R. J. Zauhar[1]
C. L. Colbert[2]
R. S. Morgan[3]
W. J. Welsh[4]

[1] *Department of Chemistry and Biochemistry, University of the Sciences in Philadelphia, 600 S. 43rd Street, Philadelphia, PA 19104-4495*

[2] *Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392*

[3] *Department of Biochemistry, Molecular and Cell Biology, The Pennsylvania State University, University Park, PA 16802*

[4] *Department of Chemistry, University of Missouri–St. Louis, 8001 Natural Bridge Road, St. Louis, MO 63121-4499*

Evidence for a Strong Sulfur–Aromatic Interaction Derived from Crystallographic Data

Received 18 September 1998; accepted 31 August 1999

Abstract: *We have uncovered new evidence for a significant interaction between divalent sulfur atoms and aromatic rings. Our study involves a statistical analysis of interatomic distances and other geometric descriptors derived from entries in the Cambridge Crystallographic Database (F. H. Allen and O. Kennard,* Chem. Design Auto. News, *1993, Vol. 8, pp. 1 and 31–37). A set of descriptors was defined sufficient in number and type so as to elucidate completely the preferred geometry of interaction between six-membered aromatic carbon rings and divalent sulfurs for all crystal structures of nonmetal-bearing organic compounds present in the database. In order to test statistical significance, analogous probability distributions for the interaction of the moiety X–CH_2–X with aromatic rings were computed, and taken a priori to correspond to the null hypothesis of no significant interaction. Tests of significance were carried our pairwise between probability distributions of sulfur–aromatic interaction descriptors and their CH_2–aromatic analogues using the Smirnov–Kolmogorov nonparametric test (W. W. Daniel,* Applied Nonparametric Statistics, *Houghton-Mifflin: Boston, New York, 1978, pp. 276–286), and in all cases significance at the 99% confidence level or better was observed. Local maxima of the probability distributions were used to define a preferred geometry of interaction between the divalent sulfur moiety and the aromatic ring. Molecular mechanics studies were performed in an effort to better understand the physical basis of the interaction. This study confirms observations based on statistics of interaction of amino acids in protein crystal structures (R. S. Morgan, C. E. Tatsch, R. H. Gushard, J. M.*

Correspondence to: Randy J. Zauhar; email: r.zauhar@usip.edu
Biopolymers, Vol. 53, 233–248 (2000)
© 2000 John Wiley & Sons, Inc.

McAdon, and P. K. Warme, International Journal of Peptide Protein Research, *1978, Vol. 11, pp. 209–217; R. S. Morgan and J. M. McAdon,* International Journal of Peptide Protein Research, *1980, Vol. 15, pp. 177–180; K. S. C. Reid, P. F. Lindley, and J. M. Thornton,* FEBS Letters, *1985, Vol. 190, pp. 209–213), as well as studies involving molecular mechanics (G. Nemethy and H. A. Scheraga,* Biochemistry and Biophysics Research Communications, *1981, Vol. 98, pp. 482–487) and quantum chemical calculations (B. V. Cheney, M. W. Schulz, and J. Cheney, Biochimica Biophysica Acta, 1989, Vol. 996, pp.116–124; J. Pranata,* Bioorganic Chemistry, *1997, Vol. 25, pp. 213–219)—all of which point to the possible importance of the sulfur–aromatic interaction. However, the preferred geometry of the interaction, as determined from our analysis of the small-molecule crystal data, differs significantly from that found by other approaches.* © 2000 John Wiley & Sons, Inc. Biopoly 53: 233–248, 2000

Keywords: *divalent sulfur atoms; aromatic rings; probability distributions; sulfur–aromatic interactions; CH_2–aromatic interactions; Smirnov–Kolmogorov nonparametric test*

INTRODUCTION

The hypothesis that a strong, favorable interaction exists between aromatic rings and divalent sulfur atoms was first proposed by Morgan and co-workers[1,2] to explain the high frequency of contacts observed in protein crystal structures between sulfur-bearing amino acids (cysteine and methionine) and those that include an aromatic ring (histidine, tryptophan, tyrosine, and phenylalanine). Not only were pairwise interactions observed, but also stacked arrangements of sulfur-bearing and aromatic residues, suggesting that a special interaction between these two categories of amino acid might have a special significance for stabilizing the folded conformation of proteins. A statistical analysis of interactions collected from structures in the Brookhaven Protein Data Bank confirmed that such interactions occurred much more frequently than would be expected under the assumption of random association between amino acids. A later study by Reid et al.[3] using a larger data set confirmed these observations, and in addition defined a preferred geometry for the interaction of aromatic rings and the sulfur-bearing side chains of cysteine and methionine. A maxima in the distribution of sulfur–aromatic distances was found at a separation of about 5.3 Å, and for those sulfurs within 6 Å of the ring centroid, the distribution of the angle of elevation of the sulfur above the plane of the ring showed a local maximum at about 30°–50° for cysteine sulfurs, and 45°–65° for methionine sulfurs. (These local maxima are superimposed on an overall maximum at 0°, which arises from the usual solid-angle weighting.)

Experimental studies of interactions between divalent sulfur-bearing and aromatic model compounds were carried out by Bodner and Morgan[4] by measuring nmr chemical shifts in mixtures of model compounds at a series of different concentrations and temperatures. Their results indicated the formation of 1:1 complexes of the sulfur-bearing and aromatic small molecules in CCl_4 solution, with a free energy of formation of about 1 kcal/mole. Computer simulations by Nemethy and Scheraga with the ECEPP molecular mechanics force field[5] were conducted using similar model compounds, the study involving energy minimization of various trial configurations of the interacting molecules. These computational results also indicated a stabilization of 1–2 kcal/mole upon bringing a divalent sulfur-bearing compound into contact with an aromatic ring.

Quantum mechanical studies have been carried out for the cysteine side-chain model methanethiol by Cheney et al.[6] Here the cysteine model interacted with a benzene ring, with a number of initial configurations used as starting points for quantum mechanical energy optimization at the HF/3-21G* level of theory. Each of the optimized configurations was used in subsequent MP2 single-point calculations, to introduce the effects of electron correlation. The most favorable configuration places the sulfur at 4.4 Å from the ring centroid, and at an elevation of 56° from the ring plane, in rough agreement with the statistical results for cysteine found by Reid et al. The energy of interaction (including correlation) for this configuration was −3.0 kcal/mole. In the minimum-energy configurations, the lone pairs of the sulfur are directed away from the center of the ring, while the methane group lies above the plane of the ring.

Additional experimental evidence of the importance of the sulfur–aromatic interaction has been found by Lebel et al.,[7] who directly measured downfield shifts of tyrosyl proton resonances induced by interaction with the disulfide bridge in oxytocin derivatives. Viguera and Serrano[8] investigated the contribution of phenyl-cysteine and phenyl-methionine interactions to α-helix stability using nmr and CD spectra, along with a computer algorithm, AGADIR,[9]

for predicting the helical content of peptides in solution. Their nmr results directly indicate a strong interaction between Cys or Met and Phe side chains, while the assumption of interaction energies of −2.0 kcal/mole between Cys and Phe residues and −0.65 kcal/mole between Met and Phe leads to an optimal prediction of peptide helical content by the AGADIR program.

The importance of the sulfur–aromatic interaction is also indicated by a recent *negative* result; Spencer and Stites[10] have performed site-directed mutagenesis to substitute a leucine for a methionine at position 32 in staphylococcal nuclease, and found a *decrease* in protein stability of 0.8 kcal/mole, where computer simulations[11] had predicted an *increase* in stability of 1.6 kcal/mole. This interesting result prompted new quantum mechanical studies by Pranata,[12] who calculated interaction energies for several optimized configurations of dimethyl sulfide and benzene at the MP2/6-31G* level of theory. Here electron correlation effects were included in the energy gradients that directed the geometric optimization, and thus the final configurations of the interacting molecules correspond to true minima at the MP2 level of theory. (In contrast, Cheney et al.[6] included electron correlation only at the end points of Hartree–Fock energy minimizations.) The best configuration identified by Pranata corresponds to an energy of interaction of −2.85 kcal/mole, with a sulfur–aromatic distance of 4.9 Å. While the angle of elevation of the sulfur above the plane of the benzene ring is not provided by the author, it is clearly at a relatively large angle, and thus this work again reproduces at least roughly the results of the statistical analysis of Reid et al. Again, the most stable configuration of the interacting molecules directs the sulfur lone pairs away from the aromatic ring.

