Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 19, Issue Number 1, (2001)

Non-hydrogen Bond Interactions Involving the Methionine Sulfur Atom†

www.adeninepress.com

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

Abstract

Of all the nonbonded interactions, hydrogen bond, because of its geometry involving polar atoms, is the most easily recognizable. Here we characterize two interactions involving the divalent sulfur of methionine (Met) residues that do not need any participation of proton. In one an oxygen atom of the main-chain carbonyl group or a carboxylate side chain is used. In another an aromatic atom interacting along the face of the ring is utilized. In these, the divalent sulfur behaves as an electrophile and the other electron-rich atom, a nucleophile. The stereochemistry of the interaction is such that the nucleophile tends to approach approximately along the extension of one of the covalent bonds to S. The nitrogen atom of histidine side chain is extensively used in these nonbonded contacts. There is no particular geometric pattern in the interaction of S with the edge of an aromatic ring, except when an N-H group in involved, which is found within 40º from the perpendicular to the sulfide plane, thus defining the geometry of hydrogen bond interaction involving the sulfur atom. As most of the Met residues which partake in such stereospecific interactions are buried, these would be important for the stability of the protein core, and their incorporation in the binding site would be useful for molecular recognition and optimization of the site's affinity for partners (especially containing aromatic and heteroaromatic groups). Mutational studies aimed at replacing Met by other residues would benefit from the delineation of these interactions.

Keywords: methionine; hydrogen bond; S···O interaction; S···aromatic interaction; protein core; molecular recognition

Introduction

There is a dichotomy in the behaviour of the sulfur-containing amino acid residue, methionine (Met). On one hand, the plot of water accessible surface area of different residues against hydrophobicity (as measured by the free energy of transfer of amino acids from water to organic solvents) puts Met not with completely nonpolar side chains (Ala, Val, Leu and Phe), but with residues containing one dipole, such as Ser, Thr, Tyr and His (1). Indeed, the dipole of Met seems to be rather strong, which could make it an ideal hydrogen bond acceptor (2). On the other hand, Met residues are not found to be adept in forming hydrogen bonds (3). The very few bonds that are observed do not indicate any particular stereochemical preference which is the hallmark of all oxygen-containing residues (4).

Unlike the protein structures one can locate the protons in the X-ray crystallographic investigations of small organic molecules, which thus offer a more reliable resource for the analysis of hydrogen bonds. Using the structural database of small molecules, Allen *et al*. (5) found that of 1811 Y-S-Z substructures (Y, Z = C, N, O

Fax: (91) 33-334-3886
Email: pinak@boseinst.ernet.in

or S) that co-occur with N-H or O-H donors, only 4.75% form S···H-N,O bonds, thus indicating that divalent sulfur is a poor hydrogen-bond acceptor. Although averse to form hydrogen bonds, there are many convincing examples to show that the sulfur atom can be in a direct contact with oxygen atom (6-9). In a landmark paper, Rosenfield *et al*. (10) found that there are two types of noncovalent interactions with their distinct geometry at the divalent sulfur atom. The electrophiles tend to approach S at angles < 40º from the perpendicular to the plane through atoms Y-S-Z (Y, Z = any atom, except H), whereas nucleophiles approach approximately along the posterior extensions of the Y→ S or Z → S bonds. The interaction of metal ions (electrophiles) with Met residues in protein structures exhibit the above stereochemistry (11), and it would be of interest to see if oxygen atoms (with lone pair of electrons) can show the geometry expected of a nucleophile in their interaction with methionine sulfur. We also analyze if the electron-rich face of aromatic rings in protein structures can act as nucleophiles when aromatic residues interact with Met.

Valuable structural information on the binding characteristics of chemical groups can be obtained from the study of their environment in the solid state, which is then used in the structure-based drug design (12-13). Consequently, our analysis will not only address the question if the environment around divalent sulfur as observed in small molecules is maintained locally around the Met side-chain during the folding process, which is driven mainly by the global requirement to bury apolar side chains, but additionally any specific geometry observed can then be used to design a more precise binding geometry of an inhibitor if the active site contains a Met residue.

Methods

The Protein Data Bank (PDB) (14) files (as of January 1999) used for the analysis were selected using PDB_SELECT (15) to collect 393 PDB chains with an R-factor below 20%, a resolution better than 2.0 Å and sequence identity less than 25%. Only the well-ordered Met side chains were retained by excluding those for which the SD atom had a temperature factor (or B-factor) > 30 $Å^2$. The polypeptide chains considered are given below (PDB code is followed by the subunit identifiers (if present) and the number of Met residues selected).

119L	3	153L	3	16PK	14	1A1I_A	0	1A28_A	4	1A2P_A	0
1A2Y_B	1	1A34_A	2	1A3C	2	1A4M_A	8	1A68	0	1A6G	2
1A7T_A	6	1A8D	2	1A8E	0	1A8I	17	1A9S	9	1A9X_B	1
1ABA	2	1AD2	1	1ADO_A	1	1ADS	5	1AFW_A	5	1AGJ_A	1
1AGQ_D	0	1AH7	5	1AHO	0	1AIE	0	1AIL	0	1AJ2	8
1AJS_B	5	1AK1	5	1AKO	6	1AKZ	0	1AL3	2	1ALO	18
1ALV_A	4	1AMM	7	1AMP	6	1AMX	0	1ANF	3	1AOC_A	0
1AOH_B	0	1AOQ_A	12	1AOZ_A	6	1APY_B	2	1AQ0_A	10	1AQ6_A	4
1AQB	3	1ARB	3	1ARV	7	1ATG	2	1ATL_A	4	1ATZ_A	4
1AVM_A	4	1AVW_B	0	1AWD	0	1AXN	4	1AYL	8	1AYO_A	1
1AZO	0	1B2N_AB	17	1BAM	4	1BBP_A	0	1BD8	2	1BDM_A	5
1BDO	2	1BEA	0	1BEB_A	4	1BEN_B	0	1BFD	11	1BFG	1
1BG0	8	1BGF	2	1BGP	2	1BIS_B	1	1BKF	2	1BKR_A	2
1BOL	4	1BRF	0	1BRT	0	1BTN	2	1BU7_A	8	1BV1	0
1BXA	5	1BXE	0	1BXO	0	1BYB	9	1BYQ_A	5	1C3D	6
1C52	2	1CBN	0	1CEM	5	1CEO	4	1CEW_I	2	1CEX	1
1CFB	2	1CHD	8	1CHM_A	12	1CKA_A	1	1CL1_A	7	1CLC	7
1CNV	1	1CPC_B	6	1CPO	2	1CSE_I	0	1CSH	8	1CTJ	1
1CYD_A	9	1CYO	0	1D2N_A	4	1DAD	2	1DEK_A	5	1DFX	0
1DKZ_A	6	1DOK_A	2	1DOS_A	5	1DPS_B	3	1DUN	0	1DUP_A	2
1DXY	4	1ECA	3	1ECP_A	10	1EDG	9	1EDM_B	0	1EDT	1
1EZM	8	1FDR	6	1FDS	3	1FIT	0	1FLE_I	1	1FNA	0
1FTR_A	7	1FUA	3	1FUR_A	10	1FUS	0	1FVK_A	1	1FWC_A	5
1G3P	1	1GAI	1	1GCI	3	1GD1_O	6	1GDO_A	1	1GIF_A	1
1GKY	2	1GSA	7	1GUQ_B	6	1HA1	3	1HFC	2	1HGX_B	0
1HOE	0	1HRD_A	11	1HTR_P	0	1HXN	2	1IDA_A	2	1IDK	1
1IDO	3	1IIB_A	2	1ISO	13	1ISU_A	1	1IXH	0	1JDW	10
1JER	0	1JFR_A	2	1JHG_A	1	1JPC	1	1KID	4	1KNB	1
1KPT_A	1	1KUH	1	1KVE_AB	3	1KVU	7	1LAM	11	1LAT_A	0

