

The Geometry of the N–H···O=C Hydrogen Bond. 3.* Hydrogen-Bond Distances and Angles

BY ROBIN TAYLOR, OLGA KENNARD AND WERNER VERSICHEL

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

(Received 27 October 1983; accepted 18 January 1984)

Abstract

A survey of 1509 N–H···O=C hydrogen bonds, observed by X-ray or neutron diffraction in 889 organic crystal structures, gave the following results. The distribution of hydrogen-bond distances shows small but significant deviations from normality. The H···O and N···O distances of intramolecular bonds tend to be longer and shorter, respectively, than those of intermolecular bonds. The H···O distances of intermolecular bonds are very sensitive to changes in the nature and environment of the donor and acceptor groups. For example, substituted ammonium ions tend to form shorter hydrogen bonds than unsubstituted ammonium ions, and amides are stronger acceptors than unionized carboxylic acids and esters. The distribution of N–H···O angles is consistent with an overall energetic preference for a linear or near-linear N–H···O arrangement. There is a highly significant tendency for short hydrogen bonds to be more linear than long bonds. Statistically significant correlations were found between the H···O hydrogen-bond distance and the donor (N–H) and acceptor (C=O) bond lengths, but may be partly due to librational effects. The C=O bond length is also correlated with the number of hydrogen bonds accepted by the C=O group.

Introduction

This is the final paper in a short series describing the results of a statistical analysis of 1509 crystallographically independent N–H···O=C hydrogen bonds. In this paper we examine the distributions of: (a) the hydrogen-bond distances, $r(\text{H}\cdots\text{O})$ and $r(\text{N}\cdots\text{O})$; (b) the hydrogen-bond angle, $\alpha(\text{N}-\text{H}\cdots\text{O})$; (c) the donor bond length, $r(\text{N}-\text{H})$; (d) the acceptor bond length, $r(\text{C}=\text{O})$ (Fig. 1).

Methodology

All crystallographic data were retrieved from the Cambridge Structural Database (1982 release; Allen *et al.*, 1979). The analysis was based on 1509 crystallographically independent bonds, taken from 889 organic crystal structures; a full list of references was

deposited previously (Taylor, Kennard & Versichel, 1983). Some 1426 of the hydrogen bonds were observed by X-ray diffraction, the remainder by neutron diffraction. The X-ray hydrogen-bond geometries were 'normalized' (Jeffrey & Lewis, 1978), *i.e.* the H-atom position was moved along the observed N–H bond direction until the N–H distance was equal to a 'standard' value (1.030 Å). This procedure corrects for systematic errors in the X-ray results (Taylor & Kennard, 1983*a*). Since the hydrogen bond is very sensitive to its crystallographic environment, unweighted means were used throughout the analysis (Taylor & Kennard, 1983*b*).

$r(\text{H}\cdots\text{O})$ and $r(\text{N}\cdots\text{O})$

Preliminary observations

Fig. 2(a) shows the distribution of $r(\text{H}\cdots\text{O})$ for the complete sample of 1509 hydrogen bonds. The mean and standard deviation of the distribution are 1.921 (4) and 0.143 Å, respectively (Table 1). Kolmogorov-Smirnov and χ^2 goodness-of-fit tests (Siegel, 1956) show that the distribution is significantly non-normal (>99% confidence level). This is illustrated by the probability plot in Fig. 2(b). The plot was constructed by selecting a series of fifteen distances (x_i , $i = 1, 2, \dots, 15$); for each distance, the observed number of bonds with $r(\text{H}\cdots\text{O}) < x_i$ was plotted against the number expected in a normal distribution with $\mu = 1.921$ Å and $\sigma = 0.143$ Å. The systematic deviations of the plotted points from the ideal straight line of unit slope are obvious. The coefficients of skewness and kurtosis of the $r(\text{H}\cdots\text{O})$ distribution are 0.80 (6) and 3.49 (13), respectively. Comparison with the values expected for a normal distribution, *i.e.* 0 and 3, respectively (Snedecor & Cochran, 1980), indicates that the $r(\text{H}\cdots\text{O})$ distances are positively skewed and exhibit positive kurtosis. The former observation implies that low values of

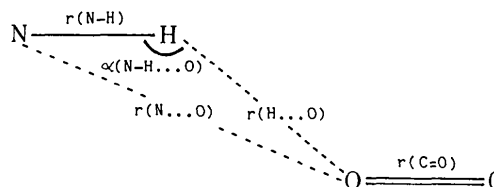


Fig. 1. Geometrical parameters used in analysis.