Theory and experiment both indicate the presence of a special and significant interaction between divalent sulfurs and aromatic rings. To further investigate this phenomenon, we have turned to the Cambridge Crystallographic Database,[13] a repository of approxi-

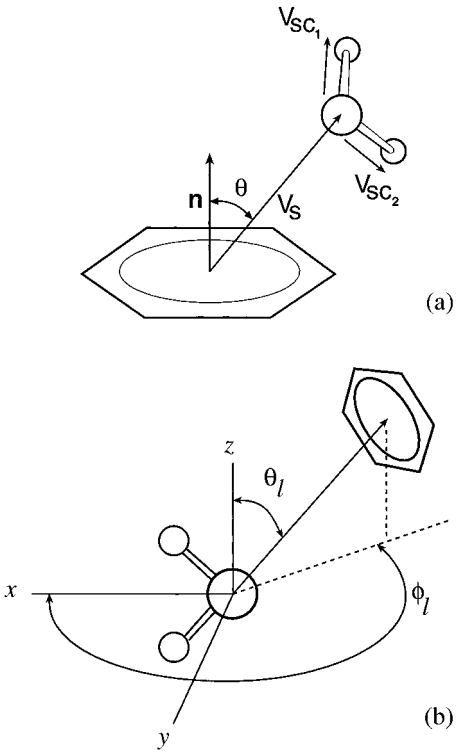


FIGURE 1 Vectors and angles defined for each database "hit." (a) Here, **n** is a unit normal to the plane of the phenyl ring, $\mathbf{V_S}$ the vector from the ring centroid to the sulfur, and $\mathbf{V_{SC_1}}$ and $\mathbf{V_{SC_2}}$ vectors from the sulfur to its bonded neighbors. The polar angle θ is defined as shown. (b) The special polar and azimuthal angles θ_l and ϕ_l are computed by the "lone pair" facility in the Cambridge QUEST program, and are relative to the local *x*–*y*–*z* coordinate system.

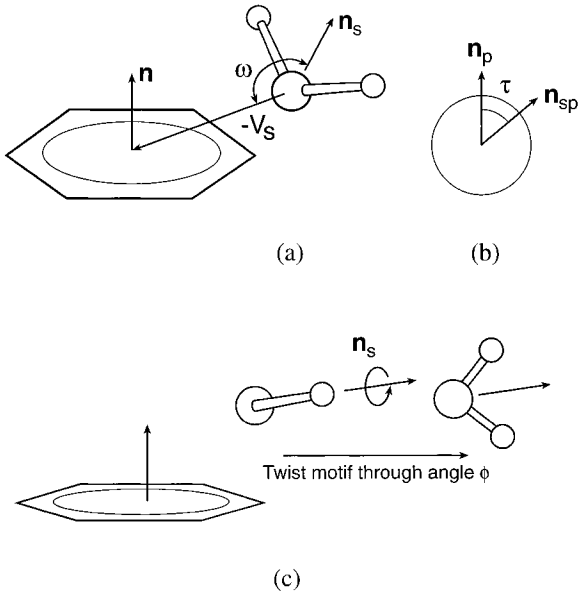


FIGURE 2 Definitions of angles used in our analysis. (a) The $\mathbf{n_s}$ is the perpendicular bisector of the C–S–C bond angle. Together, $\mathbf{V_S}$ and $\mathbf{n_s}$ define the angle ω. (b) The $\mathbf{n_p}$ and $\mathbf{n_{sp}}$ are the components of **n** and $\mathbf{n_s}$ perpendicular to $\mathbf{V_S}$; the angle between these components defines the torsional angle τ. (c) The angle ϕ is the propeller twist about $\mathbf{n_s}$ needed to make the C–S–C plane perpendicular to the plane of the aromatic ring. The angle is positive for a counterclockwise twist, negative for clockwise (as seen looking down the vector $\mathbf{n_s}$). Here the angle ϕ is $\pm 90°$. (The angle is computed by finding the rotation needed to make the vector $\mathbf{V_{SC_1}} \times \mathbf{V_{SC_2}}$ have zero projection along **n**, and can range from −180° to +180°.)

mately 189,000 crystal structures of small compounds, including a large number that contain both sulfurs and aromatic rings. We reason that if a strong sulfur–aromatic interaction does exist, then it will determine in part the crystalline form of compounds that include both these chemical motifs. Just as sulfur–aromatic interactions occur more frequently in folded proteins than would be expected on the basis of random packing, so too we expect to see an increased frequency of association between divalent sulfur atoms and aromatic rings in crystal structures of compounds that include these groups. Furthermore, given the large number of potential sulfur–aromatic contacts in the Cambridge Database, it should be possible to ascertain with some confidence the preferred geometry of this interaction.

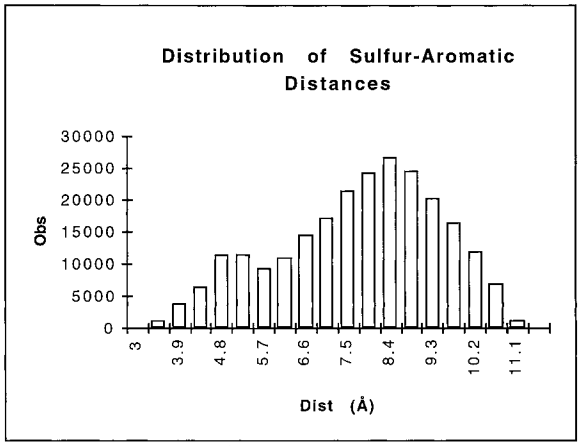


FIGURE 3 Histograms of observed distances to the aromatic ring centriod. Number of observations is plotted as a vertical bar for each radial bin (bin size = 0.45 Å). (a) Sulfur–aromatic distribution. (b) CH_2–aromatic (control) distribution.

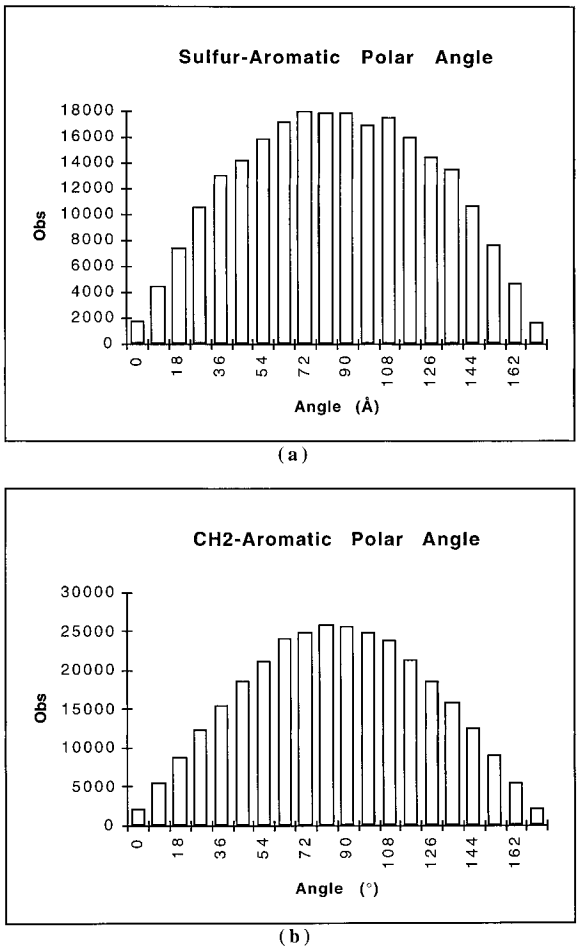


FIGURE 4 Histograms of the polar angle (θ) distribution for (a) sulfur–aromatic and (b) CH_2–aromatic (control) interactions. (Angular bin size 9°.)

METHODS

Data Set Generation

A query was designed to extract from the Cambridge Database the crystal structures of all compounds that contain at least one divalent sulfur (motif C–S–C) and at least one phenyl ring. A total of 2266 structures were selected, and for each structure all sulfur–aromatic contacts were identified with a sulfur-to-ring centroid distance of 10 Å or less (the search actually collected interactions a bit beyond the 10 Å limit). For each contact, all of the vectors and angles shown in Figure 1 were computed and saved. A file was generated, each line of which contained all of the angles defined in Figure 1 for one contact located within the periodic cell structure of one selected compound. In general, each crystal structure yielded a large number of duplicate observations, which were searched for and eliminated using a PERL script. The result of this preliminary analysis was a total of 240,310 unique sulfur–aromatic contacts.

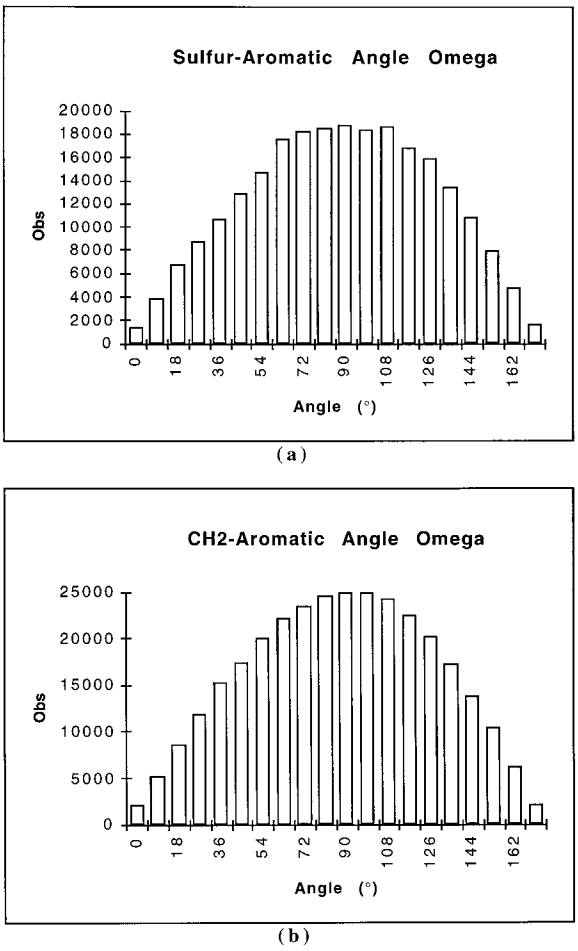


FIGURE 5 Histograms of the orientation angle ω distributions for (a) sulfur–aromatic and (b) CH_2–aromatic (control) interactions.