1LBU	4	1LCL	3	1LIT	0	1LKI	3	1LKK_A	0	1LMB_3	2
1LML	7	1LOU	1	1LTS_A	0	1LUC_A	7	1MAI	1	1MKA_A	5
1MML	0	1MOL_A	0	1MOQ	7	1MPG_A	6	1MRJ	2	1MRO_ABC	35
1MRP	1	1MSC	0	1MSI	4	1MSK	4	1MTY_BDG	20	1MUC_A	4
1MUG_A	0	1MZM	0	1NAR	1	1NBA_B	2	1NBC_A	1	1NCI_A	0
1NIF	7	1NKR	3	1NLR	1	1NLS	1	1NOX	1	1NP4	1
1NPK	2	1NUL_B	1	1NWP_A	4	1ONR_A	3	1OPD	0	1OPY	5
1ORC	0	1OVA_C	7	1OYC	6	1PBW_B	1	1PDA	2	1PDO	1
1PGS	2	1PHC	5	1PHM	4	1PHN_A	3	1PLC	2	1PMI	6
1PNE	6	1PNK_B	5	1POA	1	1POC	3	1POT	5	1PPN	0
1PRX_B	2	1PTY	7	1PUD	13	1QBA	15	1QNF	5	1RA9	2
1RCF	0	1REC	1	1REG_X	4	1RGE_A	0	1RHS	5	1RIE	2
1RMG	7	1RPO	1	1RSY	0	1RVA_A	0	1SBP	0	1SFP	1
1SGP_I	0	1SHK_A	0	1SLT_B	0	1SLU_A	0	1SMD	8	1SRA	0
1SVB	3	1SVP_A	4	1TAF_AB	3	1TCA	4	1THV	1	1TIB	0
1TML	7	1TVX_B	0	1TX4_A	1	1UAE	7	1UBI	1	1UNK_A	0
1URO_A	8	1UXY	2	1VCA_A	1	1VHH	1	1VID	7	1VIF	1
1VJS	5	1VLS	1	1VPS_B	5	1VSD	2	1WAB	2	1WDC_B	2
1WHI	1	1WHO	2	1WHT_B	2	1XGS_A	2	1XJO	1	1XNB	2
1XSO_A	0	1XYZ_A	12	1YAI_A	2	1YAS_A	2	1YCC	2	1YTB_A	1
1YVE_I	12	1ZIN	5	256B_A	3	2A0B	2	2ABK	1	2ACY	1
2ARC_B	2	2AYH	3	2BAA	2	2BBK_HL	6	2BOP_A	1	2CBA	1
2CCY_A	3	2CHS_A	6	2CTC	3	2CYP	5	2DOR_A	5	2DRI	3
2END	1	2ENG	1	2ERL	0	2FDN	0	2FHA	4	2FIV_A	1
2GAR	1	2GDM	1	2HBG	5	2HFT	0	2HMZ_A	1	2HTS	1
2I1B	2	2IGD	0	2ILK	5	2IZH_B	0	2MCM	0	2MSB_A	2
2NAC_A	7	2PGD	7	2PHY	4	2PIA	3	2PII	2	2PLC	5
2POR	3	2PSP_A	1	2PTH	1	2PVB	0	2QWC	5	2RN2	2
2RSP_B	2	2SAK	0	2SCP_A	10	2SIC_I	3	2SN3	0	2SNS	4
2SPC_A	2	2TGI	1	2TYS_A	1	2VHB_B	1	3CHB_D	3	3CHY	3
3CLA	5	3COX	12	3CYR	1	3DAA_A	4	3GCB	11	3GRS	8
3LZT	2	3NUL	0	3PCG_M	4	3PTE	5	3PVI_A	2	3SDH_A	3
3SEB	6	3SIL	5	3TDT	3	3VUB	4	4BCL	2	4MT2	0
4PGA_A	4	4XIS	6	5CSM_A	1	5HPG_A	1	5P21	3	5PTI	0
5PTP	2	6CEL	6	6GSV_A	6	7AHL_A	1	7RSA	4	8ABP	6

The main-chain carbonyl oxygen and aromatic atoms (belonging to all the chains present in PDB files, but not applying any symmetry transformation) within a distance of 3.8 Å from the atom SD of the Met side chain were identified. If more than one atom of a given aromatic side chain is within this range, the one with the shortest distance was assumed to be in contact. The stereochemistry of the nonbonded contact, X, relative to the CG-SD-CE moiety of Met was described using the spherical polar coordinates shown in Fig. 1. θ is the polar angle between the normal to the sulfide plane and the SD···X vector (if $\theta > 90^o$, θ is made equal to $180^o - \theta$, so that contacts above or below the plane are assumed to be equivalent; *i.e.*, $0^o \leq \theta \leq 90^o$). ϕ is the azimuthal angle between the extension of the bisector of the CG-SD-CE angle and the vector from SD to the projection of the atom X on the sulfide plane. No distinction is made between the directions in which ϕ is measured (*i.e.*, $0^o \leq \phi \leq 180^o$). Our convention of ϕ is slightly different from the one (ϕ_R) used by Rosenfield *et al*. (10) ($\phi = 180^o - \phi_R$). d is the distance between the sulfur and the contacting atom.

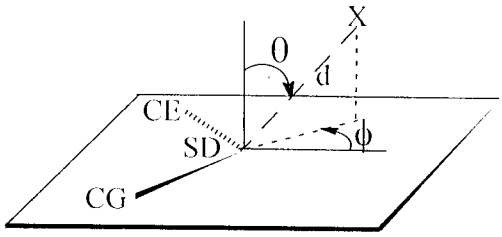


Figure 1: Spherical polar coordinates describing the position of atom X in contact with the Met sulfur atom (SD).

When the Met sulfur is in contact with an aromatic atom, a similar coordinate system, with the origin at the concerned aromatic atom, is used to define the position of SD relative to the aromatic ring. Depending on the value of θ_{Ar} being less than or greater than 45º, SD is assumed to be interacting with the face or the edge of aromatic ring, respectively. Likewise, molecular axes were also used to define the spherical polar angles (θ_{CO} and ϕ_{CO}) of SD, having an S···O contact, relative to the carbonyl plane. The origin was placed at the O atom and ϕ_{CO} measured from the extension of the C-O direction.

The solvent accessible surface area (ASA) of the Met residues was computed using the program ACCESS (16), which is an implementation of the Lee and Richards (17) algorithm. The solvent accessibility (in percentage) of each residue was calculated by dividing its ASA by the reference value (193.69 $Å^2$) obtained from an

extended tripeptide, Ala-Met-Ala. The secondary structures were assigned in accordance with the algorithm DSSP (18). The notations H and G represent helical structures (α and 3_{10}, respectively); B, isolated β-bridge; E, β-strand, S and T, turns; and C, regions of no regular structure. The molecular plots were made using MOLSCRIPT (19).