* Parts 1 and 2: Taylor, Kennard & Versichel (1983, 1984).

$r(\text{H}\cdots\text{O})$ are grouped closer to the mean than high values. Presumably, this reflects a potential-energy-distance curve of the general shape shown in Fig. 2(c). The positive kurtosis indicates that a normal distribution with $\mu = 1.921 \text{ \AA}$, $\sigma = 0.143 \text{ \AA}$ has a flatter peak than the observed $r(\text{H}\cdots\text{O})$ distribution. Although statistically significant, the deviations from normality of the $r(\text{H}\cdots\text{O})$ distribution are relatively moderate (Hamilton, 1964) and are only detectable because of the large sample size.

The $r(\text{N}\cdots\text{O})$ distribution (Fig. 2d, e) is rather similar to that of $r(\text{H}\cdots\text{O})$. Relevant statistics are given in Table 1.

Comparison of intermolecular and intramolecular bonds

Fig. 3(a, b) shows the distributions of $r(\text{H}\cdots\text{O})$ and $r(\text{N}\cdots\text{O})$ for the 152 intramolecular bonds in our

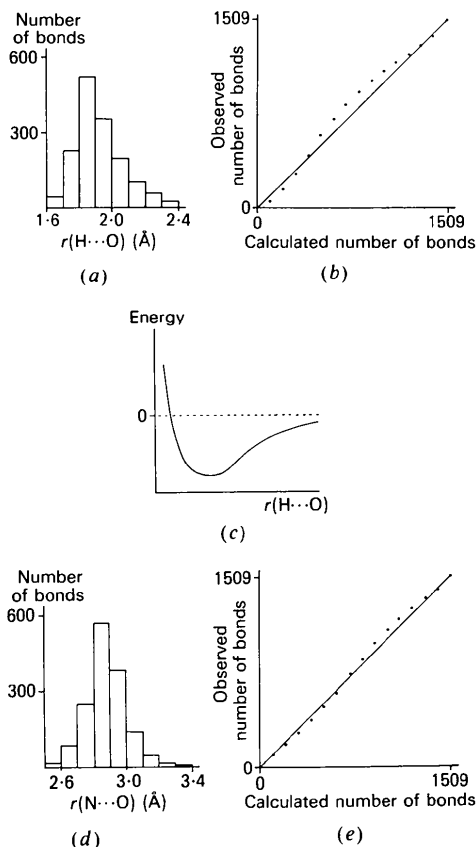


Fig. 2. (a) Distribution of $r(\text{H}\cdots\text{O})$ for complete sample of 1509 bonds. (b) Probability plot: for fifteen distances, x_i ($i = 1, 2, \dots, 15$), the observed number of bonds with $r(\text{H}\cdots\text{O}) < x_i$ is plotted against the number expected in a normal distribution with $\mu = 1.921 \text{ \AA}$, $\sigma = 0.143 \text{ \AA}$. (c) Presumed variation of hydrogen-bond energy with $r(\text{H}\cdots\text{O})$. (d) Distribution of $r(\text{N}\cdots\text{O})$ for complete sample of 1509 bonds. (e) Probability plot: for fifteen distances, x_i ($i = 1, 2, \dots, 15$), the observed number of bonds with $r(\text{N}\cdots\text{O}) < x_i$ is plotted against the number expected in a normal distribution with $\mu = 2.878 \text{ \AA}$, $\sigma = 0.122 \text{ \AA}$.

Table 1. Statistics of $r(\text{H}\cdots\text{O})$ and $r(\text{N}\cdots\text{O})$ distributions (distances in \AA)

Distribution	N^*	μ	σ	$\sqrt{b_1}$	b_2
$r(\text{H}\cdots\text{O})$, all	1509	1.921 (4)	0.143	0.80 (6)	3.49 (13)
$r(\text{H}\cdots\text{O})$, intermolecular	1357	1.913 (4)	0.138	0.84 (7)	3.70 (13)
$r(\text{H}\cdots\text{O})$, intramolecular	152	1.988 (13)	0.165	†	†
$r(\text{N}\cdots\text{O})$, all	1509	2.878 (3)	0.122	0.42 (6)	4.03 (13)
$r(\text{N}\cdots\text{O})$, intermolecular	1357	2.892 (3)	0.112	0.84 (7)	4.39 (13)
$r(\text{N}\cdots\text{O})$, intramolecular	152	2.755 (12)	0.143	†	†

* N = number of observations in sample, μ = sample mean, σ = sample standard deviation, $\sqrt{b_1}$ = sample coefficient of skewness, b_2 = sample coefficient of kurtosis.