Using appropriate vector algebra (implemented in a custom C computer program), the angles defined in Figure 1 were used to compute the geometric descriptors defined in Figure 2. As illustrated in the diagram, for each sulfur–aromatic contact we define a polar angle θ of the ring centroid-to-sulfur axis, measured with respect to the ring normal; the angle ω between the bisector of the C–S–C angle and the centroid-to-sulfur axis; the torsional angle τ of the C–S–C bisector relative to the ring normal; and finally a propeller twist ϕ, defined as the smallest angle of rotation about the C–S–C bisector needed to make the C–S–C plane perpendicular to the plane of the aromatic ring. One-dimensional (1D) histograms were generated for the centroid–sulfur distance, and for all of the angles just defined, and two-dimensional (2D) histograms were generated for the various angles, with the centroid–sulfur distance taken as the horizontal axis in each case.

Statistical Analysis

In our analysis, we are interested in the spatial distribution of the C–S–C sulfur-bearing motif about the aromatic ring, and the distributions of the angular descriptors that measure the orientation of this group. We recognize that the forms of these distributions are determined in part by steric factors that would apply to *any* motif of similar geometry; for example, when the distance of any group to the ring centroid is small, then the polar angle must also be relatively small (else the group would overlap with ring atoms). Moreover, these constraints on the forms of the various distributions are uninformative as to the presence or absence of some special, favorable interaction between divalent sulfurs and aromatic rings.

To provide a source of "control" distributions, we have repeated exactly the same analyses as were just described, but using the motif C–CH_2–C in place of C–S–C. The control query selected 7942 crystal structures from the database, which yielded a total of 317,269 unique CH_2–aromatic contacts. We assume a priori that the C–CH_2–C group has no "special" interaction with the aromatic ring, and furthermore, its similarity to the C–S–C moiety allows us to compute all of the geometric descriptors described above, by simply substituting the central carbon in place of the sulfur. The 1D and 2D histograms were generated for

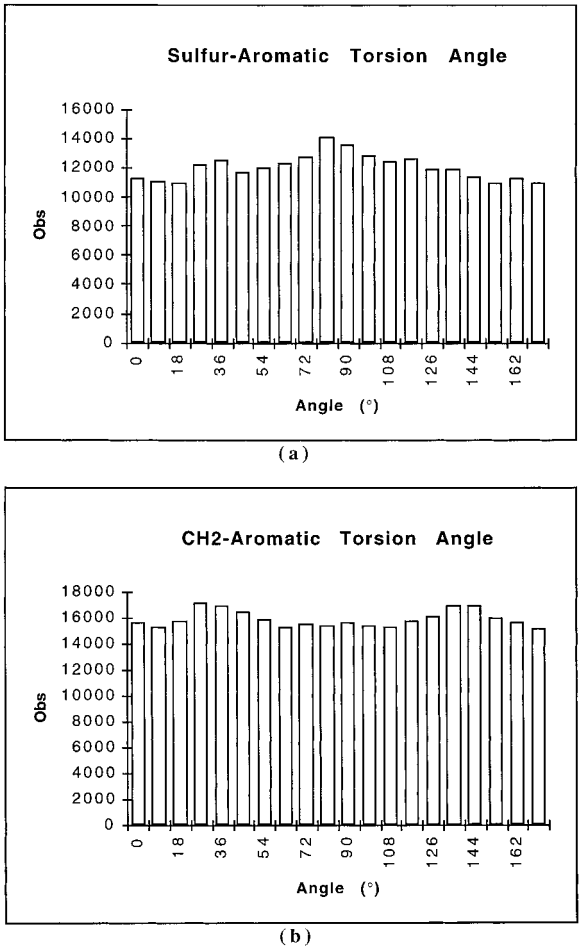
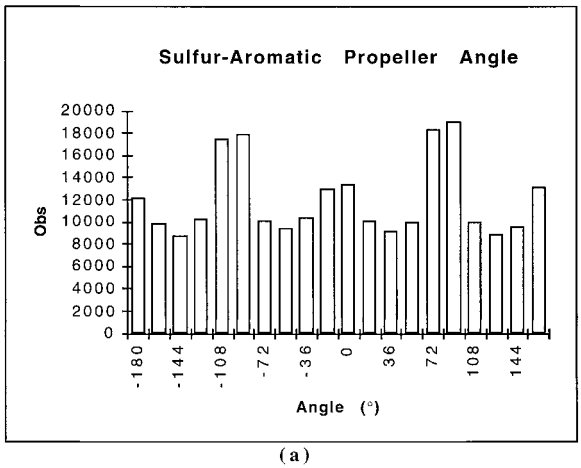


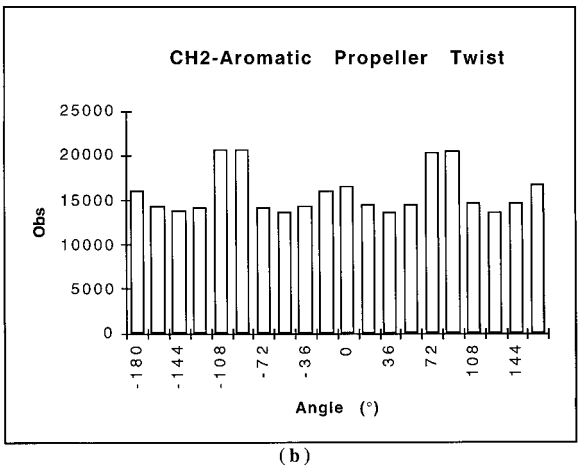
FIGURE 6 Histograms of the torsional angle (τ) distributions for (a) sulfur–aromatic and (b) CH_2–aromatic (control) interactions.

the observed C–CH_2–C distributions, in the same manner as for the sulfur-bearing motif.

We assessed the presence of a special sulfur–aromatic interaction by testing for statistically significant deviations between the distributions generated for the sulfur-bearing motif and the corresponding control distributions. To test for significant difference, we applied the Smirnov–Kolmogorov test,[14] a nonparametric test that requires only that the observations of the two distributions be merged (although still tagged by their distribution of origin), and then sorted in order using a ranking function selected by the investigator. The test directly compares the *cumulative distributions*, which are derived from the original distributions to be compared. (The cumulative distribution function specifies the number of observations for which the computed descriptor, e.g., sulfur–aromatic distance, is less than a particular value.) The statistic used in the two-sided Smirnov–Kolmogorov test employed here is the absolute value of the difference in the cumulative distributions along their common domains; the maximum value attained by the statistic is used to compute the level of confidence that the two source

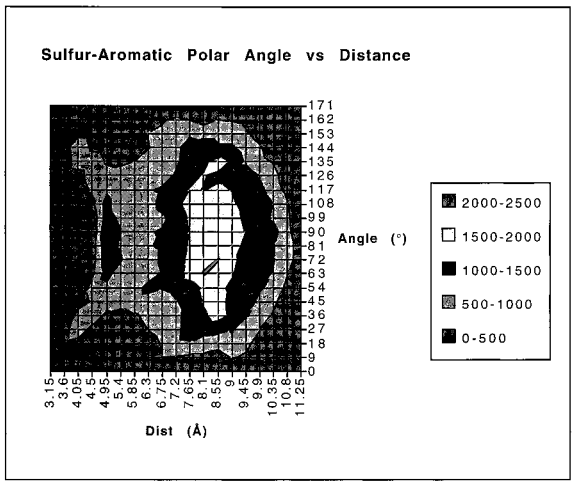


(a)

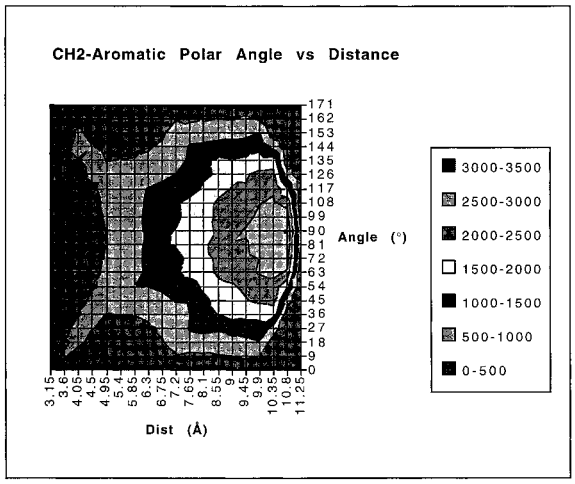


(b)

FIGURE 7 Histograms of the propeller twist angle (ϕ) distributions for (a) sulfur–aromatic and (b) CH_2–aromatic (control) interactions.