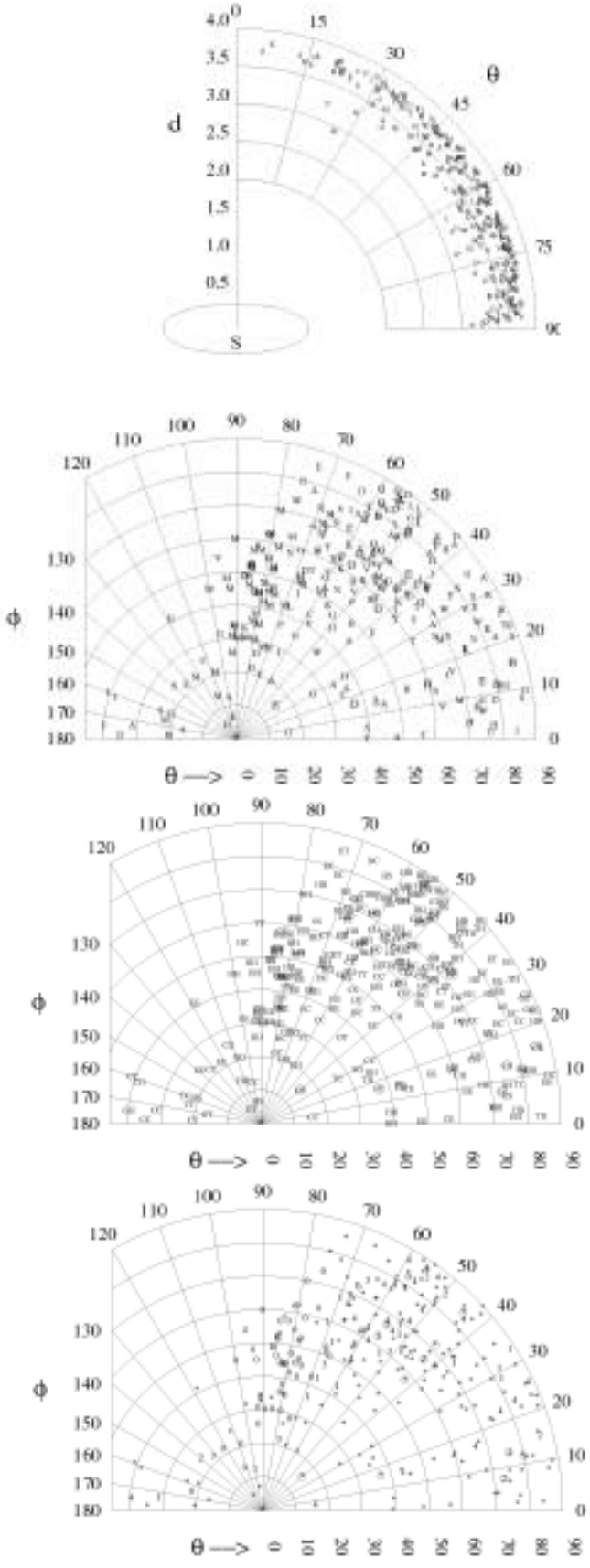


Figure 2: Polar graphs of **(a)** d (Å) vs. θ and

(b) of θ vs. ϕ (both in degrees). Each point is indicated by the one-letter amino acid code of the residue providing the oxygen atom.

(c) Same as in (b), except that a point is represented by two letters standing for the secondary structural elements of Met and the O-containing residue.

(d) Same as in (b) except that the labels, 1, 2, 3, 4 and * are used, corresponding to the absolute value of the sequence difference between the two residues (109 cases with values ≤ 4 and 166 '*' cases when they are > 4). The labels are encircled in (b) - (d) when O is from the carboxylate side chain of Asp or Glu.

Nonbonded interactions provide stability to the native structure of proteins (20). Such interactions, barring a few rare exceptions involving the cysteine residues (21-22), involve a proton. Substantiated by the observations in small molecule crystal structures, we wanted to see if there could be a direct S···O contact with specific stereochemistry involving the Met sulfur atom. To avoid any ambiguity due to the presence of any protons, we excluded hydroxyl oxygens and restricted ourselves only to the main-chain carbonyl (and a few cases of side-chain carboxylate) oxygen atoms. As a large percentage of Asn and Gln residues have the side-chain amide groups wrongly oriented (23), the oxygen atoms of these side chains were also debarred. Because of the electron-rich nature of the face of the aromatic ring, we also wondered if the Met S interacting with the face of an aromatic side chain has features similar to S···O interaction. For comparison, S interacting with the edge of aromatic rings was also analyzed. Such partitioning of the interacting groups/atoms is important, because a recent study addressing some of the issues raised here, but putting all C atoms (aliphatic as well as aromatic) in one category, and all nonhydroxyl oxygen atoms in another and also retaining all contacts up to a distance of 4.5 Å, found that the stereoelectronic requirements around the sulfur Met do not dictate the direction of approach of other atoms to the sulfur (24). It needs to be pointed out that the use of a longer cut-off distance introduces residues which are not really in contact, and this can obfuscate the visualization of any pattern in the contact geometry.

Out of a total of 1928 Met residues in 393 polypeptide chains 1276 residues passed the selection criterion of temperature factor and constitute our database. Of these, 22% (a value comparable to those Met S atoms which are within 4 Å of either a hydroxyl group or a nitrogen (3)) exhibit S···O interaction (with an average distance 3.6(2) Å), 8% interact with an aromatic face (S···aromatic-atom distance being 3.6(1) Å) and 9% are in contact with an aromatic atom at the edge (3.7(1) Å). 3% Met residues interact with more than one oxygen atoms and 2% engage both oxygen and aromatic atoms simultaneously.

(a) Geometry of S···O Interaction and the Secondary Structural Features of the Residues

Rosenfield *et al*. (10) have shown that nucleophiles (like carbonyl oxygen atoms carrying a partial negative charge) approach divalent sulfur along the prolongation of its covalent bonds, with θ lying between 60 and 90º and ϕ close to 50º (Fig. 1), thereby interacting favourably with the LUMO (lowest unoccupied molecular orbital) centred on sulfur. Results in proteins essentially reproduce these features (Fig. 2), with 70% of the 263 points having $\theta > 50º$. Barring some intra-residue interactions (discussed below) with steric constraints, the distribution is clustered with ϕ values in the range 30 to 60º. Of the twelve additional carboxylate oxygen atoms found in contact with SD, 50% have $\theta > 50º$ and ϕ in the range 30 to 60º.

From Fig. 2(b) it can be seen that most of the points with $\phi > 65º$ are Met residues (these also have lower θ values starting at about 70º and decreasing all the way to ~10º). These are the cases of Met SD interacting with the oxygen atom of the same residue and are usually located in helices (Figs. 2(c) and 3(a)). Of the 48 cases with intra-residue S···O interaction, 32 residues belong to helices (including one 3_{10}-helix), 11 to β-sheets and 5 to turns.

The most common combination of the secondary structural elements of the two interacting residues is 'HH' (60 cases). 30 examples of 'EE' combination are also observed, suggesting that the interactions would be important in stabilizing the secondary structure. For the 'HH' cases, the two dominant motifs are the methionine SD interacting with its own oxygen atom, or the atom preceding by 4 residues (i.e.,

Figure 3: (a) Intra-residue S···O interaction (with d, θ and ϕ values of 3.7 Å, 74º and 76º, respectively) involving the helical residue, Met153, in the structure, 1CPO. (b) (*i*, *i*-4) S···O interaction (with parameters of 3.32 Å, 87º and 60º, respectively) in a helix, involving Met148 and the carbonyl group (whose normal intrahelical hydrogen bond is also shown) of Gln144 in 1BG0.