† Not calculated.

sample. Fig. 3(c, d) shows the corresponding distributions for the 1357 intermolecular bonds. Means, standard deviations and coefficients of skewness and kurtosis are summarized in Table 1. Mann-Whitney and variance-ratio tests (Snedecor & Cochran, 1980) established that the following observations are statistically (>99%) significant: (a) the $r(\text{H}\cdots\text{O})$ distances of intramolecular bonds tend to be longer than those of intermolecular bonds; (b) the $r(\text{N}\cdots\text{O})$ distances of intramolecular bonds tend to be shorter than those of intermolecular bonds; (c) the variances of the intramolecular $r(\text{H}\cdots\text{O})$ and $r(\text{N}\cdots\text{O})$ distributions are larger than those of the corresponding intermolecular distributions. Observations (a) and (b) can only be reconciled if the $\alpha(\text{N}-\text{H}\cdots\text{O})$ angles of intramolecular bonds are, on average, smaller than those of intermolecular bonds. This might be due to the prevalence of the intramolecular hydrogen-

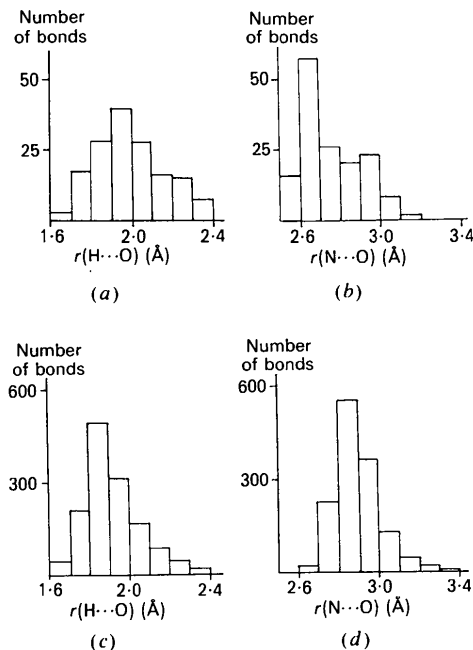
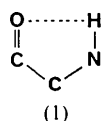


Fig. 3. Distributions of (a) $r(\text{H}\cdots\text{O})$ and (b) $r(\text{N}\cdots\text{O})$ for the 152 intramolecular bonds. Distributions of (c) $r(\text{H}\cdots\text{O})$ and (d) $r(\text{N}\cdots\text{O})$ for the 1357 intermolecular bonds.

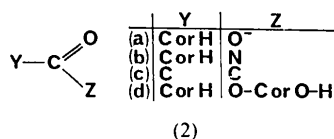
bonding arrangement (1), in which $\alpha(\text{N-H}\cdots\text{O})$ is forced to deviate considerably from linearity.



Observation (c) suggests that the geometries of intramolecular bonds are often distorted by internal geometrical constraints. Intramolecular bonds were therefore excluded from the remainder of our study of $r(\text{H}\cdots\text{O})$.

Dependence of $r(\text{H}\cdots\text{O})$ on nature of donor and acceptor

The proton-donor and -acceptor groups involved in the 1357 intermolecular bonds were divided into various categories. Donor groups were categorized as: (a) uncharged donors, $\geq\text{N-H}$; (b) charged trigonal donors, $\geq\text{N}^+-\text{H}$; (c) unsubstituted ammonium ions, NH_4^+ ; (d) monosubstituted ammonium ions, RNH_3^+ ; (e) disubstituted ammonium ions, R_2NH_2^+ ; (f) trisubstituted ammonium ions, R_3NH^+ . (The formula ' $\text{R}_2\text{:H}_2^+$ ' indicates that the N atom is bonded to two C and two H atoms, and is not meant to imply that the two substituents are chemically identical; similarly for ' R_3NH^+ '. The symbol 'R' does not necessarily indicate a simple alkyl or aryl substituent, as the dataset includes a wide variety of structural types.) Acceptor groups were categorized as: (a) carboxylate anions (2a; henceforth termed 'carboxylates'); (b) amides (2b); (c) ketones (2c); (d) unionized carboxylic acids and esters (2d; henceforth termed 'carboxyls'). Five of the 1357 intermolecular bonds were found to involve aldehyde acceptors. The $r(\text{H}\cdots\text{O})$ distances of these bonds (1.786, 1.823, 1.839, 2.053, 1.893 Å) seem to be somewhat shorter than those of bonds involving ketone acceptors. We therefore decided not to group aldehydes and ketones together, and the five bonds were excluded from this part of the study.