(a)



(b)

FIGURE 8 Two-dimensional histograms of the polar angle θ vs separation for (a) sulfur–aromatic and (b) CH_2–aromatic (control) interactions. Radial bin size is 0.45 Å, angular bin size is 9°. The amplitude of the distribution is shown by coded shading, as indicated in the legend.

distributions show a larger difference than would be explained by chance variation alone.

We computed the Smirnov–Kolmogorov statistic between the ring centroid to group (C–S–C or C–CH_2–C) distance distributions, and also between the various angle distributions (θ, ω, τ, ϕ) where we compared the sulfur–aromatic and CH_2–aromatic distributions for the same angular measure. In the case of the comparison between distance distributions, the observations were simply ranked by their radius from the ring center; in the case of the angular distributions, the observations for a particular angle Γ were ranked using the following function:

$$A_i = \text{floor}\left(\frac{R_i}{R_{\text{bin}}}\right) \cdot \Gamma_{\text{range}} + \Gamma_i \qquad (1)$$

Here R_i is the distance from the ring center for the ith observation, Γ_i is the angular value for the point, and

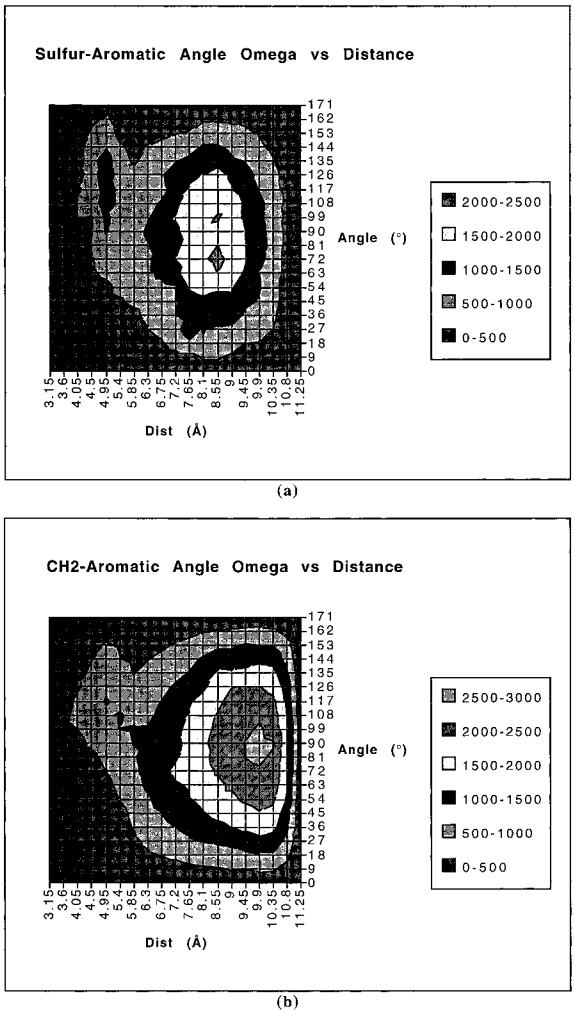


FIGURE 9 Two-dimensional histogram of the orientation angle ω vs separation for (a) sulfur–aromatic and (b) CH_2–aromatic (control) interactions.

Γ_{range} is the maximum range of values expected for a given angle (equal to 180° for θ, ω, and τ, and 360° for ϕ). R_{bin} is a radial "bin" size, and floor( ) denotes the function that returns the largest integer that is smaller than or equal to its real argument. The S statistics for the angular distributions was found to be relatively insensitive to the choice of the bin size; the results below correspond to $R_{bin} = 1.0$ Å.

Molecular Mechanics

In an effort to better understand the physical interactions that underlie the observed statistical distributions, we used the SYBYL Molecular Modelling Package, version 6.5 (Tripos, Inc., 1699 S. Hanley Road, St. Louis, MO 63144) to perform molecular mechanics minimizations and to analyze the energetics of unit cells generated for several entries extracted from the Cambridge database. Our protocol was to generate a complex of eight unit cells (occupying two unit-cell lengths along each of the three crystallographic axes), and to assign each molecule to one of two sets—those molecules containing any bond that "pierced" one of the bounding planes of the complex were assigned to the "exterior" set, the remainder to the "interior " set. Since SYBYL does not generate periodic boundary conditions for nonorthogonal unit cells (the case for all of the structures extracted from the database), we imposed range constraints on all atoms in the exterior set so that an energy penalty of 100 kcal/mol/Å^2 was applied for any atom that deviated from the original starting position by more than 1.5 Å; this was done so that the complexes could be minimized using the Tripos force field[15] included with SYBYL, while at the same time avoiding large deviations from the experimental structures. All energy calculations were performed with Gasteiger partial charges,[16] constant dielectric function, a dielectric constant of 1.0, and nonbonded cutoff of 12 Å.

After initial minimization with 1000 steps of the Powell minimizer, the unconstrained interior set was extracted from

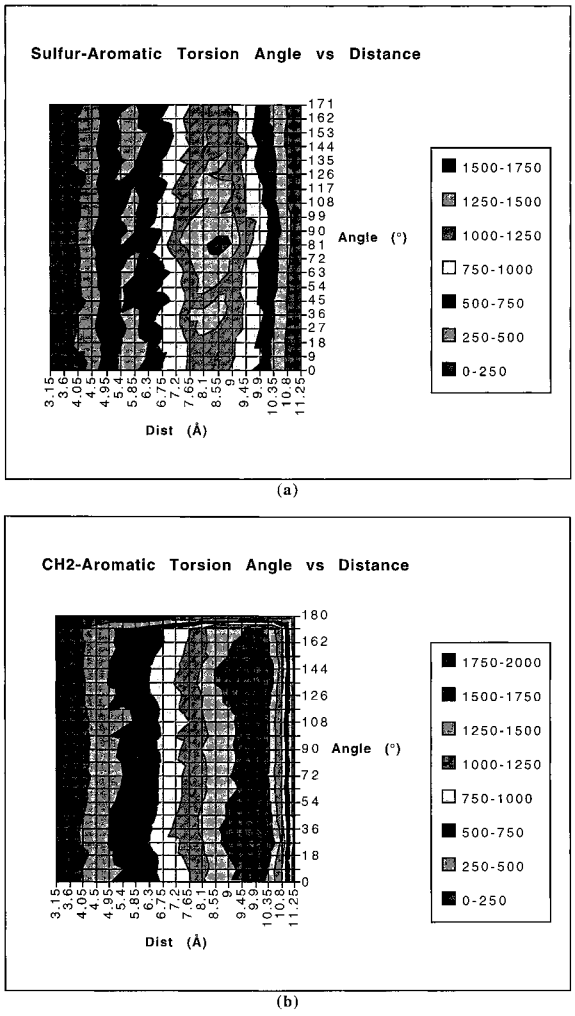


FIGURE 10 Two-dimensional histogram of the torsion angle τ vs separation for (a) sulfur–aromatic and (b) CH_2–aromatic (control) interactions.

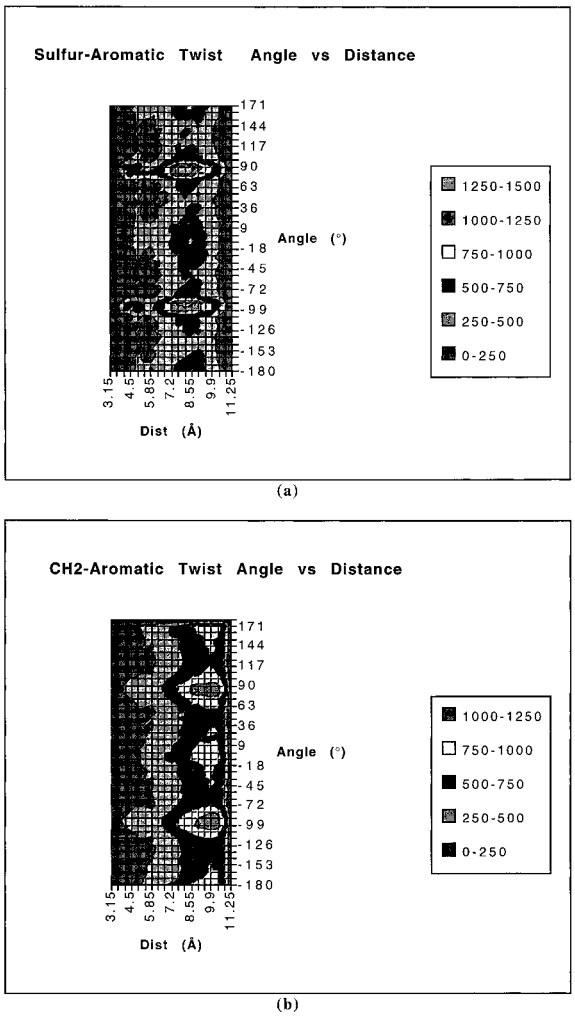


FIGURE 11 Two-dimensional histograms of the propeller twist angle ϕ vs separation for (a) sulfur–aromatic and (b) CH_2–aromatic (control) interactions.

each complex and used in computations of interaction energy. This was done by decomposing the interior atoms into three subsets: divalent sulfurs, atoms in six-membered aromatic rings, and the remainder (nonsulfur, nonsix-membered–aromatic). Energy of interaction was computed between all of the subsets for all of the complexes. In addition, pairs of molecules were selected from the interior of each complex, those chosen as representative of the most probable geometry of sulfur–aromatic interaction as determined from our statistical analyses, and were further utilized in an additional round of minimization and energy analysis.