(b)

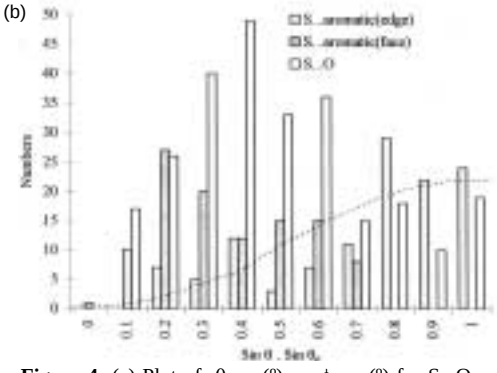
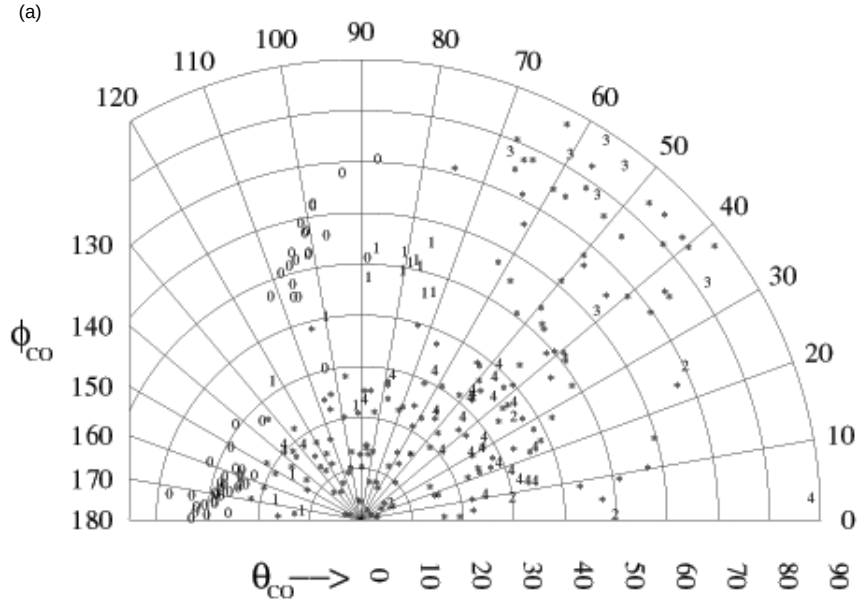


Figure 4: (a) Plot of θ_{CO} (º) vs. ϕ_{CO} (º) for S···O contacts, when the spherical polar angles of S are defined relative to the carbonyl group. Points are represented by labels corresponding to the sequence difference, as used in Fig. 2(d). (b) Histogram showing the distribution of $\sin\theta.\sin\theta_P$ (where θ_P is θ_{CO} or θ_{Ar}, the polar angle for S defined relative to the partner carbonyl or aromatic plane). The curve corresponds to the random distribution of the S···aromatic(edge) cases.

SD(*i*)···O(*i-4*)) (31 and 13 cases, respectively) (Fig. 3); in the rest, the residues are from two different helices. When the secondary structure is 'EE', the sequence difference between the two residues (Met - the one interacting) is 0 (for 11 cases), 1 (4), -1 (4) or -2 (1) when they belong to the same strand and in further 7 cases they are from two different strands of the same β-sheet.

As the oxygen atoms show pronounced directionality relative to the sulfide plane, it is also of interest to see if the reverse is also true, *i.e.*, if the sulfur atoms also have a fixed location with respect to the carbonyl group. The distribution (Fig. 4(a)) of the angles, θ_{CO} and ϕ_{CO}, which look at the position of SD from the perspective of the main-chain carbonyl group show that the majority of the sulfur atoms have $\theta_{CO} < 40^o$, i.e., they are within 40º from the normal to the carbonyl plane at the oxygen atom (as can also be seen in Fig. 3(b)), suggesting that the p_π orbital on O is directed towards SD. This geometry is different from that of a hydrogen bond donor, which is usually located along the sp^2 lone-pair direction (at $\theta_{CO} \approx 90^o$ and $\phi_{CO} \approx 60^o$) of the carbonyl oxygen atom (12). It is very interesting to see that depending on the sequence difference between Met and the CO-containing residue, the points are distributed in distinct regions in Fig. 4(a). Thus for an S···O contact, both the sulfide and the carbonyl planes have fixed orientations with respect to each other. For a random distribution of the two planar moieties, geometric consideration dictates that the number of contacts should be proportional to $\sin\theta.\sin\theta_{CO}$, which is not the case (Fig. 4(b)). We also wanted to see if the restriction in θ_{CO} to be below 40º is suggestive of the inherent requirement of these functional groups to interact, or the S atoms are made to be out of the CO plane by the presence of other in-plane groups hydrogen bonded to the carbonyl oxygen. Of the 263 S···O interactions, in 19 cases there is no hydrogen bonded partner within 3.6 Å≈ of the oxygen atom. In two-third of these θ_{CO} is below 40º, suggesting that the geometry observed is not a consequence of other competing interactions.

(a)



The observations discussed here have relevance beyond Met residues. The sulfur atoms in disulfide bonds should have similar chemical nature as Met S, and their hydrogen bond forming tendency is also very weak (4) and many of these half-cystine sulfur atoms are reported to have short contact with carbonyl oxygen atoms (3). Geometrical characterization of these interactions in an analogous manner may help us understand why some engineered disulfide linkages are more stable than

the others (25-26).

(b) Solvent Accessibility of Met Residues Showing S···O Interactions

The solvent accessibility values of Met residues (Fig. 5) indicate that most of the residues are buried (relative accessibility ≤ 10% with negligible accessible surface area for the S atoms), suggesting that the S···O interaction may have a useful role in the protein core. Indeed, an analysis aimed at finding out the frequency with which potential hydrogen bond donors and acceptors are satisfied found that 1.8% carbonyl groups, in spite of being buried, fail to hydrogen bond (27). It is possible that some of these are engaged in S···O interaction. Of the 263 S···O interactions that we have identified, the carbonyl group in 19 cases are bereft of any hydrogen bond interaction. Thus Met residues are unique in the protein core in that they can have hydrophobic interactions and also engage polar oxygen atoms at the same time.

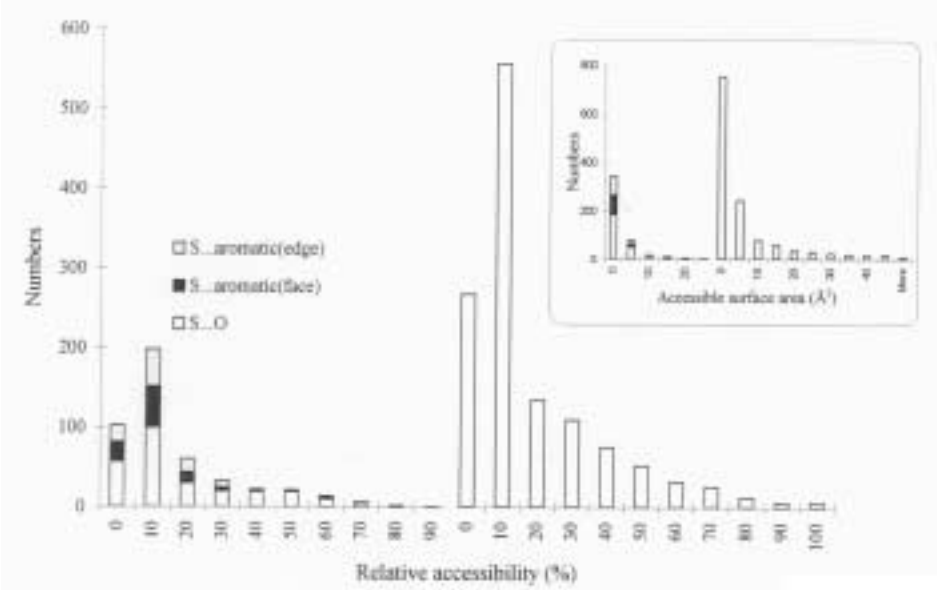


Figure 5: Relative solvent accessibility of Met residues engaged in different types of interactions. The general distribution is shown to the right. Likewise, the two distributions of the accessible surface area of the sulfur atom are given in the inset.

(c) S···aromatic(face) Interaction and Its Implications

When interacting with the face of an aromatic ring the orientation of the Met sulfide plane is essentially similar in all the structures (Fig. 6). In majority (53%) of the 107 cases θ is > 70º. ϕ is mostly concentrated between 40 and 60º, suggesting that in these interactions, as in S···O interaction, S behaves as the electrophile and the ring face a nucleophile. In the deviant cases, for example, when θ < 30º, it is found that almost 50% of the aromatic residues are Tyr. An interesting observation is in 50% cases involving His residues (27 examples), the sulfur atom is on the atom ND1 (11 cases) or NE2 (3) (Fig. 7). This is quite similar to the occurrence of an X-H group (X = C, O or N) on the face and pointing towards the N atom of a heterocyclic ring (28-30).