The mean values of $r(\text{H}\cdots\text{O})$ for the various donor-acceptor combinations are given in Table 2, together with the number of observations in each category. The marginal figures are row- and column-means, *unweighted* by the number of observations in each 'cell'. Despite the simplicity of the donor-acceptor classification scheme, some overall trends are apparent. The mean values of $r(\text{H}\cdots\text{O})$ tend to increase along the approximate series: carboxylate < amide < ketone < carboxyl. Thus, carboxylates seem

Table 2. Dependence of $r(\text{H}\cdots\text{O})$ on nature of donor and acceptor (distances in Å)

Donor	Acceptor				Row mean*	
	Carboxyl	Ketone	Amide	Carboxylate		
$\geq\text{N-H}$	<i>N</i> †	117	38	597	74	1.959
	μ	2.002 (12)	1.970 (22)	1.934 (5)	1.928 (19)	
	σ	0.131	0.133	0.123	0.163	
$\geq\text{N}^+-\text{H}$	<i>N</i>	11	2	12	36	1.888
	μ	1.983 (55)	1.844 (126)	1.858 (43)	1.869 (28)	
	σ	0.183	0.179	0.149	0.170	
NH_4^+	<i>N</i>	13	2	4	56	1.946
	μ	1.916 (41)	1.995 (110)	1.988 (75)	1.886 (18)	
	σ	0.149	0.155	0.151	0.138	
RNH_3^+	<i>N</i>	68	8	15	226	1.885
	μ	1.936 (14)	1.872 (60)	1.891 (34)	1.841 (8)	
	σ	0.119	0.169	0.132	0.114	
R_2NH_2^+	<i>N</i>	6	3	3	47	1.860
	μ	1.887 (47)	1.966 (178)	1.793 (70)	1.796 (14)	
	σ	0.114	0.308	0.120	0.096	
R_3NH^+	<i>N</i>	0	1	2	11	1.835
	μ	—	1.938 (—)	1.845 (14)	1.722 (25)	
	σ	—	—	0.020	0.084	
Column mean*	1.945	1.931	1.885	1.840		

* See text.

† *N* = number of observations in cell, μ = mean of observations in cell, σ = standard deviation of observations in cell.

to form the strongest hydrogen bonds, and ketones and carboxyls the weakest. Presumably, this is due to subtle variations in the electron density distributions around the acceptor groups. The magnitude of the partial negative charge at the O atom will be particularly important (Umeyama & Morokuma, 1977). Our results are consistent with those of Berkovitch-Yellin, Ariel & Leiserowitz (1983), who inferred that amides are better acceptors than unionized carboxylic acids from a study of the crystal-packing arrangements of *N*-acylated amino acids.

The mean $r(\text{H}\cdots\text{O})$ distances also tend to increase along the approximate series: $\text{R}_3\text{NH}^+ < \text{R}_2\text{NH}_2^+ < \text{RNH}_3^+ < \geq\text{N}^+-\text{H} < \text{NH}_4^+ < \geq\text{N-H}$. Presumably, this order parallels the variation in partial positive charge at the H atom. Many of the $\geq\text{N}^+-\text{H}$ donors will be aromatic molecules in which the formal positive charge at the N atom is extensively delocalized. This probably accounts for the comparative weakness of $\geq\text{N}^+-\text{H}$ as a proton donor. Differences between the crystallographic environments of substituted and unsubstituted ammonium ions may be relevant. The R_3NH^+ ion has only one 'active' proton and therefore forms only one hydrogen bond. In contrast, the NH_4^+ ion invariably donates all four protons in hydrogen bonds. Steric interactions between the four acceptor species surrounding NH_4^+ may therefore be expected to lengthen $\text{H}_3\text{N}^+-\text{H}\cdots\text{O}$ bonds relative to $\text{R}_3\text{N}^+-\text{H}\cdots\text{O}$ bonds. Theoretical electron density calculations (Kollman, 1977; Schuster, 1976) suggest that there may be an additional electrostatic effect. When the NH_4^+ ion forms its first hydrogen bond, the electron density is polarized so as to increase the partial

