

Weak Nonbonded S···X (X = O, N, and S) Interactions in Proteins. Statistical and Theoretical Studies.

Michio Iwaoka,* Shinya Takemoto, Mai Okada, and Shuji Tomoda*

Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902

(Received January 25, 2002)

Sulfur-containing functional groups of cystine (an SSC group) and methionine (a CSC group) are usually considered as hydrophobic moieties or weak hydrogen-bond acceptors in folded protein structures. However, database analysis as well as theoretical calculations carried out in this study have provided strong evidence for the presence of specific nonbonded interactions between the divalent sulfur atoms (S) and nearby polar non-hydrogen atoms (X). Close S···X (X = O, N, S, C, etc.) atomic contacts were statistically analyzed in 604 high-resolution heterogeneous X-ray structures selected from a protein databank (PDB_SELECT). The S···O interactions found for both SSC and CSC groups showed a specific character as a π(C=O) $\rightarrow$ σ^*(S) orbital interaction based on the directional preferences. The interactions were most frequently observed in α-helices. Ab initio calculations applying the second order Møller–Plesset perturbation theory (MP2) suggested the primary importance of electron correlations. The total stabilization energies were calculated to be $\sim$3.2 and $\sim$2.5 kcal/mol for SSC and CSC groups, respectively, including the contribution from a coexisting CH···O hydrogen bond. On the other hand, the S···N interactions observed for a CSC group exhibited structural characters as a π(N) $\rightarrow$ σ^*(S) orbital interaction and an NH···S hydrogen bond, and the S···S interactions for an SSC group showed a structural character as an n(S) $\rightarrow$ σ^*(S) orbital interaction. The S···C(π) interactions should be rather weak and long-range.

Nonbonded atomic interactions play significant roles in determining protein structures and their folding pathways. Conventional hydrogen bonds (such as NH···O and OH···O), ionic salt bridges, and van der Waals forces (both attractive and repulsive) are usually counted as such weak interactions.[1–5] However, with the data of precise protein structures now available, some unusual or previously unexpected atomic interactions, such as non-conventional CH···O hydrogen bonds,[6–8] cation···π interactions,[9–11] and CH···π interactions,[12] have been characterized recently. In this paper, nonbonded S···X (X = O, N, and S) interactions[13,14] are suggested to be another member of such non-classical interactions that weakly stabilize protein structures.

Sulfur-containing functional groups of cystine (an SSC group) and methionine (a CSC group) are normally considered as hydrophobic moieties or weak hydrogen-bond acceptors in folded protein structures. However, specific attractive interactions between a divalent sulfur atom (S) and nearby non-hydrogen atoms (X) have been well recognized in small organic compounds. A conventional mechanism for formation of a nonbonded S···X interaction is shown in Fig. 1.[15] Rosenfield et al.[16] surveyed close S···X contacts in the Cambridge crystallographic database and found that the directional preferences of X with respect to S can be explained by the presence of nonbonded S···X interactions and XH···S hydrogen bonds. Later, Row and Parthasarathy[17] indicated that nonbonded S···S interactions may be stabilized by the orbital interaction between the lone pair of one sulfur atom [n(S)] and the anti-bonding orbital of the other sulfur atom [σ^*(S)]. A similar donor-acceptor interaction mechanism of S···S interactions was also reported.[18] On the other hand, Dahaoui et al.[19] pointed out the importance of a van der Waals force for the S···S interactions. The presence of S···C(π) interactions in organic crystals was also suggested.[20]

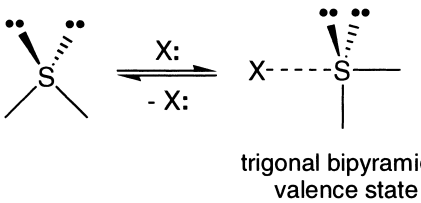

Fig. 1. A conventional mechanism for formation of nonbonded S···X interaction.

For some organic sulfur compounds, intramolecular nonbonded S···O interactions have been extensively studied in relation to the biological activities. For example, the intramolecular 1,4-S···O interaction of thiazole nucleoside analogues was found to be important for the antitumor activity.[21] Similarly, 1,5-S···O interactions were shown to play important roles in the antagonism of (acylimino)thiadiazoline derivatives to an angiotensin II receptor[22] and in the antitumor activity of leinamycin.[23] Analogous interactions between a divalent selenium atom (Se) and X (i.e., Se···X interactions) have also been well characterized in organic selenium compounds[24–26] and have been suggested to be relevant to the strong biological activity of selenium.[27–30]

In protein structures, however, nonbonded S···X interactions

have not been noticed until recently, except for S···C(π) interactions and weak NH···S and OH···S hydrogen bonds. The S···C(π) interactions in proteins were first pointed out by Morgan and co-workers,[31] who analyzed the close atomic contact between S and a π-plane in eight protein structures and found that the S atoms have a propensity to come over the π-plane. On the other hand, Reid et al.[32] suggested by using a larger set of protein structures that the close S···C(π) contact in proteins should not be explained by direct S···C(π) interactions but by CH···S interactions because the S atoms access to the π-plane from the side rather than the top. Although many experimental and theoretical studies have been reported to date,[33,34] the nature of S···C(π) interactions is still controversial.[20] As for NH···S and OH···S hydrogen bonds in proteins, it was suggested that the S atoms of cystine and methionine have only a weak character of a hydrogen-bond acceptor.[35]

For two particular proteins, the importance of an S···O interaction on the enzymatic functions has been pointed out very recently. Taylor and Markham[36] suggested that the electrostatic S···O interaction between the S atom of a methionine substrate and the carboxylate O atom of Asp118 plays a major role in the enzymatic activity of *S*-adenosylmethionine synthetase. Brandt et al.[37] reported that the cleavage of a disulfide bond in the extracellular region of the G-protein receptors, which is an important process for the receptor activation, is catalyzed by the S···O interaction between Cys121 and the carboxylic group of Asp288.

Stereochemistry of the close S···X contacts involving a methionine sulfur atom (a CSC group) was statistically analyzed in protein structures, but distinct directional preferences of the contacts were not observed.[38] It was therefore suggested that nonbonded S···X interactions in proteins are either very weak or physicochemically different from those in small molecules. However, using a larger set of protein structures (PDB_SELECT),[39,40] we have recently demonstrated that the S···X contacts in proteins exhibit obvious directionality for both SSC and CSC groups.[13] Similar directionality of the S···O interactions involving a methionine S atom was reported by Pal and Chakrabarti as well.[14]

In this paper, we present details of our previous database analysis of nonbonded S···X (X = O, N, and S) interactions.[13] In addition, we have carried out high-level quantum chemical calculations to evaluate the strength of the S···O and S···N interactions. The stabilization mechanism as well as the influence of coexisting hydrogen bonds are discussed based on the results from statistical database analysis and theoretical calculations. The S···X interactions characterized here may play roles in the functions, stability, and folding of proteins to some extent. Design of specific substrates to cleave a protein disulfide bond may also be possible by applying the novel interactions.

Method

Database Analysis. According to the PDB_SELECT database (the structural homology less than 25%),[39,40] 1085 nonredundant X-ray structures, which contained 604 high resolution (≤ 2.0 Å) and 481 low resolution (> 2.0 Å) structures, were retrieved from a protein data bank (PDB).[41] A computer program was coded for extracting all non-hydrogen atoms (X) that have a close atomic contact to a divalent sulfur atom (S) of cystine (an SSC group) or methionine (a CSC group). The data of the S···X contacts obtained were sorted by the use of structural parameters shown in Fig. 2. The S···X contacts within a single amino acid residue and the ones out of the ranges of $\theta_1 \geq 50°$ and $\theta_2 \geq 50°$ were omitted automatically.

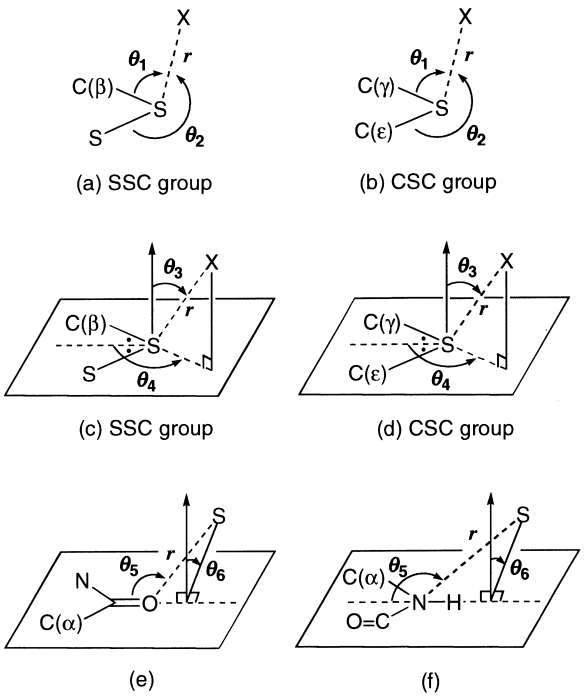

Fig. 2. Structural parameters of S···X contacts used for the statistical analysis.

In order to compare the degrees of various S···X atomic contacts with each other, a relative nonbonded distance d [$\equiv r - vdw(\mathrm{S}) - vdw(\mathrm{X})$, where $vdw(i)$ means the van der Waals radius of atom i] was also defined. The values of van der Waals radii reported by Bondi[42] were employed: 1.70 Å for C, 1.55 Å for N, 1.52 Å for O, 1.80 Å for S, etc. For Fe, the van der Waals radius reported by Allinger et al.[43] was used after the scaling [$vdw(\mathrm{Fe}) = 1.70$ Å]. For the case of S···C interactions, the distance between the S atom and the centroid of the aromatic ring (r), the vertical and horizontal distances of the S atom relative to the aromatic plane were also calculated. Angles θ_1 and θ_2 (or θ_3 and θ_4) were utilized to indicate the spatial location of X with respect to the SSC or CSC plane. We employed two different sets of angles in order to show directionality of S···X interactions in proteins more clearly. On the other hand, angles θ_5 and θ_6 were utilized to indicate the spatial location of S with respect to the X (O or N) atom.

Calculations. The Gaussian 98 program[44] was used for all quantum chemical calculations carried out in this study. The basis sets implemented in the program were employed without modification. Electron correlation energies were calculated by applying the second order Møller-Plesset perturbation theory (MP2).[45,46] Complexation energies were corrected for the basis set superposition error (BSSE) by using the counterpoise method.[47] Natural bond orbital analysis[48] was performed by the use of the NBO ver. 3.1 program linked to the Gaussian 98 program. The analysis, which has been widely used to study

wave functions and electron densities obtained from ab initio calculations, allows quantitative evaluation of the orbital interaction energies for any combinations between fully occupied localized orbitals (core, lone pair, and bonding) and vacant orbitals (anti-bonding and Rydberg).