RESULTS

Statistical Analysis

Figure 3 shows the histograms of ring centroid-to-group distance, and Figures 4–7 histograms of the four angles, for both the sulfur–aromatic and the CH_2–aromatic (control) data sets. Figures 8–11 present 2D histograms for the four angles (θ, ω, τ, ϕ) vs distance, for both the sulfur–aromatic and control data sets. Figure 12 shows plots of the *S* statistic, for the comparisons (sulfur–aromatic vs control) of the distance distributions and also the polar angle (θ) distributions. Table I summarizes the maximum observed value of the *S* statistic for all five pairwise comparisons of the sulfur–aromatic and control distributions.

The 1D distance histograms (Figure 3) for the sulfur–aromatic and the control data sets show a large maximum at relatively large separation. This is found at 8.4 Å in the sulfur–aromatic distribution, and at 9.9 Å in the control distribution. Such maxima are expected, given the sharp rise in observation volume with increasing radius, coupled with the 10 Å cutoff we imposed in the queries. However, there is an obvious local maximum in the distribution of aromatic–sulfur distances (a) at a ring centroid-to-sulfur separation of about 5.0 Å that is absent in the control distribution (b) (although the latter does appear to include an inflection point at about the same distance).

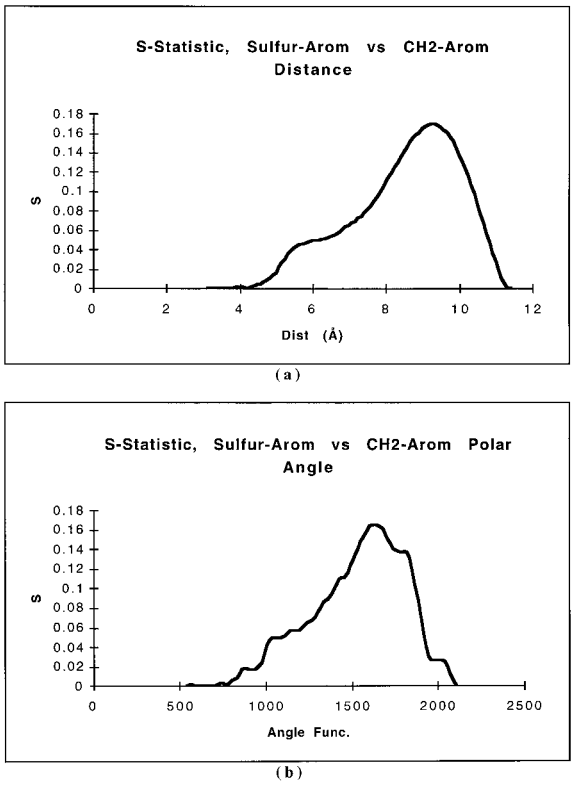


FIGURE 12 Plot of *S* statistic for comparison of sulfur–aromatic and CH_2–aromatic (control) distributions for (a) distance and (b) polar angle. The 0.0044 defines the 99% confidence level.

Table I Maximum Smirnov Statistic for each Sulfur–Aromatic/CH_2–Aromatic Distribution Pair (0.0044 = 99% Confidence Level)

Descriptor	Max *S*
Distance to ring	0.170025
Polar angle (θ)	0.168279
Orientation angle (ω)	0.169408
Torsional angle (τ)	0.169554
Propeller twist (ϕ)	0.168280

This visually apparent difference in the distributions is verified by the maximum value of the corresponding *S* statistic, 0.17 (Table I), which should be compared to the reference value of 0.0044 that defines the 99% confidence level. [It should also be pointed out that the *S* statistic "accumulates" difference as one moves along the sorted and merged distributions, and in Figure 12(a) has already reached a highly significant level at 5 Å separation.] Differences of comparable significance are found between the sulfur–aromatic and control angular distributions, although some of these are not as obvious from inspection of the plots. In every case, the difference between the sulfur–aromatic and control distributions significantly exceeds the 99% confidence level.

More detailed geometric information is found by examining the 2D histograms for the four angular descriptors (Figures 8–11), and comparing the sulfur–aromatic and the control distributions. In all of these histograms, we observe a 2D distribution that approximates the direct product of the 1D distance and the corresponding 1D angle histograms. In each of the control distributions there is one or more maxima at 9.9 Å, corresponding to the single maximum in the 1D distance histogram, and likewise the sulfur–aromatic plots all exhibit a large maximum at about 8.4 Å. But again, the 2D histograms for the sulfur–aromatic interactions all reveal additional maxima at approximately 5 Å separation that are absent in the control distributions.

The 2D histograms for polar angle vs separation (Figure 8) are similar in form for the sulfur–aromatic and control data sets, with a maximum at 90° angle and large separation, and with the most probable angle approaching 0° or 180° at small separation. The maximum at 90° is expected from solid-angle considerations, while the asymptotic behavior for small separations is easily seen to correspond to steric effects (at very small separation the interacting motif *must* be either directly above or below the ring). In addition, the sulfur–aromatic distributions shows a second broad maximum at 5.0 Å separation, covering the angular range of approximately 60°–115°. (This would correspond to an elevation of within 30° above or below the plane of the ring.) This second maximum is elongated in shape, covering a relatively large angular range, but only about 0.5 Å along the distance axis.

The sulfur–aromatic and control distributions for the angle ω (Figure 9) are again similar in form, each exhibiting a maximum at large separation and 90° angle, and with the most probable angle approaching 180° at small separation. (The value of 180° at small separation is easily understood in terms of sterics; at small separation, steric bulk associated with the sulfur must be directed away from the ring, leading to the observed angle.) Again, the sulfur–aromatic distribution shows a second maximum at 5.0 Å separation, in this case covering an angular range of approximately 90°–144°, and a distance of 0.5 Å. (There is some evidence of additional maxima in the control distribution at small separation, but these are small and poorly defined.)

The 1D distributions for the torsion angles are relatively uniform (Figure 6), and the 2D distributions (Figure 10) show little obvious similarity in form, with the sulfur–aromatic distribution exhibiting a number of shallow maxima at various separations that are not found in the control distribution. Both plots appear relatively uniform in the angular dimension at small separation. (For reasons discussed below, we expect the torsional angle to be uninformative as to the preferred geometry of the sulfur–aromatic interaction.)

Finally, the plots of the propeller angle ϕ (Figure 11) again are similar in form, both including maxima

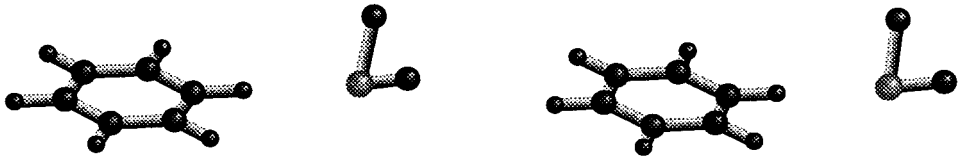


FIGURE 13 Stereo view of the "ideal" divalent sulfur–aromatic ring interaction geometry, as determined in this study.