It is generally believed that in the protein interior the contacts between apolar side chains do not have any spatial orientation (31). In contrary, we find that when sulfur interacts with the face of an aromatic ring, the orientation of the sulfide plane is also fixed (the two rings essentially being perpendicular). This means that the hydrophobic interactions (*i.e.*, the burial of nonpolar surface) - which would be the maximum had the two rings being parallel is not the primary factor; rather the optimum overlap of molecular orbitals offered by this relative orientation between the two groups lead to a better stabilization. Biochemists have long argued on the underlying factor that causes the association of two nonpolar moieties in aqueous solution - whether the aggregation results because they are mutually excluded from water (entropy-driven) or because the nonpolar moieties experience a selective attraction for one another (favourable enthalpy). There are results showing the

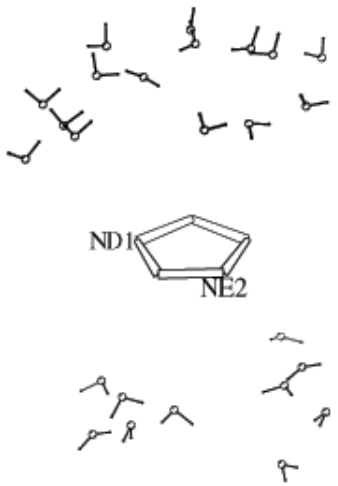


Figure 7: Scatterplot showing the disposition of different Met thioether groups against the face of the His ring. The Met side-group has been reduced five times its original size.

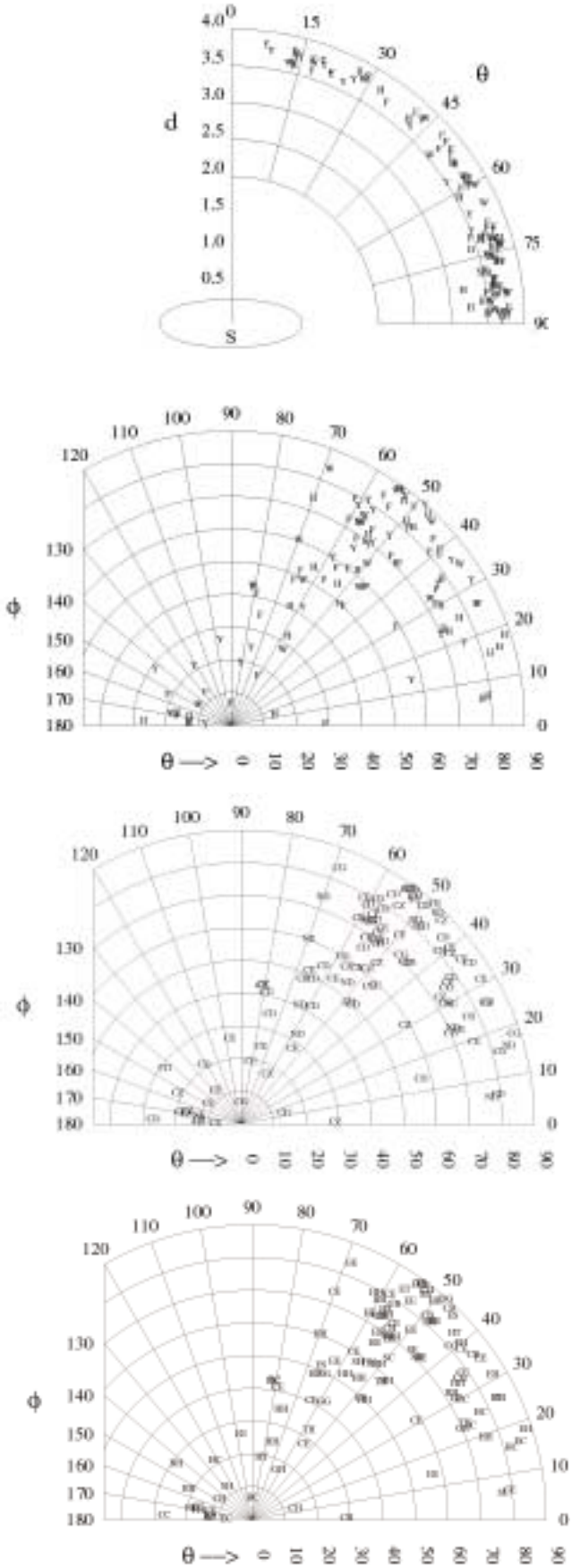


Figure 6: Plots of (a) d (Å) vs. θ (º) and (b - d) θ (º) vs. ϕ (º) for S···aromatic(face) contacts. Points are represented by one-letter amino acid code of the aromatic residue in (a) and (b), the label of the closest aromatic atom in (c). For clarity only the first two letters of the label are shown (leaving out the numeral which usually occupies the third position); this diagram shows the preponderance of the atoms ND1 (shown as ND) when His is interacting. In (d) the two letters stand for the secondary structural elements of Met and the aromatic residue.

manifestation of large enthalpy changes for some complexation phenomena involving nonpolar surfaces in aqueous solution (for details, see reference 32). The specific geometry observed for the interaction of the thioether group of Met with an

aromatic face, and likewise, the positioning of a C-H group on top of the N atom of heterocyclic rings, as observed in the interaction of protein residues with adenine (28) and the indole ring of Trp (30), may indicate a considerable contribution of enthalpy towards the free energy of binding.

As in S···O interaction, two helical residues are most likely to show S···aromatic(face) interaction (Fig. 6(d)). Of the 20 'HH' cases, in 5 both are from the same helix and have the aromatic residue and Met in (*i*, *i*+4) disposition (Fig. 8(a)). Though Klingler and Brutlag (33) reported 17 (*i*, *i*+4) Phe-Met pairs (however, using a larger contact distance cut-off of 5 Å), mostly in helices in different structures, the relative orientation of the two groups and the distance between them were not analyzed. Using short helical polypeptides Viguera and Serrano (34) showed that the interaction between the side chains of Phe and Cys or Met at positions *i* and *i*+4, respectively, can contribute up to 2 kcal/mol to the stability of the α-helix; additionally, it was suggested that the S atom is located at the edge of the aromatic ring. However, our results show that Met sulfur can favourably interact with the face of any aromatic residue, His in particular, and also provide a rationale for this observation. The interaction between cysteine sulfur and aromatic residues also showed a similar stereochemistry (22). Except one, 12 'EE' cases have the two residues in different strands of the same or different β-sheets.

Besides stabilizing helical structures, S···aromatic(face) interaction can also be useful in the association between different subunits in oligomeric proteins, as can be seen in Fig. 8(b). Analysis of protein-protein recognition sites has mainly dealt with the number of hydrogen bonds present in the interface (35). But it needs to be pointed out that interactions identified here can also be as important as hydrogen bonds in the formation of the quaternary structure and protein complexes.

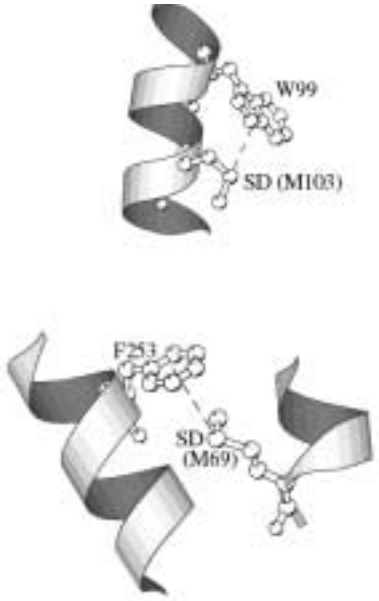


Figure 8: (a) Aromatic(*i*)···Met(*i*+4) interaction (with d, θ and ϕ values of 3.68 Å, 72º and 33º, respectively) in the structure, 1MTY (subunit D). (b) Intersubunit S···aromatic(face) interaction (with parameters of 3.72 Å, 53º and 57º, respectively), taken from the structure, IMRO, where Met69 belongs to the subunit C and the aromatic ring to subunit B. Both the residues are located in helices, part of which are shown.