Stable structures of two model complexes, CH_3SSCH_3 + $CH_3CONHCH_3$ and CH_3SCH_3 + $CH_3CONHCH_3$, were exhaustively sought by applying various initial values of access angles θ_3–θ_6 at the HF/6-31G(d) level. Seven and six structures were located for the CH_3SSCH_3 and CH_3SCH_3 complexes, respectively. They were subsequently allowed for geometry optimization at the higher calculation level [MP2/6-31G(d)]. Some structures were further optimized at the MP2/6-311G(d,p) level, but significant structural changes were not observed. We therefore utilized the MP2/6-31G(d) level of theory to obtain the stable structures of the complexes.

Potential surfaces of the complexes were obtained as follows. The structures of CH_3SSCH_3, CH_3SCH_3, and $CH_3CONHCH_3$ were optimized separately at the MP2/6-31G(d) level. By using the resulting structures, various complex structures, which have fixed values of d (−0.22, −0.02, 0.18, 0.38, 0.58 Å), θ_3 (0°, 30°, 60°, 90°, 120°, 150°, 180°), θ_4 (90°, 120°, 150°, 180°, 210°, 240°, 270°), θ_5 (60°, 90°, 120°, 150°, 180°), and θ_6 (0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, 330°), were computationally generated. For each structure, the complexation energy was calculated at the MP2/6-31G(d) level without optimization of the geometry. It should be noted that the fully optimized structure of the complex is different from the complex structure of the energy minimum on the potential surface. This caused a slight discrepancy between the complexation energy for the fully optimized structure (3.21 kcal/mol for Structure **A**) and that obtained from the potential surfaces (∼3 kcal/mol).

Results and Discussion

Effect of Resolution. In the process of determining protein structures from the X-ray diffraction data, some of the structural parameters are usually restrained to the empirical values:[49] the extent of the restraints depends on the resolution of diffraction patterns. Therefore, unexpected interactions could be more frequently observed in high-resolution protein structures than in low-resolution ones. The statistical data listed in Tables 1 and 2 show that the S···O interaction may be such a case.

For the SSC group (Table 1), the ratio of the nonbonded S···O contact to the total S···X (X = O, N, S, C, etc.) contacts in the range of $d \leq 0.0$ Å was 35.2% in the high-resolution (≤ 2.0 Å) class, which was about 9% larger than that (26.1%) in the low-resolution (< 2.0 Å) class. For the CSC group (Table 2), the ratio of the short ($d \leq 0.0$ Å) S···O contact in the high-

Table 1. Summary of S···X Contacts Observed for the SSC Group in Proteins

Class	High resolution (≤ 2.0 Å)		Low resolution (> 2.0 Å)	
Number of cystine residues	790		448	
Ranges of d[a)]	$d \leq 0.0$ Å	$0.0 < d \leq 0.5$ Å	$d \leq 0.0$ Å	$0.0 < d \leq 0.5$ Å
Number of S···X contacts[b)]				
total	284 (100.0)	2569 (100.0)	326 (100.0)	1213 (100.0)
X = O	100 (35.2)	664 (25.8)	85 (26.1)	261 (21.5)
N	33 (11.6)	359 (14.0)	37 (11.3)	167 (13.8)
S	15 (5.3)	68 (2.6)	27 (8.3)	20 (1.7)
C	134 (47.2)	1478 (57.5)	177 (54.3)	765 (63.1)
others[c)]	2 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)

a) d [$\equiv r - vdw(S) - vdw(X)$] represents a nonbonded S···X distance relative to the sum of the van der Waals radii.
b) The values in parentheses mean the percentages to the total number of S···X contacts.
c) Fe and K are involved.

Table 2. Summary of S···X Contacts Observed for the CSC Group in Proteins

Class	High resolution (≤ 2.0 Å)		Low resolution (> 2.0 Å)	
Number of methionine residues	2124		2222	
Ranges of d[a)]	$d \leq 0.0$ Å	$0.0 < d \leq 0.5$ Å	$d \leq 0.0$ Å	$0.0 < d \leq 0.5$ Å
Number of S···X contacts[b)]				
total	457 (100.0)	5049 (100.0)	606 (100.0)	3988 (100.0)
X = O	153 (33.5)	692 (13.7)	123 (20.3)	507 (12.7)
N	68 (14.9)	474 (9.4)	90 (14.9)	419 (10.5)
S	9 (2.0)	66 (1.3)	14 (2.3)	72 (1.8)
C	209 (45.7)	3815 (75.6)	372 (61.4)	2990 (75.0)
others[c)]	18 (3.9)	2 (0.0)	7 (1.2)	0 (0.0)

a) d [$\equiv r - vdw(S) - vdw(X)$] represents a nonbonded S···X distance relative to the sum of the van der Waals radii.
b) The values in parentheses mean the percentages to the total number of S···X contacts.
c) Fe, Cu, Zn, Hg, Pt, and Cl are involved.

resolution class (33.5%) was about 13% larger than that in the low-resolution class (20.3%). These distinct trends strongly suggested the presence of specific nonbonded interactions between S and O atoms. Similar trends were observed for the long ($0.0 < d \leq 0.5$ Å) S···O contacts for the SSC and CSC groups.

In contrast, the ratio of the S···C contact in the high-resolution class was significantly lower than that in the low-resolution class for both SSC and CSC groups. The observation suggested that S···C(π) interactions in proteins[31,32] may be weaker than S···O interactions. On the other hand, the ratios of S···N and S···S contacts were not affected by the resolution class. Since the frequencies of these contacts in protein structures were less than those of S···O and S···C contacts, the presence of the S···N and S···S interactions could not be assessed based only on the small resolution effects observed.

Effect of a Relative S···X Distance (*d*). A nonbonded S···X distance relative to the sum of the van der Waals radii (d) can be a convenient measure of the strength of the S···X contact. When a particular S···X interaction is attractive and short-range, the ratio of the corresponding S···X contact would increase in the short contact range. Table 1 shows that, in the high-resolution class, the ratios of S···O (35.2%) and S···S (5.3%) contacts for an SSC group in the short ($d \leq 0.0$ Å) contact range are significantly larger than those in the long ($0.0 < d \leq 0.5$ Å) contact range (25.8% and 2.6%, respectively). Similar trends were observed for the data of the low-resolution class. The observations suggested the presence of S···O and S···S interactions for an SSC group in proteins. For a CSC group (Table 2), a significant increase of the ratios of S···O and S···N contacts was observed in both high-resolution and low-resolution classes when the contacts were short-range. It was, therefore, suggested that the S···O and S···N interactions for a CSC group may exist in proteins. These interactions are analyzed in detail in the following subsections.

As for the S···C contacts, the ratios to the total S···X contacts in the short contact range ($d \leq 0.0$ Å) were smaller than those in the long contact range ($0.0 < d \leq 0.5$ Å) for both SSC and CSC groups. The trend suggested that S···C(π) interactions in proteins[31,32] may be much weaker than S···O, S···N, and S···S interactions or may work only at long-range. The details of the observed S···C contacts are also discussed later.

Since the atomic positions of high-resolution protein structures are more reliable than those of low-resolution ones, only the data in the high-resolution class were used for the following statistical analyses. We have also carried out the same analyses using the low-resolution data. However, the results were not largely different from those described below although the directional preferences of the interactions were slightly weaker.

S···O Interaction for an SSC Group. A nonbonded interaction between a cystine sulfur atom (an SSC group) and an oxygen atom is most important among the S···X interactions observed in proteins, not only because the ratio of the S···O contact for an SSC group to the total S···X contacts distinctly increases in the short contact range (Table 1) but also because the frequency of the S···O contact with respect to the number of cystine residues is significant (100/790 in the range of $d \leq 0.0$ Å and 764/790 in the range of $d \leq 0.5$ Å in the high-resolution class). The number of the S···O contacts observed in 604 high-resolution protein structures is graphically shown in Fig. 3a as a function of the relative nonbonded distance (d). When the number was corrected by a factor of $1/r^2$ (which means the normalization by the probable spherical area having a radius of r), a broad peak appeared at around $d = 0.5$ Å (Fig. 3b). This clearly supports the presence of S···O interactions for an SSC group in proteins.

The observed S···O interactions may operate in a wide range of distance according to Fig. 3b: the range may cover at least -0.5 Å $\leq d \leq 0.5$ Å. This is one of the unique features of S···O interactions. Another feature is the types of the O atoms involved in the S···O interactions. Figure 3a shows that main-chain carbonyl O atoms occupied more than half of the cases and also that the ratio is even larger in the short contact range ($d \leq 0.0$ Å). Thus, the S···O interactions for an SSC group contain mostly S···O=C interactions. This type of S···O interactions is analyzed more precisely below. On the other hand, the S···O(water) interactions seem to be rather long-range and may possibly involve direct S···O interactions and OH···S hydrogen bonds. Since the positions and orientations of water molecules are ambiguous in protein structures, no further analysis was performed.

Directional preferences of the observed S···O=C interactions were analyzed by using angles θ_1–θ_6. Figure 4a shows the spatial distribution of the main-chain O atoms relative to the SSC S atom by using access angles θ_1 and θ_2. The scattergram clusters in two areas when the relative nonbonded S···O distance (d) is less than 0.2 Å; one is the area of $65° \leq \theta_1 \leq 95°$ and $150° \leq \theta_2 \leq 180°$, corresponding to the orientation of the backside of the disulfide bond [the direction of antibonding orbital σ^*(SS)], and the other is the area of $145° \leq \theta_1 \leq 175°$ and $75° \leq \theta_2 \leq 105°$, corresponding to the orientation of the backside of the S–C bond [the direction of antibonding orbital σ^*(SC)]. The observed directionality is consistent with the rule for the S···X interactions in small molecules that a nucleophile tends to approach S along the direction of the S–S or S–C bond (Fig. 1). It should be noted that the former cluster is more dense than the latter one. This suggests that linear S–S···O interactions are stronger than linear C–S···O interactions; this tendency is consistent with the normal observation that the S–S bond is weaker and hence more reactive than the S–C bond.

The directionality with respect to the S atom was further analyzed by using θ_3 and θ_4 (Fig. 5a). Two clusters, corresponding to linear S–S···O interactions ($60° \leq \theta_3 \leq 120°$ and $220° \leq \theta_4 \leq 250°$) and C–S···O interactions ($60° \leq \theta_3 \leq 120°$ and $120° \leq \theta_4 \leq 150°$), appeared in the scattergram ($d \leq 0.2$ Å). The areas of the clusters indicated that the spatial distribution of the O atoms largely extends from the SSC plane. In addition, the scattergram showed that the appearance of the clusters is not strongly affected by the distance between the amino acid residues along the sequence [Δ(O,S)], suggesting that the linearity of the S···O interactions may be intrinsic in nature.