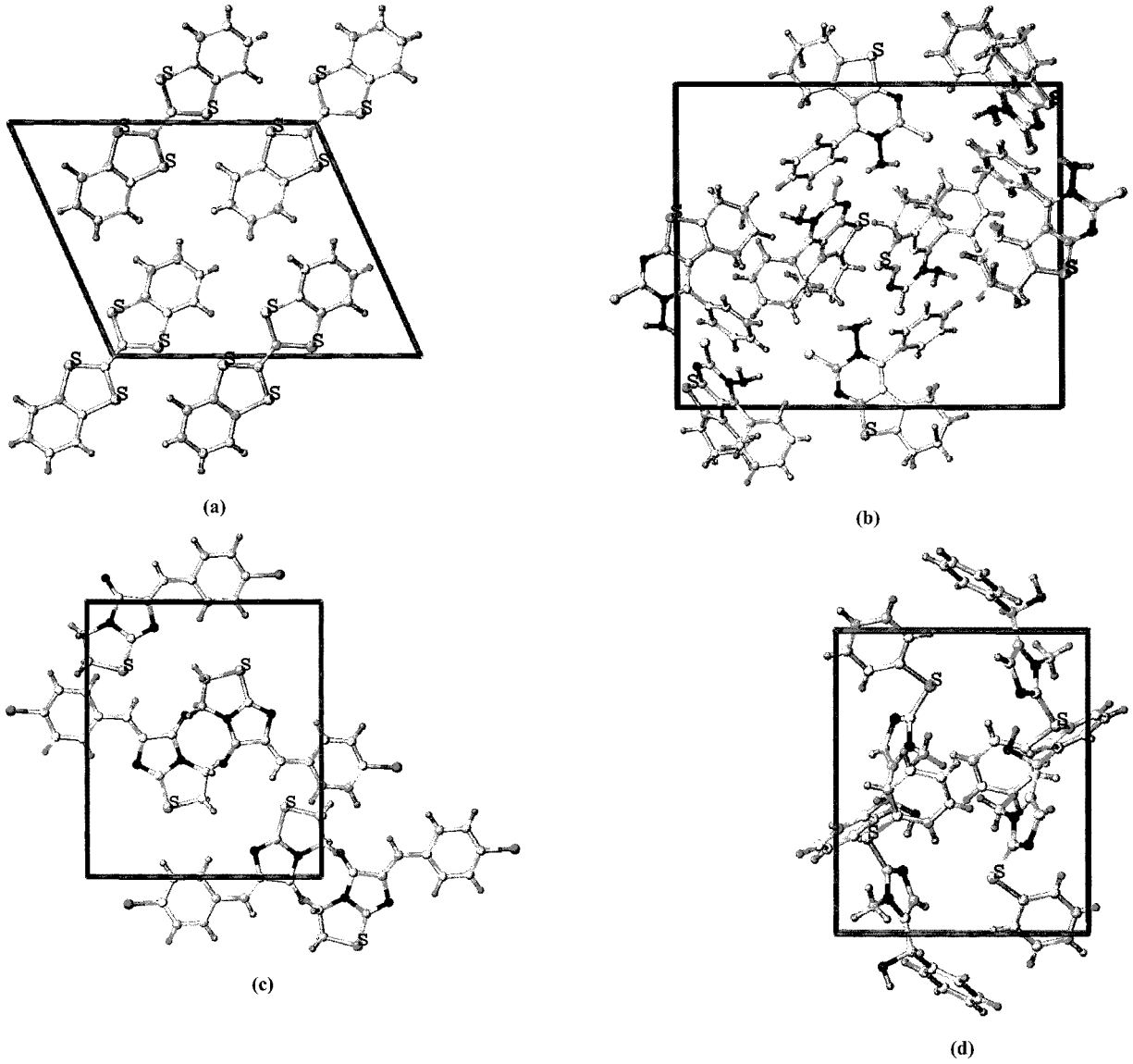


FIGURE 14 Four unit cells that include sulfur–aromatic interactions that approximate out "ideal" geometry: Cambridge entries BIRKIW,[17] GAPYOL,[18] SOJMUZ,[19] and VUSKID.[20] The unit cell is outlined.(It should be noted that none of the unit cells are orthogonal, although some appear so in these views.)

at −90° and 90° at large separation. Here the sulfur–aromatic distribution shows two additional maxima at a separation of about 5.0 Å, at −90° and 90° angles. These angles place the groups attached to the sulfur directly above and below the plane of the ring; it is likely that this orientation is favored sterically.

Given the positions of the various angular maxima, we can construct a three-dimensional (3D) model of the most probable position and orientation of the sulfur-bearing motif, when it is in close contact with an aromatic ring. To improve the visualization of the maxima for the angles θ and ω, we repeated Figures 8(a) and 9(a) as 3D surface plots (not shown), and located the local maxima for these distributions at 90° and 126°, respectively. The propeller twist angle was set to 90° and the torsion angle to 0°, leading to the structure shown in Figure 13. One piece of information that our analysis does *not* provide is the phase of the interaction of the motif and the ring (i.e., we cannot determine if the sulfur lies between ring hydrogens, or along the axis of a C_{ring}—H_{ring} bond). Observation of a number of specific examples from the database led to the structure in the diagram, where we have placed the sulfur atom in contact with one of

Table II Energy Components for Unit-Cell Complexes (All energies in kcal/mole)

	For Interior Set			Interaction					
Structure	No. Molecules	No. Atoms		Sulfur–Sulfur	Sulfur–Arom	Sulfur–NSNA[a]	Arom–Arom	Arom–NSNA	NSNA–NSNA
BIRKIW	4	104	Electrostatic	29.01	−29.55	−26.07	7.41	13.18	5.94
			van der Waals	−4.58	−13.35	−2.18	−11.68	−2.92	−0.43
			Total	24.43	−42.90	−28.25	−4.27	10.26	5.51
GAPYOL	28	1008	Electrostatic	41.59	−23.16	−57.13	3.16	14.23	8.75
			van der Waals	−2.79	−20.35	−55.86	−28.03	−136.05	−187.74
			Total	38.80	−43.51	−112.98	−24.87	−121.81	−175.98
SOJMUZ	8	208	Electrostatic	6.70	0.74	−14.91	−0.61	0.15	5.51
			van der Waals	−0.18	−8.82	−3.74	−14.67	−45.65	−5.20
			Total	6.52	−8.08	−18.65	−15.28	−45.51	0.31
VUSKID	12	444	Electrostatic	5.39	−10.15	−4.98	2.08	2.05	−5.50
			van der Waals	−0.21	−10.31	−5.10	−66.26	−67.54	−11.12
			Total	5.18	−20.46	−10.08	−64.18	−65.49	−16.62

[a] NSNA: nonsulfur, nonaromatic.

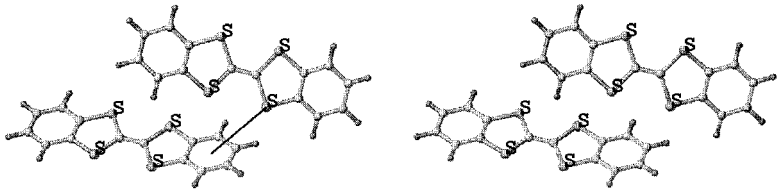
the ring hydrogens. (We intend to address better the question of the relative orientation of the ring in future work.)

Molecular Mechanics

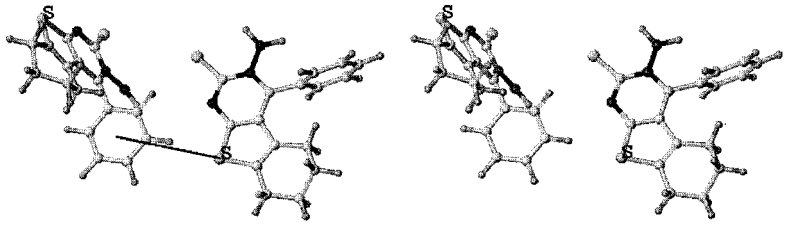
In Figure 14 we display unit cells for four members of the sulfur–aromatic data set, the entries with database IDs BIRKIW,[17] GAPYOL,[18] SOJMUZ,[19] and VUSKID.[20] Each of these contains at least one sulfur–aromatic interaction that is a close approximation to the ideal geometry derived from this study (Figure 13). Energy minimizations and computation of interaction energy components were carried out as described above for the $2 \times 2 \times 2$ unit-cell complexes, and the results are summarized in Table II. Here we show interaction energies between all the three defined atom subsets (sulfur, aromatic, and nonsulfur, nonaromatic), and break down the energies into their electrostatic and van der Waals components. It should be stressed that all energy values represent interactions *between* molecules in the unit cell; all intramolecular interactions (and all bonded energies) were omitted.

It can seen that the sulfur–aromatic interaction is always stabilizing and of significant size, but of course the relative contribution of this term to the total energy of the cell depends upon the chemical structure of the molecules and the relative proportion of sulfur atoms and phenyl rings. It is most important in BIRKIW, where the sulfur–aromatic interaction (−42.90 kcal/mole) is actually larger in magnitude than the total energy of the unit-cell complex (−35.22 kcal/mole), and least important in GAPYOL where it is 10% of the size of the total energy of the unit-cell complex. One of the referees has suggested that the sulfur–aromatic interaction we have identified may in fact be a form of hydrogen bond[21] between the divalent sulfur (acting as acceptor) and ring carbon (acting as donor). While this is an interesting and useful hypothesis, the computational data in Table II suggest the situation may be more complex. Where we would expect a hydrogen bond to be largely electrostatic in character, only for the case of BIRKIW is the sulfur–aromatic interaction energy mostly electrostatic; in the other structures analyzed the van der Waals contribution either predominates (SOJMUZ) or is of nearly equal magnitude to the electrostatic energy (GAPYOL, VUSKID). On the other hand, while the sulfur–aromatic energies tend to include a large van der Waals contribution, in most of the structures considered here the electrostatic component is still of significant size. This would indicate a directionality greater than expected in a purely van der Waals interaction, and would suggest that the sulfur–aromatic interaction may in fact have significant hydrogen bond character.

