sulphur-aromatic interactions in proteins. *FEBS Lett* 190:209–213.
- Solomon EI, Penfield KW, Wilcox DE. 1983. Active sites in copper proteins. An electronic structure overview. *Struct Bonding* 53:1–57.
- Sykes AG. 1991. Plastocyanin and the blue copper proteins. *Struct Bonding* 75:175–224.
- Wilker JJ, Lippard SJ. 1995. Modeling the DNA methylphosphotriester repair site in *Escherichia coli* ada. Why zinc and four cysteines? *J Am Chem Soc* 117:8682–8683.
- Zhang Z, Palfey BA, Wu L, Zhao Y. 1995. Catalytic function of the conserved hydroxyl group in the protein tyrosine phosphatase signature motif. *Biochemistry* 34:16389–16396.