On the other hand, directional preferences of S···O=C interactions ($d \leq 0.2$ Å) with respect to the O atom were analyzed by using θ_5 and θ_6 (Fig. 6a). The scattergram shows two clusters. One is the area around $\theta_5 = 90°$ and $\theta_6 = 0°$ (or 360°), and the other is the area around $\theta_5 = 90°$ and $\theta_6 = 180°$, corre-

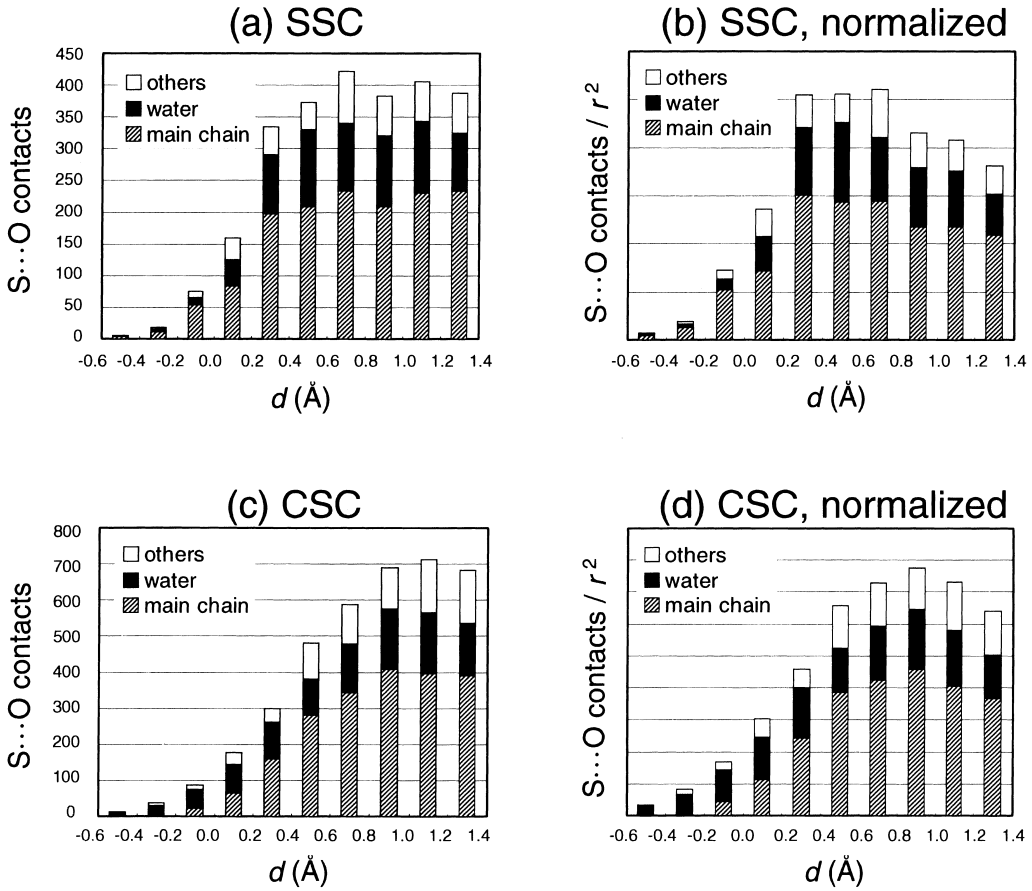

Fig. 3. Histograms of the S···O contacts observed in 604 high-resolution protein structures as a function of relative nonbonded distance d [$\equiv r - vdw(\mathrm{S}) - vdw(\mathrm{O})$]. (a) The number of S···O contacts observed for an SSC group. (b) The number of S···O contacts for an SSC group normalized by a factor of $1/r^2$. (c) The number of S···O contacts observed for a CSC group. (d) The number of S···O contacts for a CSC group normalized by a factor of $1/r^2$.

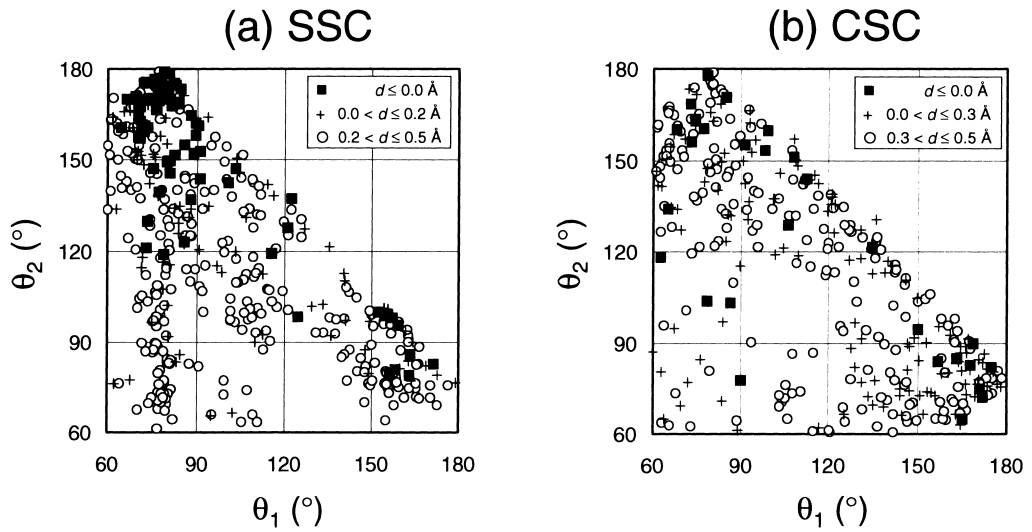

Fig. 4. Spatial distribution of the O atoms relative to the S atom for the S···O=C interactions ($d \leq 0.5$ Å) observed in 604 high-resolution protein structures by using access angles θ_1 and θ_2. The short-range ($d \leq 0.0$ Å), medium-range (0.0 Å $< d \leq$ 0.2 or 0.3 Å), and long-range (0.2 or 0.3 Å $< d \leq 0.5$ Å) interactions are distinguished by symbols ■, +, and ○, respectively. (a) S–S···O=C interactions for an SSC group. (b) C–S···O=C interactions for a CSC group.

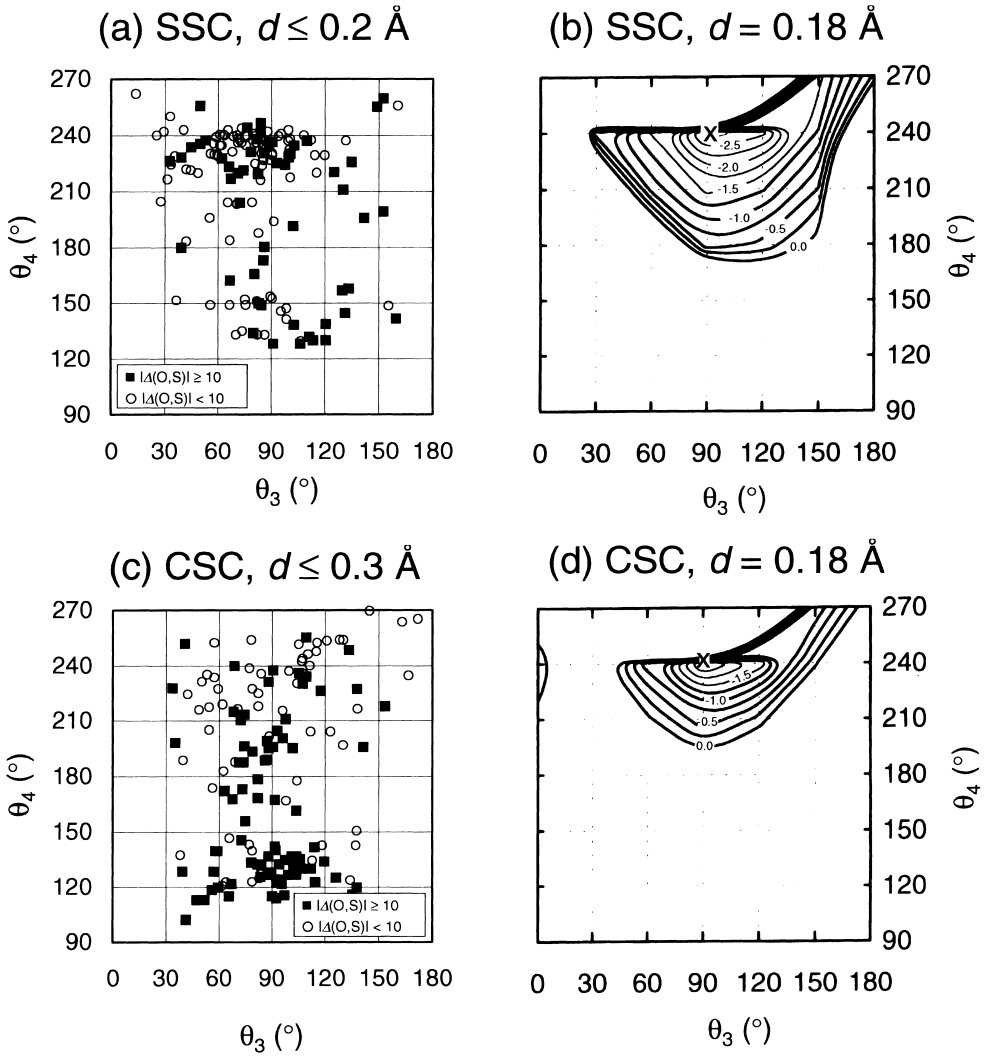

Fig. 5. Directional preferences of S···O=C interactions. (a) Spatial distribution of the O atoms relative to the S atom for the S–S···O=C interactions ($d \leq 0.2$ Å) observed in 604 high-resolution protein structures by using access angles θ_3 and θ_4. Δ(O,S) means the difference of an amino acid residue number between the donating O atom and the accepting S atom. (b) A θ_3–θ_4 potential surface calculated for the CH_3SSCH_3 + $CH_3CONHCH_3$ complex at the MP2/6-31G(d) level under the conditions of $d = 0.18$ Å, $\theta_5 = 90°$, and $\theta_6 = 0°$. The complexation energy is shown in kcal/mol. Symbol x (at $\theta_3 = 90°$, $\theta_4 = 240°$) means the energy minimum. (c) Spatial distribution of the O atoms relative to the S atom for the C–S···O=C interactions ($d \leq 0.3$ Å) observed in 604 high-resolution protein structures by using access angles θ_3 and θ_4. (d) A θ_3–θ_4 potential surface calculated for the CH_3SCH_3 + $CH_3CONHCH_3$ complex at the MP2/6-31G(d) level under the conditions of $d = 0.18$ Å, $\theta_5 = 90°$, and $\theta_6 = 0°$. The complexation energy is shown in kcal/mol. Symbol x (at $\theta_3 = 90°$, $\theta_4 = 240°$) means the energy minimum.

sponding to the orientations just above and below the carbonyl O atom, respectively [the directions of the π(C=O) orbital]. The observed directional preferences were a little surprising because S···O=C interactions in small organic compounds are usually formed in the directions of the lone pairs of the carbonyl O atom,[21–23] i.e., the directions on the carbonyl plane ($\theta_6 \sim 90°$ or 270°).