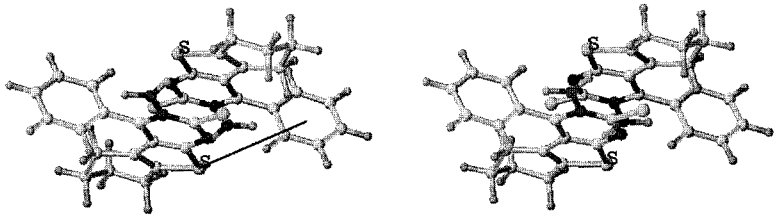
To further refine our analysis, we extracted interacting pairs of molecules from the unit-cell complexes that were representative of our ideal geometry (Figure 13), and computed the total interaction energy as well as the sulfur–aromatic contribution (including van der Waals and electrostatic components). The extracted pairs are shown in Figure 15, and the energy analysis is presented in Table III. In the case of GAPYOL, two orientations were extracted [Figure 15(b,c)], the first



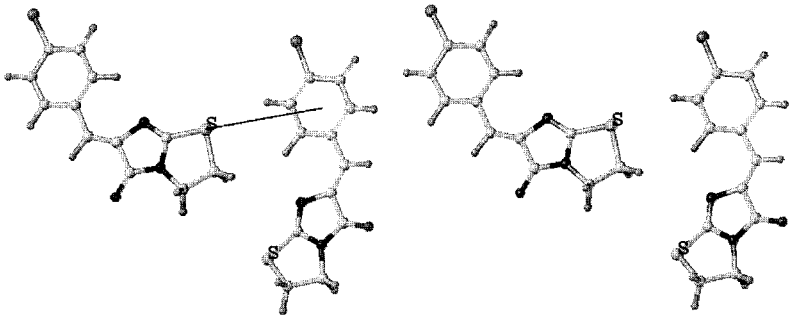
(a)



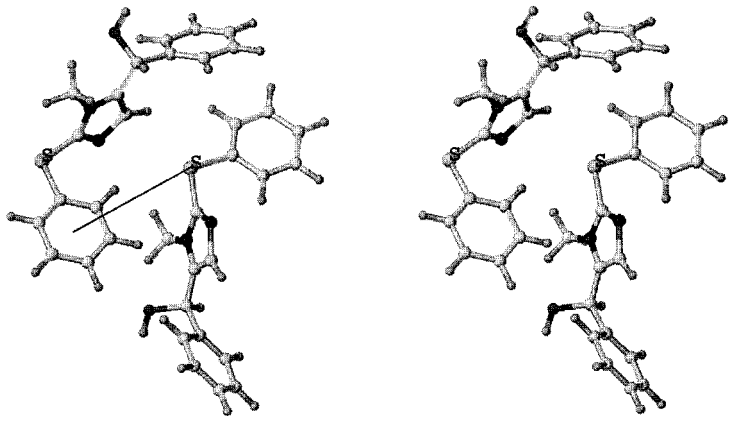
(b)



(c)



(d)



(e)

Table III Total and Sulfur–Aromatic Interaction Energies for Molecule Pairs, Before and After Minimization

	Structure	Pair Interaction Total Energy	Sulfur–Arom Total Energy	Sulfur–Arom Electrostatic	Sulfur–Arom van der Waals
Before pair minimization	BIRKIW[a]	−4.41	−7.47	−5.26	−2.21
	GAPYOL (1)	−2.11	−0.44	−0.22	−0.21
	GAPYOL (2)	−10.48	−0.77	−0.29	−0.48
	SOJMUZ	−1.99	−0.57	−0.17	−0.4
	VUSKID	−8.51	−1.07	−0.24	−0.83
After pair minimization	BIRKIW[a]	−19.48	−10.97	−5.93	−5.04
	GAPYOL (1)	−11.68	−0.92	0.01	−0.94
	GAPYOL (2)	−12.15	−0.9	−0.25	−0.65
	SOJMUZ	−4.04	−0.81	−0.22	−0.59
	VUSKID	−11.67	−1.48	−0.19	−1.29

[a] The BIRKIW pair includes two sulfur–aromatic interactions.

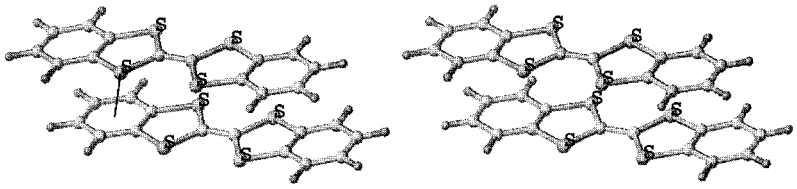
corresponding closely to our ideal geometry, while in the second the orientation of the divalent sulfur with respect to the ring plane is nearly perpendicular to that of the ideal configuration. It is found that the energy for each sulfur–aromatic interaction ranges from −3.7 kcal/mole (BIRKIW, corresponding to a total energy of −7.47 in two equivalent interactions) up to −0.44 kcal/mole for pair 1 of GAPYOL. This is consistent with prior experimental and computational studies, which placed the size of the interaction at −1 to −2 kcal/mole. Finally, we allowed each of the extracted pairs to energy minimize (in isolation from the unit-cell complexes) under the Tripos force field. The interaction energies were recomputed, and are reported in Table III. The minimized structures are shown in Figure 16. In most of the complexes (the two exceptions are GAPYOL, pair 2, and SOJMUZ) the divalent sulfur moves to a position above the aromatic ring, bringing the geometry in better accord with the results of ab initio calculations and the statistics derived from protein crystal structures. It can be seen that the relative proportion of electrostatic and van der Waals contributions to the total sulfur–aromatic interaction depends on the structure considered, both before and after energy minimization; however, in every case the proportion of the energy in the van der Waals term increases relative to the electrostatic upon energy minimization.

DISCUSSION

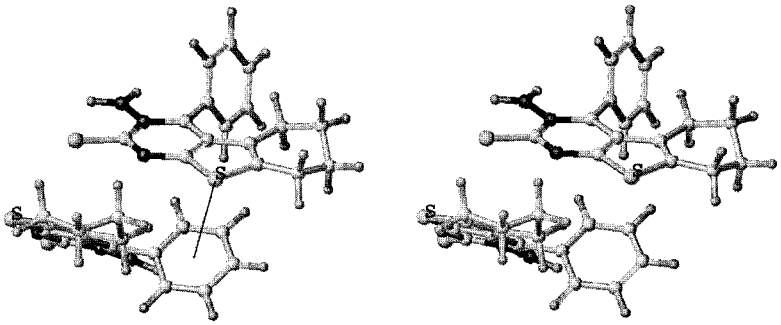
We have determined an "ideal" geometry for the sulfur–aromatic interaction by analyzing a large sample of such interactions present in the Cambridge Crystallographic Database, and determining a preferred interaction distance, as well as values for four geometric descriptors. Three of these (the polar angle θ, orientation angle ω, and propeller twist ϕ) were used to define the ideal geometry. The distribution of the remaining descriptor, the torsion τ, did not exhibit an obvious maxima near the preferred sulfur–aromatic distance of 5 Å. Given our final preferred geometry, and the definition of the torsion presented in Figure 2, we note that a reflection of the motif through the plane of the ring would change the torsion angle from 0° to 180°, and yet would present the same physical interaction. This suggests that the torsion is simply not a useful descriptor, given the preferred geometry we finally discovered.

The ideal geometry we have determined differs significantly from that suggested by the previous analyses of protein crystallographic structures, and by quantum chemical studies. Our ideal structure places the sulfur in the plane of the ring, with the lone pairs interacting closely with a ring hydrogen. In contrast, sulfur–aromatic interactions in proteins typically involve the sulfur at a significant angle of elevation

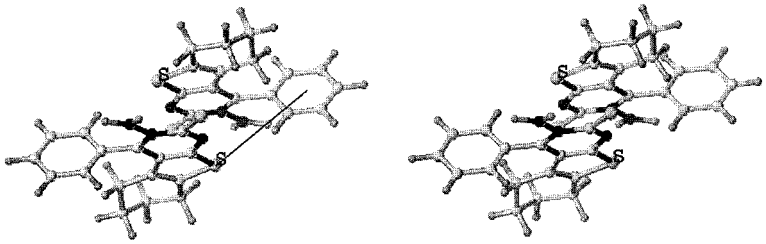
FIGURE 15 Stereo views of extracted molecule pairs that include close sulfur–aromatic contacts. (A) BIRKIW; (b) GAPYOL, "ideal" geometry; (c) GAPYOL, alternative interaction geometry; (d) SOJMUZ; (e) VUSKID. These structures are all taken from the interior set of the corresponding 2 × 2 × 2 unit-cell complexes, as described in the text. The solid line indicates a sulfur–aromatic contact.