(d) S···aromatic(edge) Interaction and Some Examples of Hydrogen Bonding

Earlier analyses on the interaction of S with the aromatic ring gave conflicting results. While Morgan *et al*. (36) found a preference for the π electron cloud of the aromatic ring, Reid *et al*. (37) found a preponderance of S atoms at the edge. In our analysis, S atoms are almost equally located on the face and edge (in the ratio 107:120). Given that more space is available at edge than on face for interaction with the aromatic ring, a random distribution would suggest a ~ 4:10 distribution of S atoms on face and edge. Consideration of the expected numbers of occurrences at different relative orientations of the two planar groups also suggests that the interaction with the face (unlike the edge) is significantly different from the random distribution (Fig. 4(b)), which is also in accordance with what has been found for Cys···aromatic interaction (22). That the face may provide a greater interaction energy is indicated by the occurrence of shorter contacts at higher θ values (Fig. 6(a)), while there is essentially no variation of d with θ at the edge (Fig. 9(a)).

Unlike in the face, S atoms are not restricted to higher (> 70º) θ values while at the edge (Fig. 9). Similarly ϕ values are also spread over the 0 to 60º range. This suggests that at the edge of an aromatic ring the interaction between the aromatic residue and Met is more of hydrophobic nature with no particular stereochemistry.

Though we set out to analyze non-hydrogen bond interaction, a few well-defined cases of hydrogen bonds have been encountered and their geometry can be used to characterize other possible hydrogen bonds involving the Met S atom. This will be particularly useful, as the earlier studies on hydrogen bonds involving Met sulfur could not discern any angular preferences (3-4). The ring N-H group of Trp and His side chains can act as proton donor along the edge of the aromatic ring and 14 cases (7 examples each of Trp and His) of S···N contacts may represent hydrogen bond interactions. Though proton positions have not been used in the analysis, it can be seen from the one example shown in Fig. 10 that the N-H group points towards the

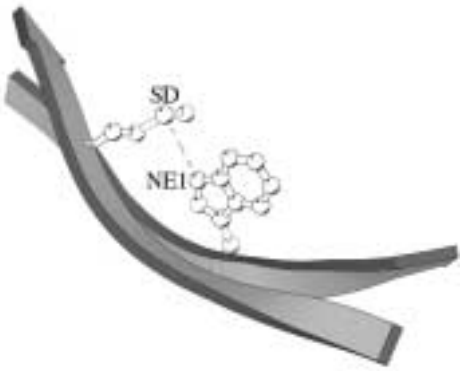


Figure 10: Diagram showing the relative orientation of the aromatic side chain of Trp223 and the sulfide group of Met149, linked by NH···S hydrogen bond (proton not shown; d, θ and ϕ values are 3.64 Å, 10º and 146º, respectively) in the structure, 4BCL. The two residues are from two strands of a β-sheet.

S atom. Majority of the cases have θ < 40º, which fits into the framework of protons (*i.e*., electrophiles) interacting with the lone-pair orbital on S atoms (nucleophiles) as proposed by Rosenfield *et al*. (10)

From this it can be inferred that all cases of S···N contact involving planar >N-H groups (where the position of the proton is fixed within the group), can be taken as hydrogen bonds if the condition on θ is satisfied. It is debatable if a hydroxyl group, in which the position of the proton is not fixed in space and whose function as a proton donor would involve a reduction in conformational entropy, would have

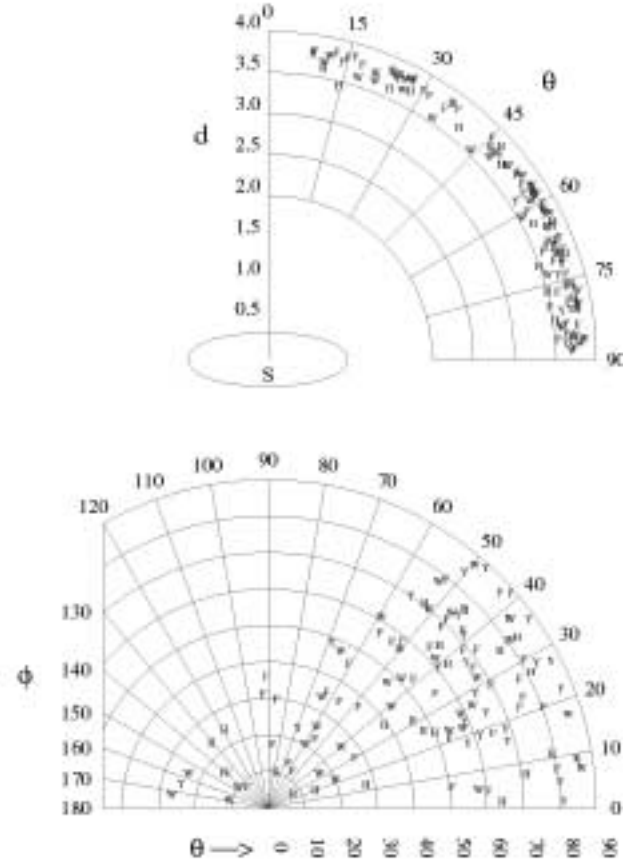


Figure 9: Plot depicting geometrical features of S···aromatic(edge) interaction.

enough enthalpic gain so as to form a hydrogen bond with Met S or it would rather have an S···O contact.

As can be seen from Fig. 10, hydrogen bond between Met sulfur and a heteroaromatic side chain can be a stabilizing force in holding two β-strands together. In 35% cases of hydrogen bonding, the two residues are from two different strands. Of the 120 S···aromatic(edge) interactions, there are 27 'HH' and 21 'EE' cases. In the former, 4 belong to the same helix and the sequence difference between the two residues is 4 in two cases and 1 in two others. Thus the specific (*i, i*+4) aromatic-Met pairs observed in helices when SD interacts with the face of the aromatic ring is not found when the interaction is with the aromatic edge.

(e) χ_3 Torsion Angle and Met Rotamers for the Interacting Residues

We wanted to see if different interactions involving the sulfur atom have any effect on the χ_3 torsion angle which spans the thioether group. The general distribution (Fig. 11, right) does not get altered significantly by the presence of any specific interaction (Fig. 11, left). Using the terminology *gauche*$^+$ (g^+), *gauche*$^-$ (g^-) and *trans* (*t*) to represent the bin of angles in the range -120 to 0º, 0 to 120º and 120 to

-120°, *t*: g^+ : g^- ratios for the above two distributions are 1:2.0:1.7 and 1:2.2:2.1. A ratio of 1:1.8:1.5 has been reported by Word *et al*. (38)

The preference of the *gauche* conformation over *trans* has been commented upon

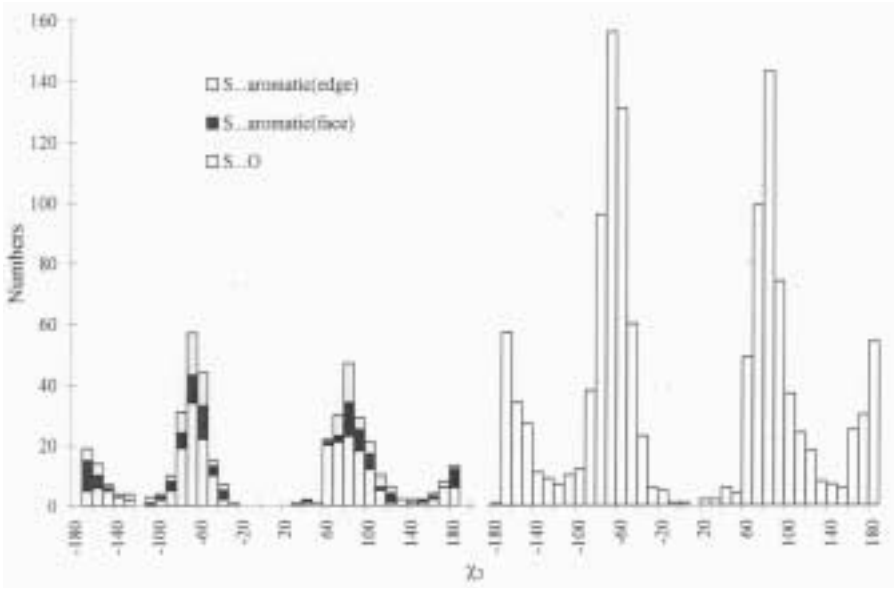


Figure 11: Histogram showing the distribution of the torsion angle, χ_3 (CB-CG-SD-CE) of Met residues having different types of interaction (left) and in the whole dataset (right).

by Gellman (32) and Word *et al*. (38) This may be because the orientation of the sulfide plane (which affects χ_3) is chosen to achieve the optimum overlap of its molecular orbital with that of the interacting group, and for many of the short range interactions (see below) the *trans* conformation is not needed.