There are some possible reasons for the observed discrepancy in the directional preferences between the S···O interactions in proteins and those in small molecules. One is the conformational restrictions of protein structures. However, this is not plausible because the clusters are more concentrated in the directions of the π(C=O) orbital when the amino acid residues that form the S···O interaction are separated by more than ten residues along the sequence [$|\Delta(O,S)| \geq 10$] (Fig. 6a). One other possible reason is the existence of NH···O=C hydrogen bonds that might disturb the formation of the S···O interaction in the directions of the lone pairs of the O atom. We chose the S···O interactions, in which the O atom does not have an NH···O hydrogen bond, and investigated the directionality. The result did not show any significant difference from the scattergram shown in Fig. 6a. Hence, the directional preferences of the S···O interactions observed in proteins should not be affected by the existence of NH···O=C hydrogen bonds. One plausible reason for the discrepancy is that the S···O interactions in proteins usually involve an amide O atom, while

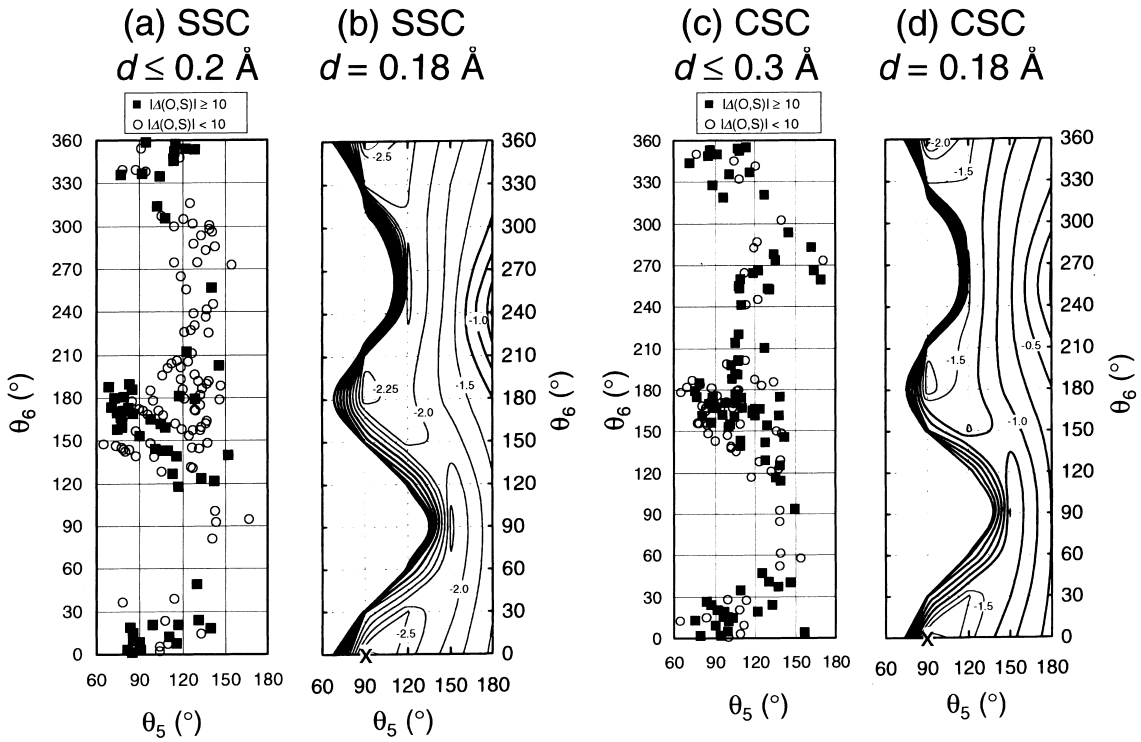

Fig. 6. Directional preferences of S···O=C interactions. (a) Spatial distribution of the S atoms relative to the O atom for the S–S···O=C interactions ($d \leq 0.2$ Å) observed in 604 high-resolution protein structures by using access angles θ_5 and θ_6. Δ(O,S) means the difference of an amino acid residue number between the donating O atom and the accepting S atom. (b) A θ_5–θ_6 potential surface calculated for the CH_3SSCH_3 + $CH_3CONHCH_3$ complex at the MP2/6-31G(d) level under the conditions of $d = 0.18$ Å, $\theta_3 = 90°$, and $\theta_4 = 240°$. The complexation energy is shown in kcal/mol. Symbol x (at $\theta_5 = 90°$, $\theta_6 = 0°$) means the energy minimum. (c) Spatial distribution of the S atoms relative to the O atom for the C–S···O=C interactions ($d \leq 0.3$ Å) observed in 604 high-resolution protein structures by using access angles θ_5 and θ_6. (d) A θ_5–θ_6 potential surface calculated for the CH_3SCH_3 + $CH_3CONHCH_3$ complex at the MP2/6-31G(d) level under the conditions of $d = 0.18$ Å, $\theta_3 = 90°$, and $\theta_4 = 240°$. The complexation energy is shown in kcal/mol. Symbol x (at $\theta_5 = 90°$, $\theta_6 = 0°$) means the energy minimum.

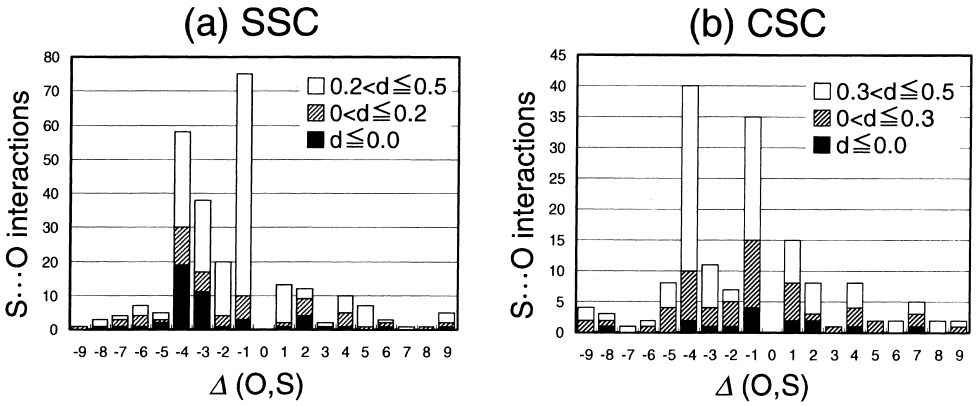

Fig. 7. Histograms of the S···O=C interactions as a function of the difference of an amino acid residue number between the donating O atom and the accepting S atom [Δ(O,S)]. Only the data in the range of $|\Delta(O,S)| < 10$ are shown. (a) S···O interactions for an SSC group. (b) S···O interactions for a CSC group.

those in small molecules involve various types of carbonyl O atoms. Details of the analysis on the observed discrepancy will be disclosed in due course.

Figure 7a shows how the S···O interactions are located along the amino acid sequence in the range of $|\Delta(O,S)| < 10$. When the S···O interaction is short ($d \leq 0.0$ Å), the amino acid residues are mostly separated by four or three residues [Δ(O,S) $= -4$ or -3]. These interactions reside in α- and 3_{10}-helices, respectively, indicating that the S···O interaction stabilizes the secondary structures to some extent. When the S···O interaction is long ($0.2 < d \leq 0.5$ Å), the number of the S···O interactions with Δ(O,S) $= -1$, which correspond to 1,6-S···O inter-

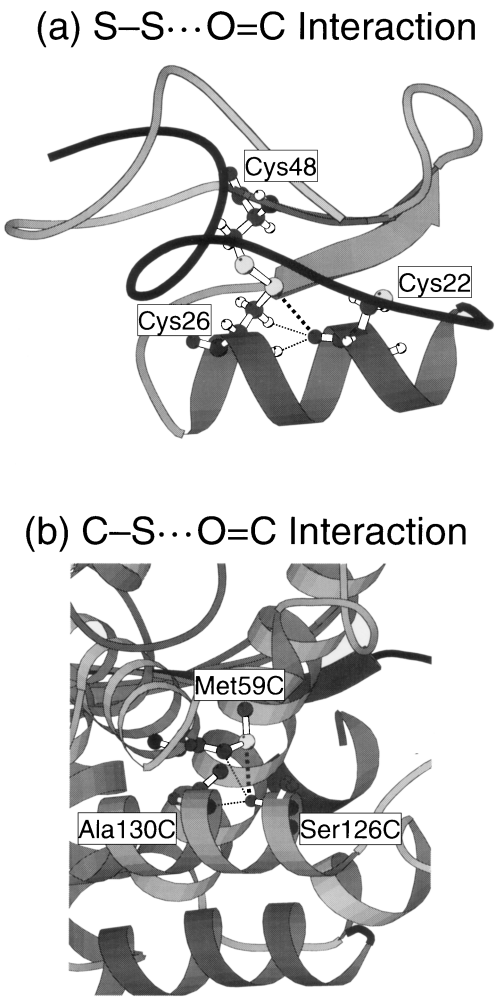

Fig. 8. Typical examples of S···O=C interactions found in protein structures. The figures are generated using MOLSCRIPT.[57] Coexisting NH···O and CH···O hydrogen bonds are also shown with thin broken lines. (a) An S–S···O=C interaction in scorpion protein toxin (1 aho).[50] (b) A C–S···O=C interaction in cryptophyte phycoerythin (1 qgw).[52] See the text for details of the structural parameters.

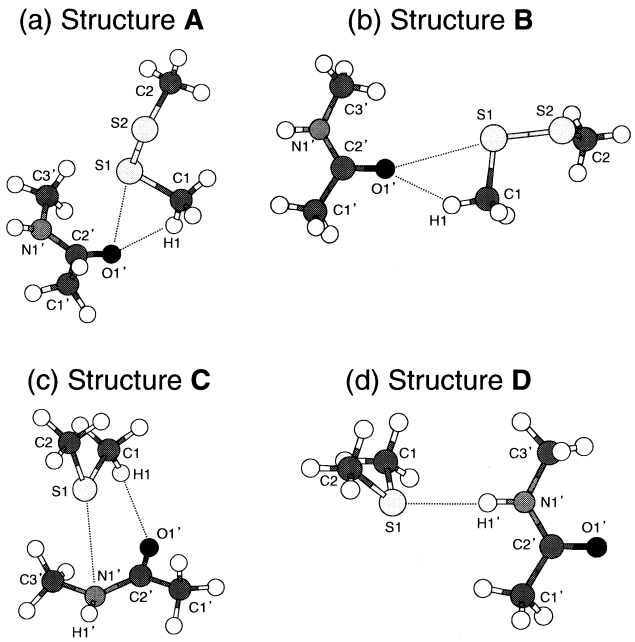

Fig. 9. Optimized structures obtained for model complexes by ab initio calculations at the MP2/6-31G(d) level. Coexisting CH···O hydrogen bonds are also shown with broken lines. (a) A global energy minimum structure for the CH_3SSCH_3 + $CH_3CONHCH_3$ complex. (b) The most stable structure having a coplanar S···O interaction to the carbonyl plane for the CH_3SSCH_3 + $CH_3CONHCH_3$ complex. (c) A global energy minimum structure for the CH_3SCH_3 + $CH_3CONHCH_3$ complex. (d) The most stable structure having an NH···S hydrogen bond for the CH_3SCH_3 + $CH_3CONHCH_3$ complex.

actions separated by five covalent bonds, is significantly enhanced. The 1,6-S···O interactions have structural characters as distorted S–S···O=C interactions due to the conformational restraint: i.e., $\theta_1 \sim 80°$, $\theta_2 = 70° \sim 180°$, $\theta_5 \sim 80°$, and $\theta_6 \sim 140°$.