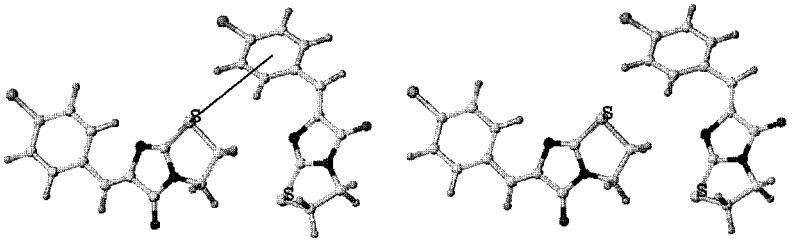
(a)



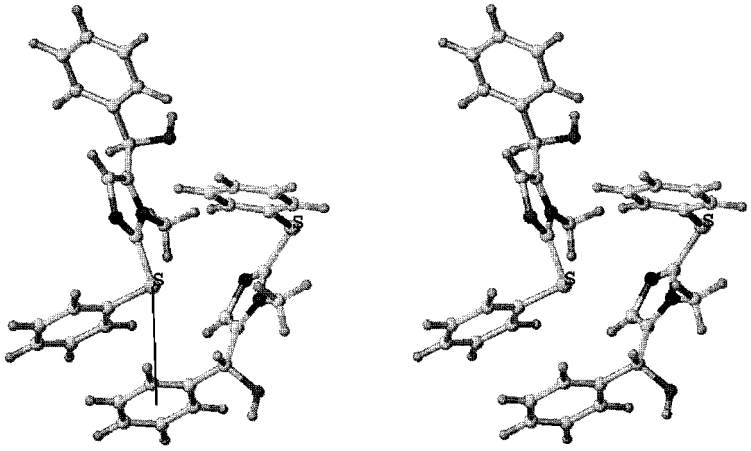
(b)



(c)



(d)



(e)

above the plane of the ring. In the case of quantum mechanical studies of small sulfur-bearing compounds interacting with an aromatic ring, it has been observed that the most stable configurations place the sulfur at a large elevation above the ring plane (in agreement with protein data), but with the sulfur lone pairs directed *away* from the ring centroid, an orientation impossible for interacting amino acids.

The results of quantum-mechanical studies are not directly comparable to the analysis presented here or to studies involving protein structures. The sulfur-bearing compounds that have been used in quantum calculations lack significant steric bulk, and there are few constraints on the geometry of interaction with the ring. In contrast, the compounds present in the Cambridge Database typically include significant steric bulk in the vicinity of each divalent sulfur (for the compounds shown in Figures 14–16, the sulfur is part of a ring). As a consequence, it is impossible for a sulfur to closely contact a ring in such a way that the lone pairs are directed away from the ring centroid.

Our energy analysis of unit-cell complexes and of isolated interacting molecules supports the results of prior experimental and computational work, which indicates that a single sulfur–aromatic interaction can stabilize a complex by 1–2 kcal/mole, and in the structures considered here the sulfur–aromatic energy can be a significant fraction of the total interaction energy. Our computational evidence suggests that the sulfur–aromatic interaction is intermediate between a purely van der Waals interaction and a hydrogen bond, and is characterized by a significant electrostatic component. However, the relative size of the electrostatic and van der Waals energies clearly depends upon the structure considered, lending a certain ambiguity to this interaction.

We are still left with the question as to why the preferred geometry for sulfur–aromatic contact derived from this study places the sulfur in the plane of the ring, in contradiction to the observed preference in proteins. For small planar molecules, stacking interactions may lead to a crystal structure in which the compounds are all nearly parallel to each other, and this arrangement would clearly tend to favor edge-on interactions. In fact, this arrangement is observed in the unit cell for the compound BIRKIW shown in Figure 15(a). At the same time, the compound VUSKID shown in Figure 15(e), which is more flexible and does not exhibit this packing pattern, yields essentially the same sulfur–aromatic contact geometry. An alternative explanation is that the geometry observed in the small-molecule crystal structures is energetically preferred, but that steric constraints of the folded polypeptide chain disfavor this configuration. Our energy minimizations of isolated pairs of molecules lead in most cases to a significant shift in geometry, which places the sulfur above the plane of the ring, and this is in accord with the ab initio and protein database results. While that would seem to favor the hypothesis that crystal packing forces contribute significantly in determining the ideal geometry we have discovered in this study, we would prefer not to rush to a hasty conclusion; the significance of electrostatics in the sulfur–aromatic interaction means that the charge model chosen will have a great impact on the most favorable geometry, and here we have only employed a simple semiempirical method available in SYBYL. At the same time, dispersion interactions are difficult to compute accurately with molecular orbital methods, and it is possible that previous ab initio studies have not adequately estimated the magnitude and geometric dependence of this other critical component of the sulfur–aromatic interaction, especially for compounds with greater steric bulk. At this point we consider the relative utility of molecular mechanics and ab initio methods in accurately reproducing these energy contributions to be an unresolved issue.

We plan further work aimed at addressing these questions, specifically (1) a new and detailed survey of the sulfur–aromatic interactions present in the protein structures of the Brookhaven Database, with analysis of the geometry of the sulfur orientation, as carried out here; (2) refined molecular mechanics studies that include more accurate charge models and the contribution of electrostatic polarization; and (3) high-level quantum chemical studies of interactions between aromatic rings and sulfur-bearing compounds that include significant steric bulk, in order to determine the effect on the preferred geometry and energy of the interaction.

CONCLUSION

An accumulation of experimental and computational evidence points to the possible existence of a special and significant interaction between divalent sulfurs

FIGURE 16 Stereo views of extracted pairs after unconstrained energy minimization. Starting points for the minimizations were the corresponding geometries extracted from the unit-cell complexes (FIGURE 15). (a) BIRKUW; (b) GAPYOL (ideal); (c) GAPYOL (alternate); (d) SOJMUZ; (e) VUSKID. The solid line indicates a sulfur–aromatic contact.

and aromatic rings. We have presented new and compelling evidence to support this hypothesis, derived from crystal structures for small organic compounds available in the Cambridge Crystallographic Database. This class of interaction, which is frequently observed in protein structures, may be important for protein stability and for directing the folding of polypeptide chains. Additional work is needed to fully ascertain the energetics and intrinsic preferred geometry of this interaction.

Our thanks to Dr. Robert Clark (Tripos, Inc.) for suggesting the statistical test used in this work, and for many helpful discussions.

REFERENCES

1. Morgan, R. S.; Tatsch, C. E.; Gushard, R. H.; McAdon, J. M.; Warme, P. K. Int J Peptide Protein Res 1978, 11, 209–217.
2. Morgan, R. S.; McAdon, J. M. Int J Peptide Protein Res 1980, 15, 177–180.
3. Reid, K. S. C.; Lindley, P. F.; Thornton, J. M. FEBS Lett 1985, 190, 209–213.
4. Bodner, B. L.; Jackman, L. M.; Morgan, R. S. Biochem Biophys Res Commun 1980, 94, 807–813.
5. Nemethy, G.; Scheraga, H. A Biochem Biophys Res Commun 1981, 98, 482–487.
6. Cheney, B. V.; Schulz, M. W.; Cheney, J. Biochim Biophys Acta 1989, 996, 116–124.
7. Lebel, M.; Sugg, E. E.; Hruby, V. J. Int J Peptide Protein Res 1987, 29, 40–45.
8. Viguero, A. R.; Serrano, L. Biochemistry 1995, 34, 8771–8779.
9. Munoz, V.; Serrano, L. Proteins Struct Funct Genet 1994, 20, 301–311.
10. Spencer, D. S.; Sites, W. E. J Mol Biol 1996, 257, 497–499.
11. Yamaotsu, N.; Moriguchi, I.; Hirono, S. Biochim Biophys Acta 1993, 1203, 243–250.
12. Pranata, J. Bioorg Chem 1997, 25, 213–219.
13. Allen, F. H.; Kennard, O. Chem Design Auto News 1993, 8, 1 and 31–37.
14. Daniel, W. W. Applied Nonparametric Statistics; Houghton-Mifflin: Boston, MA, 1978; pp276–286.
15. Clark, M.; Cramer, R. D., III; van Opdenbosch, N. J Comp Chem 1989, 10, 982–1012.
16. Gasteiger, J.; Marsili, M. Tetrahedron 1980, 36, 3219–3228.
17. Shibaeva, R. P.; Lobkovskaya, R. M.; Klyuevo, V. N. Cryst Struct Commun 1982, 189, 835.
18. Florencio, F.; Martinez-Carrera, S.; Garcia-Blanco, S. Z Kristallogr 1987, 179, 395.
19. Karolak-Wojciechowska, J.; Kiec-Kononowicz, K. Acta Crystallogr Sect C (Cr Str Comm) 1991, 47, 2371.
20. Ohta, S.; Yamamoto, T.; Kawasaki, I.; Yamashita, M.; Katsuma, H.; Nasaka, R.; Kobayashi, K.; Ogawa, K. Chem Pharm Bull 1992, 4, 2681.
21. Gregoret, L. M.; Rader, S. D.; Fletterick, R. J.; Cohen, F. E. Proteins 1991, 9, 99–107.