Some recent studies find (g^+ g^+ g^+) rotamer to be the most common for Met residues (38-40). However, this rotameric state is not found when the Met sulfur interacts with its own oxygen atom or another one close in sequence. For example, in the case of intra-residue S···O interaction (Figs. 2(d) and 3(a)), when the residue is in helix there are 13 cases of (t g^+ g^+), 17 of (t g^- g^-) and a lone case of (t g^- t); the first one is also one of the rotamers found when the residue is in β-sheet, the other being (g^- g^- g^-) (6 and 5 instances, respectively). Of the 6 remaining cases (the residues being in turn conformation), in half the rotamer observed is one of the above three predominant ones. In all the 13 cases of intrahelical SD(*i*)···O(*i-4*) interaction (Figs. 2(d) and 3(b)) the (g^+ t g^-) rotamer is observed, which is also preferred in 7 (4 cases of the secondary structure 'TH', 2 of 'HC' and 1 of 'HT') out of 11 remaining cases where Met precedes the interacting oxygen by 4 residues.

(f) Implications for Mutational Studies

Various protein engineering experiments have replaced Met by other residues or vice versa, and studied the effect of mutation on the stability. Gassner *et al*. (41) have replaced up to 10 adjacent core residues (like Leu, Ile and Phe) of T4 lysozyme by Met which occupies roughly the same volume. The stability of the single mutants was lowered (0.4 to 1.9 kcal/mol), as expected from the extra side-chain flexibility and different shapes. The most conserved replacement of Met is by norleucine, Nle (substitution of -S- by a -CH_2- group), which occupies similar volume and has similar side-chain conformational properties. The replacement of Met13 in ribonuclease S does not perturb the stability or function of the protein (42). Such substitutions in other proteins have yielded similar results (43-45). There are some examples where the replacement of Met by other residues made significant difference to the local structure or the stability of the protein. It would be of interest to see if these residues have any specific interaction which got dis-

rupted on substitution. In one example, two Met residues (192 and 213) in the primary specificity pocket of α-lytic protease, when replaced by Ala, gave mutant proteases with extraordinary broad specificity profiles (46). It has been suggested that the wild-type enzyme is a rigid molecule which prefers Ala in the P_1 position of the substrate. Mutations make the molecule more flexible so that large as well as small P_1 side chains can be accommodated with equal ease. In another example (47), azurin from *A. denitrificans*, which has a Cu site, is coordinated by 3 strong ligands (two histidines, one cysteine) arranged in a nearly trigonal planar configuration. A fourth weak ligand (Met121) occupies an axial position. The crystal structures of the Met121His mutant at pH 6.5 and pH 3.5 show that the metal binding cavity is more flexible than expected. At the high pH His121 is added as the fourth strong ligand, whereas in the low pH structure His121 moves away from the copper and a nitrate molecule takes its position. In the third example, Met6Leu mutant of T4 lysozyme has been studied (48). The substitution of Met with a leucine residue within the interior of a protein is expected to increase stability both because of a more favourable solvent transfer term as well as the reduced entropic cost of holding a Leu side chain in a defined position. Together, these terms are expected to contribute about 1.4 kcal/mol to protein stability. However, M6L is significantly destabilized (-2.8 kcal/mol). Though globally the structure is similar to the wild-type molecule, it has large local structural perturbations. The common feature in all these cases is the loss of rigidity which can not simply be explained by a change in

Table I
Specific interactions involving Met residues in the wild-type structure whose replacement by other residues cause structural changes

PDB file	Protein name	Met residue	Interacting atom	Parameters d(Å)	θ(º)	θ(º)	Comment
2ALP	α-Lytic protease	213	NE1-W199	3.43	18	23	hydrogen bonding
1AZU	Azurin	121	ND1-H46[a]	3.23	83	53	S···aromatic(face)
206L	T4 lysozyme	6	ND2-N101	3.37	6	148	hydrogen bonding

[a] There are two other S O contacts, but the geometries are not good.

the shape and size at the point of mutation. However, in all these (Table I), the wild-type structure has a specific interaction with a good geometry, which is lost when the residue is replaced, thus introducing flexibility into the local environment. From this it can be suggested that the replacement of a Met residue involved in any of the stabilizing interactions identified here by other hydrophobic residues may not lead to isostructural or isoenergetic mutants.

Conclusions

Though considered a hydrophobic residue which can conservatively be replaced by norleucine (*n*-butyl side chain) in mutational studies and which does not show much inclination to form hydrogen bond, methionine in protein structure can still be close to carbonyl (and even carboxylate) oxygen atoms. Bereft of any proton, these are the manifestation of direct S···O contact with the oxygen atom (acting as a nucleophile) approaching the divalent sulfur (electrophile), CG-SD-CE, along the extension of the CG-SD or CE-SD bond (Figs. 1, 2 and 3). Even with respect to the carbonyl plane, SD is located on top of the oxygen atom (Fig.4), so that the sulfide and the carbonyl planes linked by an S···O contact have well-defined relative orientation. Distinct stereochemistry is observed even in the interaction of the divalent sulfur with the aromatic side chains of nonpolar residues. If located on the π-electron rich aromatic face (which thus has a nucleophilic character), the sulfide plane is so oriented as to make one C-S bond point towards the face (Fig. 6), and in particular towards the N atom in His ring (Fig. 7). No fixed orientational geometry is seen when S is located at the edge of an aromatic ring (Fig. 9). The N-H group of a Trp or His ring can form hydrogen bond if it is directed towards the S atom within 40º from the perpendicular to the sulfide plane (Figs. 9 and 10). In such a situation, as in the interaction of metal ions with the Met S (11), the proton acts as an electrophile and the divalent sulfur a nucleophile. Such stereochemical selectivities are reflective of what have already been observed in small molecule struc-

tures (10).

The S···O interactions, mostly observed in the protein core (Fig. 5), serve the important role of engaging a polar atom, which otherwise might have been left with its bonding potential unsatisfied in a nonpolar milieu. Despite the normally accepted view that nonpolar residues do not exhibit any specific contact geometry, we find that the divalent sulfur of Met residues has a preferred orientation when interacting with an aromatic face, suggestive of the attractive nature of such interaction not generally seen between nonpolar surfaces. Thus Met residues are tailored for strong interactions with nonpolar surfaces (specifically, the aromatic face, as shown in Fig. 8) on binding partners, and also capable of engaging oxygen atoms through S···O interaction and N-H groups through hydrogen bonding**.** The flexibility offered by the three side-chain torsion angles, as revealed by the capability of the side-chain to fold onto itself so as to make an intra-residue S···O contact (Fig. 3(a)), makes the residue adapt itself to partners of different shapes. Thus Met residues should be particularly useful in molecular recognition, and indeed the binding sites of several proteins are rich in Met residues (see reference 32). As Met and Trp side chains can be in stereospecific contact through hydrogen bonding and S···aromatic(face) interactions, the former has been found to have a high propensity to be in close association with Trp (30) and interestingly, tryptophanyl-tRNA synthetase (49) has a Met in the active site, while methionyl-tRNA synthetase (50) has Trp. Analysis of specific interactions of the sulfur atom would be useful in understanding results from studies that deal with the replacement or introduction of Met residues in protein structures.