A typical example of S–S···O=C interactions in α-helices [Δ(O,S) = −4] is shown in Fig. 8a. The protein is scorpion protein toxin (1aho).[50] A short-range S···O interaction (d = −0.23 Å) is formed between the S atom of Cys26 and the O atom of Cys22. The values of θ_1–θ_6 are 81°, 150°, 60°, 228, 134°, and 186°, respectively, indicating linearity of the S–S···O alignment and perpendicularity of the access of the S atom toward the C=O plane. There are two additional interactions around the O atom of Cys22: the NH···O hydrogen bond of the α-helix (H···O = 1.96 Å, ∠N–H···O = 161°) and the weak CH···O hydrogen bond with the H(β) atom of Cys26 (H···O = 2.78 Å, ∠C–H···O = 114°). These interactions would cooperatively stabilize the local protein structure with the S···O interaction. A number of similar coordination structures around the O atom were found in various proteins, not only for the S···O interactions with Δ(O,S) = −4.

In order to elucidate the intrinsic structural preferences as well as the strength of S···O interactions, we performed quantum chemical calculations on the model complex, CH_3SSCH_3 + $CH_3CONHCH_3$, at the MP2/6-31G(d) level of theory. Two stable structures were located (Figs. 9a,b). Structure **A** is the global energy minimum having a vertical S···O interaction to the amide plane (d = 0.02 Å, θ_1 = 71.3°, θ_2 = 169.7°, θ_3 = 98.3°, θ_4 = 237.5°, θ_5 = 81.5°, and θ_6 = 2.3°). The structural features are in good agreement with those observed in proteins. The total stabilization energy due to the complexation with the correction for BSSE[47] was 3.21 kcal/mol including the contribution from a coexisting CH···O hydrogen bond. We have previously reported that the additional coordination of a water molecule to the carbonyl O atom slightly shortens the S···O contact (d = −0.02 Å).[13] Thus, the importance of a coexisting NH···O (or OH···O) hydrogen bond for the S···O interaction (Fig. 8a) was supported theoretically.

On the other hand, Structure **B**, which has a horizontal S···O interaction to the amide plane (d = 0.06 Å, θ_1 = 68.6°, θ_2 = 170.4°, θ_3 = 91.6°, θ_4 = 240.4°, θ_5 = 160.7°, and θ_6 = 66.6°), was 1.69 kcal/mol less stable than Structure **A**. The complexation energy with the correction for BSSE was 2.18 kcal/mol,

which is about 1 kcal/mol smaller than that of Structure **A**. It is important to note that Structure **B** would be more preferred by the CH···O hydrogen bond.[6] Therefore, the S···O interaction should be stronger than the CH···O hydrogen bond because the globally stable structure of the complex (Structure **A**) was not determined by the directional preference of the CH···O hydrogen bond but by that of the S···O interaction.

Within the framework of the present MP2 theory, the difference between MP2 and HF calculation results is assigned to the effects of high-order electrostatic interactions such as a dispersion force.[51] Since the S···O atomic distance became larger for the structure optimized at the HF/6-31G(d) level (d = 0.37 Å) and the stabilization energy reduced to 2.18 kcal/mol, the main stabilization element of the S···O interaction may arise from electron correlations. However, the directional preferences of the S···O interaction observed in proteins as well as in Structure **A** suggested the significant contribution from the charge-transfer interaction between the π(C=O) and σ^*(SS) orbitals. Natural bond orbital analysis[48] suggested that the $\pi \rightarrow \sigma^*$ orbital interaction energy is 0.64 kcal/mol, which corresponds to about 20% of the total S···O interaction energy (3.21 kcal/mol). Other types of intermolecular orbital interactions were found to be negligible except for those due to the coexisting CH···O hydrogen bond. We suggest at present that the S···O interactions in proteins may be primarily stabilized by the dispersion and/or long-range electrostatic forces but that the observed directional propensity is controlled by the HOMO–LUMO-type orbital interaction.

The potential surfaces of the S···O interaction were also calculated. The results are given in Figs. 5b and 6b. It is clearly seen that the distribution patterns of the S···O interactions in proteins almost perfectly follow the potential surfaces of the S···O interaction calculated for the isolated model complex even though the surfaces are rather flat. Thus, it is likely that the S···O interaction is one of the important forces that determine protein structures.

Dependence of the potential surfaces on the relative S···O distance (d) was also analyzed by changing the value of d from -0.22 Å to $+0.58$ Å with an interval of 0.2 Å. The obtained potential surfaces (not shown) revealed that the profiles are only slightly affected by the value of d. The energies of the potential minima on the θ_3 versus θ_4 planes (at $\theta_3 = 90°$ and $\theta_4 = 240°$, see Fig. 5b) were -0.88, -2.34, -2.90, -2.97, and -2.80 kcal/mol at $d = -0.22$, -0.02, 0.18, 0.38, and 0.58 Å, respectively. Similarly, the energies of the potential minima on the θ_5 versus θ_6 planes (at $\theta_5 = 90°$ and $\theta_6 = 0°$, see Fig. 6b) were -2.06, -2.82, -2.97, -2.88, and -2.62 kcal/mol at the corresponding distance. The observations suggested that the S···O interaction functions over a wide range of distance, being consistent with the data shown in Fig. 3b.

S···O Interaction for a CSC Group. Directional preferences of S···O interactions for a CSC group are basically similar to those observed for an SSC group, although the S···O interactions for a CSC group exhibit weaker propensities. This is reasonable in light of the larger bond energy of the S–C bond than that of the S–S bond, as discussed earlier.

Figure 3c shows the number of the S···O contacts observed in 604 high-resolution protein structures as a function of d. When the number was corrected by a factor of $1/r^2$, a weak and broad peak appeared at around d = 0.9 Å (Fig. 3d), suggesting that the S···O interactions for a CSC group must be weaker and more long-range than the S···O interactions for an SSC group. As seen for an SSC group, most S···O interactions for a CSC group contained S···O=C interactions, although the ratio of S···O(water) interactions was large in the short contact range ($d \leq 0.0$ Å).

Directional preferences of the observed S···O=C interactions were analyzed extensively by using angles θ_1–θ_6. Figure 4b shows spatial distribution of the main-chain O atoms relative to the CSC S atom by using access angles θ_1 and θ_2. The scattergram clusters in two areas when the value of d is less than 0.3 Å; one is the area of $65° \leq \theta_1 \leq 95°$ and $150° \leq \theta_2 \leq 180°$, and the other is the area of $150° \leq \theta_1 \leq 180°$ and $65° \leq \theta_2 \leq 95°$, corresponding to the orientations of the backside of the S–C bonds [the directions of antibonding orbital σ^*(SC)]. Figure 5c also shows directionality of the S···O=C interactions by using θ_3 and θ_4 in the range of $d \leq 0.3$ Å. One cluster corresponding to linear C(γ)–S···O interactions ($60° \leq \theta_3 \leq 120°$ and $110° \leq \theta_4 \leq 140°$) appeared clearly, whereas the other cluster corresponding to linear C(ε)–S···O interactions ($60° \leq \theta_3 \leq 120°$ and $220° \leq \theta_4 \leq 250°$) appeared widely spread. It is obvious that the linearity of the S···O interactions for a CSC group is weaker than that for an SSC group. It is also notable that the appearance of the clusters is not strongly affected by the distance between the amino acid residues, forming the S···O interactions, along the sequence [Δ(O,S)].

The observed directional preferences of the S···O=C interactions for a CSC group were in good agreement with the statistical results reported by Pal and Chakrabarti[14] but were quite different from the results reported by Carugo.[38] He analyzed directional preferences of the atomic contacts around the S atoms of methionine residues in proteins and found the slight tendency that the carbonyl O atoms approach the S atom in the direction from θ_3 = 25–30° (the directions of the lone pairs of the S atom). The discrepancy may arise from the smaller number of protein structures employed for the database analysis.

On the other hand, Fig. 6c shows the directional preference of S···O=C interactions ($d \leq 0.3$ Å) with respect to the O atom by using θ_5 and θ_6. The scattergram clusters in two areas, corresponding to the orientations just above ($\theta_5 \sim 90°$ and $\theta_6 \sim 0°$) and below ($\theta_5 \sim 90°$ and $\theta_6 \sim 180°$) the carbonyl O atom [the directions of the π(C=O) orbital]. The propensity is similar to the case of the S···O interactions for an SSC group (Fig. 6a). However, their appearance did not depend on the value of Δ(O,S) so clearly as that observed for an SSC group.

Figure 7b shows how the S···O interactions are located along the amino acid sequence in the range of $|\Delta(O,S)| < 10$. When the S···O interaction is short-range ($d \leq 0.0$ Å), the number of the S···O contacts does not show any dependence on the value of Δ(O,S). However, when the interaction is long-range ($0.3 < d \leq 0.5$ Å), the numbers of S···O interactions with $\Delta(O,S) = -4$ and -1 significantly increase. These S···O interactions correspond to those in α-helices and 1,7-S···O interactions (separated by six covalent bonds), respectively.

A typical example of C–S···O=C interactions is shown in Fig. 8b. The protein is cryptophyte phycoerythrin (1qgw).[52] An S···O interaction ($d = -0.00$ Å) is formed between the S atom of Met59 and the O atom of Ser126 [$\Delta(O,S) = 67$]. The

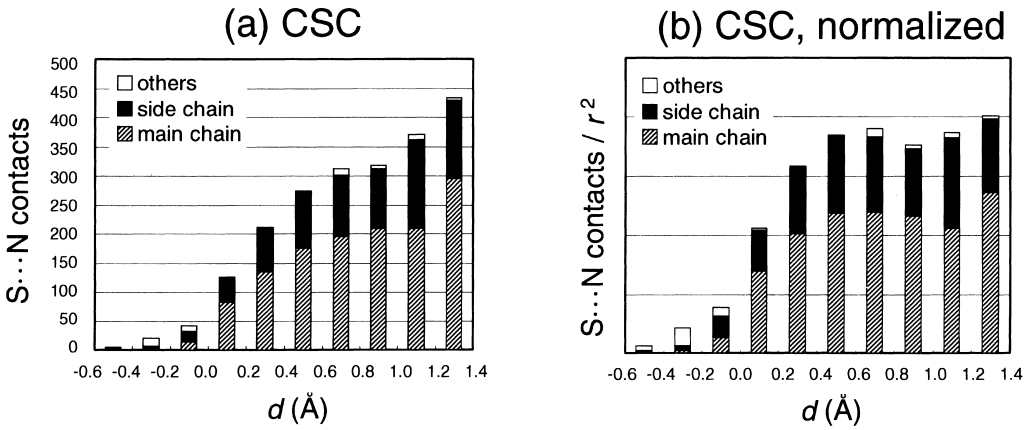

Fig. 10. Histograms of the S···N contacts observed for a CSC group in 604 high-resolution protein structures as a function of relative nonbonded distance d [$\equiv r - vdw(S) - vdw(N)$]. (a) The number of S···N contacts. (b) The number of S···N contacts normalized by a factor of $1/r^2$.

values of θ_1–θ_6 are 78°, 178°, 88°, 231°, 103°, and 188°, respectively, indicating a linear C–S···O alignment and a vertical access of the S atom toward the C=O plane. There is an α-helical NH···O hydrogen bond between the O atom of Ser126 and the N atom of Ala130 (N···O = 3.00 Å). In addition, a weak CH···O hydrogen bond may exist between the O atom of Ser126 and the H(γ) atom of Met59 since the C(γ) atom is present close to the O atom [r(C···O) = 3.45 Å]. These interactions would cooperatively stabilize the local protein structure with the S···O interaction, as in the case of the S–S···O=C interaction shown in Fig. 8a.

Ab initio calculations were carried out on the model complex, CH_3SCH_3 + $CH_3CONHCH_3$, at the MP2/6-31G(d) level of theory. We found that the potential surfaces are extremely flat relative to the S···O distance (r) and also that the global energy minimum is the structure with a close S···N contact (Structure **C** in Fig. 9c), not the one with a close S···O contact. However, the optimized structure with a fixed close S···O distance (r = 3.30 Å, i.e., d = −0.02 Å), which is not a local energy minimum structure, was only 0.32 kcal/mol less stable than the global minimum structure (Structure **C**). The complexation energy (2.47 kcal/mol) was still large. Therefore, the C–S···O=C interaction in proteins may be attractive.

Potential surfaces calculated for the CH_3SCH_3 + $CH_3CONHCH_3$ complex are shown in Figs. 5d and 6d. Since we applied the same orientation for CH_3SCH_3 as that of CH_3SSCH_3 in Structure **A** (Fig. 9a), only one stable area could be located on the θ_3 versus θ_4 plane (Fig. 5d). This is because the fixed θ_5 and θ_6 angles during the calculations of the potential surfaces caused unavoidable steric repulsion when the value of θ_4 was small. The potential surfaces should be intrinsically symmetric with respect to the axes of θ_3 = 90° and θ_4 = 180° if θ_5 and θ_6 were flexible. Therefore, the obtained potential surfaces (Figs. 5d and 6d) are consonant with the directional preferences of the C–S···O=C interactions observed in proteins. The S···O interaction for a CSC group would also be an important factor for determining protein structures.

The mechanism of the S···O interaction for a CSC group would be similar to that for an SSC group. Dispersion and/or long-range electrostatic forces should be the main stabilization elements, but the contribution from the π(C=O)→σ^*(SC) orbital interaction would also be involved according to the observed directional preferences. The S···O interaction may operate in a wide range of distance since the potential surfaces do not change significantly independent of the relative S···O distance (d): The energies of the potential minima on the θ_3 versus θ_4 planes (at θ_3 = 90° and θ_4 = 240°, see Fig. 5d) were −0.36, −1.75, −2.30, −2.40, and −2.28 kcal/mol at d = −0.22, −0.02, 0.18, 0.38, and 0.58 Å, respectively. Similarly, the energies of the potential minima on the θ_5 versus θ_6 planes (at θ_5 = 90° and θ_6 = 0°, see Fig. 6d) were −1.20, −2.10, −2.37, −2.33, and −2.15 kcal/mol at the corresponding distance of d.

S···N Interaction for a CSC Group. The number of the S···N contacts for a CSC group is shown in Fig. 10a as a function of the relative nonbonded distance (d). When the number was corrected by a factor of $1/r^2$, a weak and broad peak appeared at around d = 0.5 ∼ 0.7 Å (Fig. 10b), suggesting the existence of the long-range S···N interactions in proteins. Since more than a half of the S···N interactions were S···N(main chain) interactions, structural features of these interactions ($d \leq$ 0.5 Å) were analyzed by using θ_1, θ_2, θ_5, and θ_6. The results are shown in Fig. 11.

Figure 11a shows spatial distribution of the main-chain N atoms with respect to the CSC S atom by using access angles θ_1 and θ_2. Figure 11b displays directional preferences of the S···N(main chain) interactions with respect to the N atom by using angles θ_5 and θ_6. The scattergrams indicate that the S···N interactions involve two types of interactions; one is the linear C–S···N interaction perpendicular to the amide plane, corresponding to the clusters around θ_1/θ_2 ∼ 90°, θ_2/θ_1 ∼ 150°, θ_5 ∼ 75°, and θ_6 ∼ 0° or 180° (shown as open circles), and the other is the linear N–H···S hydrogen bond, corresponding to the clusters around θ_1 ∼ 100°, θ_2 ∼ 100°, θ_5 ∼ 180° (shown as filled squares).

The former S···N interaction shows weak linearity of the C–S···N atomic alignment and perpendicularity to the amide plane. The observed directional preferences were similar to those of the linear C–S···O interaction, suggesting a contribution from the π(N)→σ^*(SC) orbital interaction to the stability.

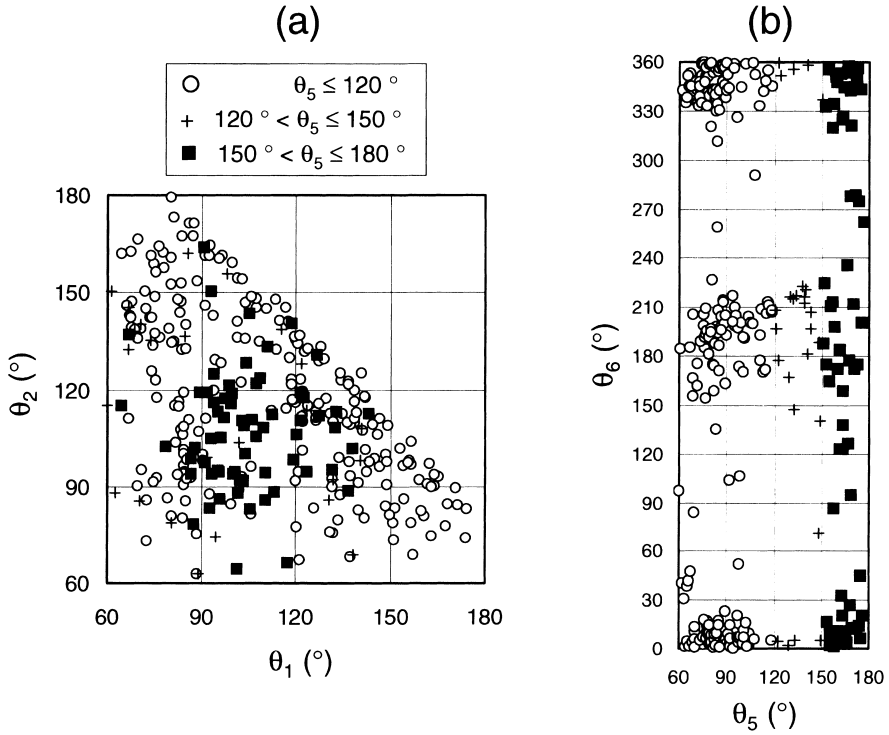

Fig. 11. Directional preferences of S···N(main chain) interactions for a CSC group. Symbols ○, +, and ■ are utilized for the S···N interactions in the ranges of $\theta_5 \leq 120°$, $120° < \theta_5 \leq 150°$, and $150° < \theta_5 \leq 180°$, respectively (a) Spatial distribution of the N atoms relative to the S atom ($d \leq 0.3$ Å) observed in 604 high-resolution protein structures by using access angles θ_1 and θ_2. (b) Spatial distribution of the S atoms relative to the N atom by using access angles θ_5 and θ_6.

The propensities are consistent with the S···N interactions characterized in small molecules.[53] Our statistical data indicate that a similar S···N interaction may also exist in protein structures. Similar directional preferences were previously reported for the specific interaction between an SH group and a main-chain or side-chain amide group.[54] The S(H)···N interactions may involve direct S···N interactions as well.

On the other hand, the structural preferences of the latter interaction (an NH···S hydrogen bond) suggested a contribution from the n(S)→σ*(NH) orbital interaction to the stability because the N atom tends to approach the S atom in the directions of the lone pairs. The directional preferences are consistent with those found in the previous statistical analysis of hydrogen bond stereochemistry in protein structures.[35]

A typical example of the C–S···N(main chain) interactions is shown in Fig. 12a. The protein is *Escherichia coli* glyoxalase I (1f9z).[55] An S···N interaction ($d = -0.20$ Å) is formed between the S atom of Met7A and the N atom of Leu8A [Δ(N,S) = 1]. The values of θ_1–θ_6 are 80°, 160°, 71°, 228°, 90°, and 193°, respectively, indicating the linear C–S···N alignment and the vertical access of the S atom toward the amide plane. Similar 1,6-S···N interactions [Δ(N,S) = 1] were most frequently observed in protein structures: the number of observations was 66 in the range of $d \leq 0.5$ Å for 2124 methionine residues.

Figure 12b shows an example of the NH···S hydrogen bond. The protein is type Ⅲ antifreeze protein (1msi).[56] An almost linear NH···S hydrogen bond is formed between the S atom of Met59 and the main chain NH group of Met56 [Δ(N,S) = −3]. The values of d and θ_1–θ_6 are 0.15 Å, 96°, 105°, 17°, 200°, 172°, and 354°, respectively, indicating that the hydrogen bond is formed in the direction of the n(S) and σ*(NH) orbitals.

Ab initio calculations reasonably reproduced the observed structural features of the S···N interactions as shown in Figs. 9c,d. Structure **C** is the global energy minimum having a vertical S···N interaction to the amide plane ($d = 0.22$ Å, $\theta_1 = 166.8°$, $\theta_2 = 93.1°$, $\theta_3 = 95.7°$, $\theta_4 = 142.5°$, $\theta_5 = 79.2°$, and $\theta_6 = 186.1°$). It should be noted that θ_5 is slightly sharper than the right angle because the S atom comes over the amide N–C bond. This is consistent with the appearance of the corresponding clusters in Fig. 11b: the centers of clusters are located at $\theta_5 \sim 75°$ and $\theta_6 \sim 0°$ and 180°. The total stabilization energy due to the complexation with the correction for BSSE was 2.88 kcal/mol, including the contribution from a coexisting weak CH···O hydrogen bond. On the other hand, Structure **D**, which has a linear NH···S hydrogen bond on the amide plane ($d = 0.24$ Å, $\theta_1 = 110.9°$, $\theta_2 = 111.0°$, $\theta_3 = 146.6°$, $\theta_4 = 180.1°$, $\theta_5 = 179.0°$, and $\theta_6 = 85.6°$), was 0.78 kcal/mol less stable than Structure **C**. It is important to note that the structural features of both Structures **C** and **D** are consonant with those observed in protein structures, suggesting that the directionality of the S···N interactions is due to the intrinsic nature.