Acknowledgements

The authors would like to thank the Council of Scientific and Industrial Research for a fellowship (to DP) and a grant (to PC), and the Department of Biotechnology for supporting the computational facilities.

References and Footnotes

1. C. Chothia, *Nature 248*, 338-339 (1974).
2. G.E. Schulz and R.H. Schirmer, *Principles of Protein Structure*, Springer-Verlag. New York, pp 12-14 (1984).
3. L.M. Gregoret, S.D. Rader, R.J. Fletterick and F.E. Cohen, *Proteins 9*, 99-107 (1991).
4. J.A. Ippolito, R.S. Alexander and D.W. Christianson, *J. Mol. Biol. 215*, 457-471 (1990).
5. F.H. Allen, C.M. Bird, R.S. Rowland and P.R. Raithby, *Acta Crystallogr. B53*, 696-701 (1997).
6. F.T. Burling and B.M. Goldstein, *Acta Crystallogr. B49*, 738-744 (1993).
7. F. Iwasaki, *Acta Crystallogr. C42*, 121-124 (1986).
8. Y. Nagao, T. Hirata, S. Goto, S. Sano, A. Kakehi, K. Iizuka and M. Shiro, *J. Am. Chem. Soc. 120*, 3104-3110 (1998).
9. S.S. Surange, G. Kumaran, S. Rajappa, D. Pal and P. Chakrabarti, *Helv. Chim. Acta 80*, 2329-2336 (1997).
10. R.E. Rosenfield, Jr, R. Parthasarathy and J.D. Dunitz, *J. Am. Chem. Soc. 99*, 4860-4862 (1977).
11. P. Chakrabarti, *Biochemistry 28*, 6081-6085 (1989).
12. G. Klebe, *J. Mol Biol. 237*, 212-235 (1994).
13. C. Pascard, *Acta Crystallogr. D51*, 407-417 (1995).
14. E.E. Abola, J.L. Sussman, J. Prilusky and N.O. Manning, *Methods Enzymol. 277*, 556-571 (1997).
15. U. Hobohm and C. Sander, *Protein Sci. 3*, 522-524 (1994).
16. S. Hubbard, *ACCESS: A program for calculating accessibilities.* Department of Biochemistry and Molecular Biology. University College of London, (1992).
17. B. Lee and F.M. Richards, *J. Mol. Biol. 55*, 379-400 (1971).
18. W. Kabsch and C. Sander, *Biopolymers 22*, 2577-2637 (1983).
19. P.J. Kraulis, *J. Appl. Crystallogr. 24*, 946-950 (1991).
20. S.K. Burley and G.A. Petsko, *Adv. Prot. Chem. 39*, 125-189 (1985).
21. P. Chakrabarti and D. Pal, *Protein Sci. 6*, 851-859 (1997).

22. D. Pal and P. Chakrabarti, *J. Biomol. Struct. Dynam. 15*, 1059-1072 (1998).
23. J.M. Word, S.C. Lovell, J.S. Richardson and D.C. Richardson, *J. Mol. Biol. 285*, 1735-1747 (1999).
24. O. Carugo, *Biol. Chem. 380*, 495-498 (1999).
25. S.F. Betz, *Protein Sci. 2*, 1551-1558 (1993).
26. B. Van den Burg, B.W. Dijkstra, B. Van der Vinne, B.K. Stupl, V.G.H. Eijsink and G. Venema, *Prot. Engng. 6*, 521-527 (1993).
27. I.K. M^{c}Donald and J.M. Thornton, *J. Mol. Biol. 238*, 777-793 (1994).
28. P. Chakrabarti and U. Samanta, *J. Mol. Biol. 251*, 9-14 (1995).
29. U. Samanta, P. Chakrabarti and J. Chandrasekhar, *J. Phys. Chem. A102*, 8964-8969 (1998).
30. U. Samanta, D. Pal and P. Chakrabarti, *Proteins 38*, 288-300 (2000).
31. J.M. Thornton, D.T. Jones, M.W. MacArthur, C.M. Orengo and M.B. Swindells, *Phil. Trans. R. Soc. Lond. B348*, 71-79 (1995).
32. S.H. Gellman, *Biochemistry 30*, 6633-6636 (1991).
33. T.D. Klingler and D.L. Brutlag, *Protein Sci. 3*, 1847-1857 (1994).
34. A.R. Viguera and L. Serrano, *Biochemistry 34*, 8771-8779 (1995).
35. L. Lo Conte, C. Chothia and J. Janin, *J. Mol. Biol. 285*, 2177-2198 (1999).
36. R.S. Morgan, C.E. Tatsch, R.H. Gushard, J.M. McAdon and P.K. Warme, *Int. J. Pept. Protein Res. 11*, 209-217 (1978).
37. K.S.C. Reid, P.F. Lindley and J.M. Thornton, *FEBS Lett. 190*, 209-213 (1985).
38. J.M. Word, S.C. Lovell, T.H. LaBean, H.C. Taylor, M.E. Zalis, B.K. Presley, J.S. Richardson and D.C. Richardson, *J. Mol. Biol. 285*, 1711-1733 (1999).
39. R.L. Dunbrack and F.E. Cohen, *Protein Sci. 6*, 1661-1681 (1997).
40. D. Pal and P. Chakrabarti, *Proteins 36*, 332-339 (1999).
41. N.C. Gassner, W.A. Baase and B.W. Matthews, *Proc. Natl. Sci. Acad. USA. 93*, 12155-12158 (1996).
42. J. Thomson, G.S. Ratnaparkhi, R. Varadarajan, J.M. Sturtevant and F.M. Richards, *Biochemistry 33*, 8587-8593 (1994).
43. A.M. Gilles, P. Marliere, T. Rose, R. Sarfati, R. Longin, A. Meier, S. Fermandjian, M. Monnot, G.N. Cohen and O. Barzu, *J. Biol. Chem. 263*, 8204-8209 (1988).
44. Z.I. Randhawa, H.E. Witkowska, J. Cone, J.A. Wilkins, P. Hughes, K. Yamanishi, S. Yasuda, Y. Masui, P. Arthur, C. Kletke, F. Bitsch and C.H.L. Shackleton, *Biochemistry 33*, 4352-4362 (1994).
45. T. Yuan and H.J. Vogel, *Protein Sci.* 8, 113-121 (1999).
46. R. Bone, A. Fujishige, C.A. Kettner and D.A. Agard, *Biochemistry 30*, 10388-10398 (1991).
47. A. Messerschmidt, L. Prade, S.J. Kroes, J. Sanders-Loehr, R. Huber and G.W. Canters, *Proc. Natl. Acad. Sci. USA. 95*, 3443-3448 (1998).
48. L.A. Lipscomb, N.C. Gassner, S.D. Snow, A.M. Eldridge, W.A. Baase, D.L. Drew and B.W. Matthews, *Protein Sci. 7*, 765-773 (1998).
49. V.A. Ilyin, B. Temple, M. Hu, G. Li, Y. Yin, P. Vachette and C.W. Carter, Jr. *Protein Sci. 9*, 218-231 (2000).
50. Y. Mechulam, E. Schmitt, L. Maveyraud, C. Zelwer, O. Nureki, S. Yokoyama, M. Konno and S. Blanquet. *J. Mol. Biol. 294*, 1287-1297 (1999).

Date Received: November 30, 2000

Communicated by the Editor Ramaswamy H. Sarma