S···S Interaction for an SSC Group. Although only a small number of S···S contacts were observed for an SSC group, the S···S contacts showed a unique character as an n(S)→σ*(S) orbital interaction. Having analyzed the observed 83 S···S contacts in the range of $d \leq 0.5$ Å (Table 1),

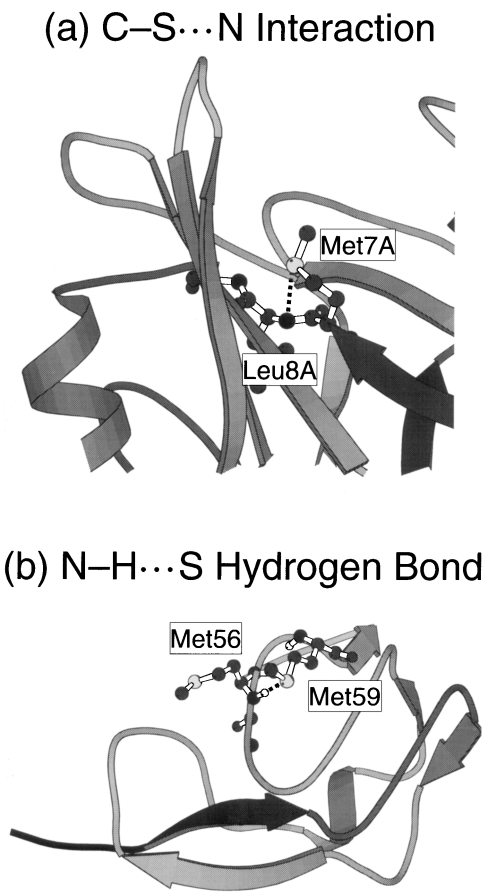

Fig. 12. Typical examples of S···N(main chain) interactions found in protein structures. The figures are generated using MOLSCRIPT.[57] (a) A C–S···N interaction in *Escherichia coli* glyoxalase I (1 f9z).[55] (b) An N–H···S hydrogen bond in type III antifreeze protein (1 msi).[56] See the text for details of the structural parameters.

we found that most of the contacts (72 out of 83) are formed between the S atoms pertaining to an SSC group.

Figure 13 shows the scattergram of the S(SSC)···S(SSC) contacts observed in protein structures by using access angles θ_3 and θ_4. The definitions of the angles are the same as the ones previously utilized to analyze the directionality of S···S contacts in organic crystals.[17] According to the literature method, the S atoms were categorized to two classes: type I is the S atom having a value of θ_3 closer to 0° or 180° (shown as open circles) and type II is the S atom having a value of θ_3 closer to 90° (shown as filled squares) in each pair of the S···S contact. It is seen that the S atoms of type I are mostly present in the regions with $\theta_3 < 60°$ and $\theta_3 > 120°$ while the S atoms of type II make clusters at $\theta_3 \sim 90°$ and $\theta_4 \sim 135°$ or 225°. Thus, the type I S atom would donate the lone pair (n) to the type II S atom, which would accept the electrons in the antibonding orbital of the S–S or S–C bond [σ^*(S)]. The directionality is the same as that observed in organic crystals,[17] suggesting the presence of S···S interactions for an SSC group in protein structures. Although the structural features of the S···S interactions are in accord with stereochemical preferences of the n(S)→σ^*(S) orbital interaction, it should be stated that the

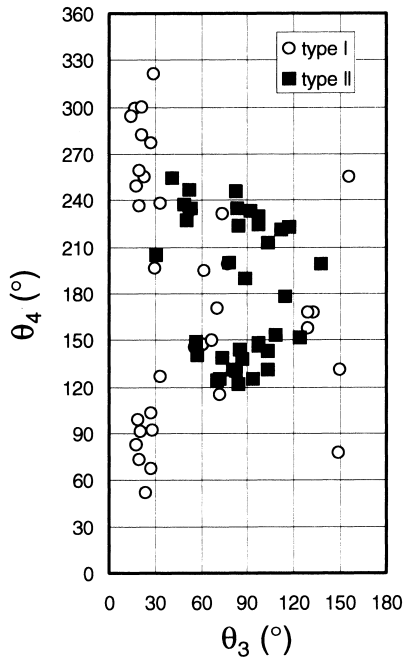

Fig. 13. A scattergram of the S(SSC)···S(SSC) contacts ($d \leq$ 0.5 Å) observed in 604 high-resolution protein structures by using access angles θ_3 and θ_4. The S atoms are categorized to two classes according to the literature method:[17] Type I is the S atom having a value of θ_3 closer to 0° or 180° (shown as open circles) and type II is the S atom having a value of θ_3 closer to 90° (shown as filled squares) in each pair of the S···S contact.

stabilization mechanism of the S···S interactions may also include other factors, such as a van der Waals force, as suggested by Dahaoui et al.[19]

S···C(π) Interaction for SSC and CSC Groups. Tables 1 and 2 suggested that the S···C interactions are weaker or more long-range than other S···X (X = O, N, and S) interactions. However, the number of S···C contacts was largest. This is not only because the C atoms are most abundant in proteins but also because the specific interaction would exist between the S atom and the π systems. The S···C(π) interactions in proteins have been frequently discussed in the literature, as mentioned in the introductory remarks. Therefore, we analyzed the observed S···C contacts in view of the S···C(π) interaction.

Figure 14 shows the number of the S atoms that are present near the phenylalanine (Phe) and tyrosine (Tyr) residues in protein structures as a function of the distant (r) between the S atom and the centroid of the aromatic ring. When the number was corrected by a factor of $1/r^2$, sharp peaks appeared at $r \sim$ 4.5 Å for an SSC group (Fig. 14b) and at $r \sim$ 5.5 Å for a CSC group (Fig. 14d). The results are consistent with the previous studies,[31,32] supporting the presence of specific S···C(π) interactions. The results also indicate that the S···C(π) interactions for an SSC group are more short-range than those for a CSC group.

Figure 15 shows spatial distribution of the S atoms relative to the aromatic plane as a side view. The origin is set to the centroid of the aromatic ring. The standard positions and van

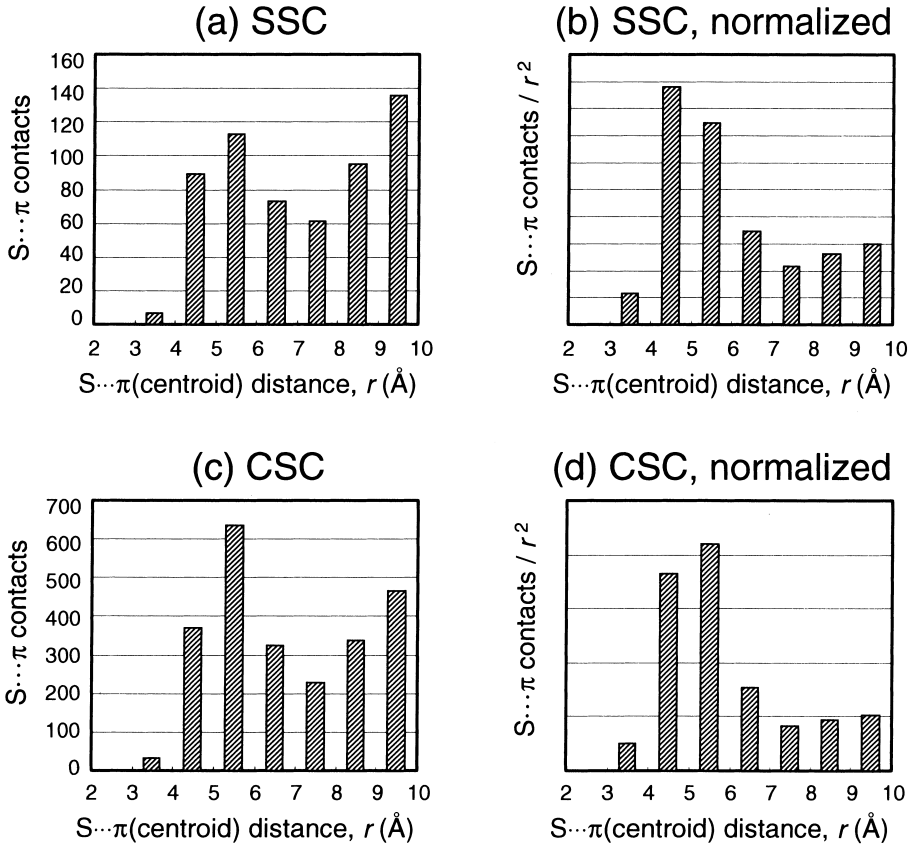

Fig. 14. Histograms of the S···C(π) contacts observed in 604 high-resolution protein structures as a function of the S···π(centroid) distance (r). Only side chains of phenylalanine and tyrosine residues are counted for the π systems. (a) The number of S···C(π) contacts observed for an SSC group. (b) The number of S···C(π) contacts for an SSC group normalized by a factor of $1/r^2$. (c) The number of S···C(π) contacts observed for a CSC group. (d) The number of S···C(π) contacts for a CSC group normalized by a factor of $1/r^2$.

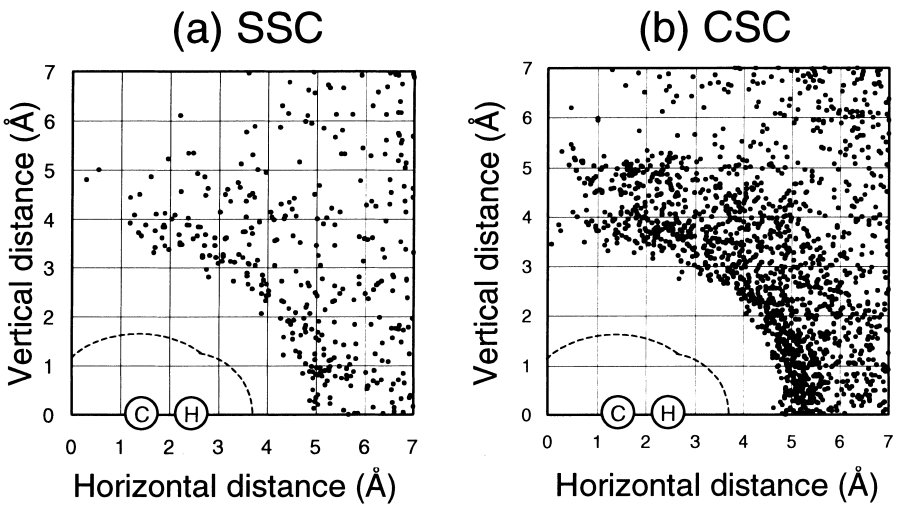

Fig. 15. Spatial distribution of the S atoms relative to the aromatic plane (Phe or Tyr) as a side view. The origin is set to the centroid of the aromatic ring. The standard positions and van der Waals surfaces of the aromatic C and H atoms are depicted for convenience.

der Waals surfaces of the aromatic C and H atoms are depicted for convenience. It is seen that the S atoms tend to approach the aromatic ring not from the top but from the edge. Reid et al.[32] previously pointed out this feature of the S···C(π) interactions using 36 non-redundant protein structures. Figure 15 supports their consideration that most of S···C(π) contacts in

protein structures can be explained by the CH···S interactions. It should be noted, however, that there are still a number of S···C(π) contacts in the range of the horizontal distance ≤ 3.0 Å, which might be explained by the direct nonbonded S···C(π) interactions.

Conclusion

Close S···X (X = O, N, S, C, etc.) atomic contacts were thoroughly surveyed in 604 high-resolution (≤ 2.0 Å) heterogeneous X-ray structures selected from a protein databank (PDB_SELECT) and were statistically analyzed in terms of the nonbonded S···X distance, the directionality relative to the S and X atoms, and the location along the amino acid sequence. Ab initio calculations were also carried out on the model complexes (CH_3SSCH_3 + $CH_3CONHCH_3$ and CH_3SCH_3 + $CH_3CONHCH_3$) to investigate the nature of the observed S···X interactions. Several types of nonbonded interactions in proteins have been characterized.

(1) S–S···O=C interactions: A majority of close S···O contacts for an SSC group in proteins can be assigned to this type of interaction, which is characterized by the linear S–S···O atomic alignment and the vertical access of the S atom to the carbonyl plane. The S···O interactions would be as strong as 3.2 kcal/mol, based on the ab initio calculations, and should be predominantly stabilized by dispersion and/or long-range electrostatic forces with a significant contribution from a $\pi(C{=}O)\rightarrow\sigma^*(S)$ orbital interaction. The interactions most frequently reside in helices, suggesting that they support the stability of the secondary structures. The presence of a CH···O hydrogen bond as well as that of a conventional hydrogen bond (such as NH···O or OH···O) at the main-chain carbonyl O atom would cooperatively stabilize the protein structures with the S···O interaction.

(2) C–S···O=C interactions: Such interactions have similar characters to the S–S···O=C interactions, although the interaction energy is smaller than that of the S–S···O=C interactions. The C–S···O=C interactions would be as strong as 2.5 kcal/mol.

(3) C–S···N interactions: Most close S···N contacts for a CSC group in proteins can be assigned to the C–S···N(main-chain) interactions or C–S···H–N(main-chain) hydrogen bonds. The C–S···N interactions are characterized by the linear C–S···N atomic alignment and the vertical access of the S atom to the amide plane, suggesting a contribution from a $\pi(N)\rightarrow\sigma^*(S)$ orbital interaction. The interactions would be as strong as 2.9 kcal/mol based on the calculations.

(4) S–S···S–S interactions: A majority of close S···S contacts in proteins can be assigned to this type of interaction. Directional preferences of the interactions suggested a contribution from an $n(S)\rightarrow\sigma^*(S)$ orbital interaction to the stability.

Structural features of known C–S···H–N hydrogen bonds and S···C(π) interactions have also been characterized. According to these observations, it can be concluded that sulfur-containing functional groups of cystine (an SSC group) and methionine (a CSC group) are not just hydrophobic moieties but are able to form specific nonbonded interactions with nearby polar non-hydrogen atoms (X) in folded protein structures. Remarkable agreement between the directional preferences of the S···O interactions observed in protein structures and the profiles of the potential surfaces calculated for the isolated model complexes strongly suggested that these nonbonded interactions are important for determining protein structures.

We thank the Computer Center, Institute for Molecular Science, Okazaki National Research Institutes for the use of IBM SP2 (Project code, ep5). This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (C) "Genomic Information Science" from the Ministry of Education, Culture, Sports, Science and Technology.

References

1 W. F. Van Gunsteren, H. J. C. Berendsen, J. Hermans, W. G. J. Hol, and J. P. M. Postma, *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 4315 (1983).

2 B. M. Pettitt and M. Karplus, *Chem. Phys. Lett.*, **121**, 194 (1985).

3 S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.*, **7**, 230 (1986).

4 P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler, *Proteins*, **4**, 31 (1988).

5 I. K. Roterman, M. H. Lambert, K. D. Gibson, and H. A. Scheraga, *J. Biomol. Struct. Dynam.*, **7**, 421 (1989).

6 Z. S. Derewenda, L. Lee, and U. Derewenda, *J. Mol. Biol.*, **252**, 248 (1995).

7 J. Bella and H. M. Berman, *J. Mol. Biol.*, **264**, 734 (1996).

8 R. Vargas, J. Garza, D. A. Dixon, and B. P. Hay, *J. Am. Chem. Soc.*, **122**, 4750 (2000).

9 M. M. Flocco and S. L. Mowbray, *J. Mol. Biol.*, **35**, 709 (1994).

10 J. P. Gallivan and D. A. Dougherty, *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 9459 (1999).

11 J. P. Gallivan and D. A. Dougherty, *J. Am. Chem. Soc.*, **122**, 870 (2000).

12 Y. Umezawa and M Nishio, *Bioorg. Med. Chem.*, **6**, 493 (1998); M. Brandl, M. S. Weiss, A. Jabs, J. Sühnel, and R. Hilgenfeld, *J. Mol. Biol.*, **307**, 357 (2001).

13 M. Iwaoka, S. Takemoto, M. Okada, and S. Tomoda, *Chem. Lett.*, **2001**, 132.

14 D. Pal and P. Chakrabarti, *J. Biomol. Struct. Dynam.*, **19**, 115 (2001).

15 E. Ciuffarin and G. Guaraldi, *J. Org. Chem.*, **35**, 2006 (1970).

16 R. E. Rosenfield, Jr. and R. Parthasarathy, *J. Am. Chem. Soc.*, **99**, 4860 (1977).

17 T. N. G. Row and R. Parthasarathy, *J. Am. Chem. Soc.*, **103**, 477 (1981).

18 G. R. Desiraju and V. Nalini, *J. Mater. Chem.*, **1**, 201 (1991).

19 S. Dahaoui, V. Pichon-Pesme, J. A. K. Howard, and C. Lecomte, *J. Phys. Chem. A.*, **103**, 6240 (1999).

20 R. J. Zauhar, C. L. Colbert, R. S. Morgan, and W. J. Welsh, *Biopolymers*, **53**, 233 (2000).

21 F. T. Burling and B. M. Goldstein, *J. Am. Chem. Soc.*, **114**, 2313 (1992).

22 Y. Nagao, T. Hirata, S. Goto, S. Sano, A. Kakehi, K. Iizuka, and M. Shiro, *J. Am. Chem. Soc.*, **120**, 3104 (1998).

23 S. Wu and A. Greer, *J. Org. Chem.*, **65**, 4883 (2000).

24 M. Iwaoka and S. Tomoda, *J. Am. Chem. Soc.*, **118**, 8077 (1996).

25 M. Iwaoka, H. Komatsu, and S. Tomoda, *Chem. Lett.*, **1998**, 969; M. Iwaoka, H. Komatsu, T. Katsuda, and S. Tomoda, *J. Am. Chem. Soc.*, **124**, 1902 (2002).
26 H. Komatsu, M. Iwaoka, and S. Tomoda, *Chem. Comm.*, **1999**, 205.
27 B. M. Goldstein, S. D. Kennedy, and W. J. Hennen, *J. Am. Chem. Soc.*, **112**, 8265 (1990).
28 D. H. R. Barton, M. B. Hall, Z. Lin, S. I. Parekh, and J. Reibenspies, *J. Am. Chem. Soc.*, **115**, 5056 (1993).
29 M. Iwaoka and S. Tomoda, *J. Am. Chem. Soc.*, **116**, 2557 (1994).
30 G. Mugesh, A. Panda, B. B. Singh, N. S. Punekar, and R. J. Butcher, *J. Am. Chem. Soc.*, **123**, 839 (2001).
31 R. S. Morgan, C. E. Tatsch, R. H. Gushard, J. M. McAdon, and P. K. Warme, *Int. J. Peptide Protein Res.*, **11**, 209 (1978).
32 K. S. C. Reid, P. F. Lindley, and J. M. Thornton, *FEBS Lett.*, **190**, 209 (1985).
33 A. R. Viguera and L. Serrano, *Biochemistry*, **34**, 8771 (1995).
34 B. V. Cheney, M. W. Schulz, and J. Cheney, *Biochim. Biophys. Acta*, **996**, 116 (1989).
35 J. A. Ippolito, R. S. Alexander, and D. W. Christianson, *J. Mol. Biol.*, **215**, 457 (1990).
36 J. C. Taylor and G. D. Markham, *J. Biol. Chem.*, **274**, 32909 (1999).
37 W. Brandt, A. Golbraikh, M. Täger, and U. Lendeckel, *Eur. J. Biochem.*, **261**, 89 (1999).
38 O. Carugo, *Biol. Chem.*, **380**, 495 (1999).
39 U. Hobohm, M. Scharf, R. Schneider, and C. Sander, *Protein Sci.*, **1**, 409 (1992).
40 U. Hobohm and C. Sander, *Protein Sci.*, **3**, 522 (1994).
41 J. L. Sussman, D. W. Lin, J. S. Jiang, N. O. Manning, J. Prilusky, O. Ritter, and E. E. Abola, *Acta Crystallogr. Sect. D*, **54**, 1078 (1998).
42 A. Bondi, *J. Phys. Chem.*, **68**, 441 (1964).
43 N. L. Allinger, X. Zhou, and J. Bergsma, *J. Mol. Struct.*, **312**, 69 (1994).
44 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, "Gaussian 98, Revision A.4.," Gaussian Inc., Pittsburgh PA, 1998.
45 C. Møller and M. S. Plesset, *Phys. Rev.*, **46**, 618 (1934).
46 M. Head-Gordon, J. A. Pople, and M. J. Frisch, *Chem. Phys. Lett.*, **153**, 503 (1988).
47 D. W. Schwenke and G. Truhlar, *J. Chem. Phys.*, **82**, 2418 (1985).
48 A. E. Reed, L. A. Curtiss, and F. Weinhold, *Chem. Rev.*, **88**, 899 (1988).
49 A. T. Brünger and M. Nilges, *Rev. Biophys.*, **26**, 49 (1993).
50 G. D. Smith, R. H. Blessing, S. E. Ealick, J. C. FontecillaCamps, H. A. Hauptman, D. Housset, D. A. Langs, and R. Miller, *Acta Crystallogr. Sect. D*, **53**, 551 (1997).
51 B. Jeziorski, R. Moszynski, and K. Szalewicz, *Chem. Rev.*, **94**, 1887 (1994); S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, and K. Tanabe, *J. Phys. Chem. A*, **103**, 8265 (1999).
52 K. E. Wilk, S. J. Harrop, L. Jankova, D. Edler, G. Keenan, F. Sharples, R. G. Hiller, and P. M. G. Curmi, *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 8901 (1999).
53 K. I. Varughese, A. C. Storer, and P. R. Carey, *J. Am. Chem. Soc.*, **106**, 8252 (1984).
54 D. Pal and P. Charkrabarti, *J. Biomol. Struct. Dynam.*, **15**, 1059 (1998).
55 M. M. He, S. L. Clugston, J. F. Honek, and B. W. Matthews, *Biochemistry*, **39**, 8719 (2000).
56 Z. C. Jia, C. I. DeLuca, H. M. Chao, and P. L. Davies, *Nature*, **384**, 285 (1996).
57 P. J. Kraulis, *J. Appli. Crystallogr.*, **24**, 946 (1991).