Table 3. $r(\text{H}\cdots\text{O})$ analysis of variance (distances in Å)(a) Mean $r(\text{H}\cdots\text{O})$ distances

Donor		Acceptor		Row mean*
		Amide/ketone/ carboxyl	Carboxylate	
$\geq\text{N}-\text{H}$	N^\dagger	752	74	1.9449
	μ	1.947 (5)	1.928 (19)	
	σ	0.127	0.163	
$\geq\text{N}^+-\text{H}$	N	25	36	1.8863
	μ	1.912 (34)	1.869 (28)	
	σ	0.172	0.170	
NH_4^+	N	19	56	1.8996
	μ	1.940 (33)	1.886 (18)	
	σ	0.146	0.138	
RNH_3^+	N	91	226	1.8642
	μ	1.923 (13)	1.841 (8)	
	σ	0.127	0.114	
R_2NH_2^+	N	12	47	1.8137
	μ	1.883 (50)	1.796 (14)	
	σ	0.173	0.096	
R_3NH^+	N	3	11	1.7551
	μ	1.876 (32)	1.722 (25)	
	σ	0.056	0.084	
Column mean*		1.9420	1.8553	1.9131‡

(b) Analysis of variance

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Nature of donor/acceptor	11	3.2427 Å ²	0.2948 Å ²	17.6
Residual (within cells)	1340	22.4376	0.0167	
Total	1351	25.6803	0.0190	

* See text.

† N = number of observations in cell, μ = mean of observations in cell, σ = standard deviation of observations in cell.

‡ Grand mean.

positive charge at the donated H atom and decrease the charges at the other three H atoms. The ability of the latter to participate in hydrogen bonds is consequently weakened.

A formal statistical analysis of the results in Table 2 is difficult because of the small number of observations in some of the categories. We therefore combined the results for amides, ketones, and carboxyls, producing Table 3(a). The marginal figures in this table are row- and column-means, *weighted* by the number of observations in each cell. Table 3(a) shows that the shortening of $r(\text{H}\cdots\text{O})$ due to increasing the partial negative charge at O (*i.e.* changing from an amide, ketone or carboxyl acceptor to a carboxylate acceptor) decreases along the approximate series: $\text{R}_3\text{NH}^+ > \text{R}_2\text{NH}_2^+ > \text{RNH}_3^+ > \text{NH}_4^+ > \geq\text{N}^+-\text{H} > \geq\text{N}-\text{H}$. This is consistent with a simple electrostatic view of hydrogen bonding as being largely due to the Coulombic attraction between H and O.

An analysis of variance (Wetherill, 1981) was performed on the results in Table 3(a). The total variance of the intermolecular $r(\text{H}\cdots\text{O})$ distances was assumed to be given by:

$$\sigma^2(\text{total}) = \sigma^2(\text{model}) + \sigma^2(\text{residual}), \quad (1)$$

where $\sigma^2(\text{model})$ is the variance accounted for by the donor-acceptor classification scheme used in Table 3(a), and $\sigma^2(\text{residual})$ is the remaining variance. The

$r(\text{H}\cdots\text{O})$ distance of the k th hydrogen bond in category ij (*i.e.* row i , column j of Table 3a) was then assumed to be given by:

$$r(\text{H}\cdots\text{O})_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \quad (2)$$

where μ_{ij} is the mean $r(\text{H}\cdots\text{O})$ distance of the hydrogen bonds in category ij , and ε_{ijk} is a Gaussian random variable with mean = 0, variance = $\sigma^2(\text{residual})$. The statistical treatment is based on two approximations. Firstly, it is assumed that $r(\text{H}\cdots\text{O})$ is normally distributed. As shown above, this is not exactly true but is a reasonably good approximation. Secondly, it is assumed that the variance of ε_{ijk} is constant for all ij categories. The within-cell standard deviations in Table 3(a) suggest that this is a satisfactory approximation. However, it is unlikely to be exactly true. For example, different types of donors vary in their ability to form three-centre (*i.e.* bifurcated) bonds (Taylor, Kennard & Versichel, 1984) and this may well produce systematic variations in $\sigma^2(\varepsilon_{ijk})$.

The results of the analysis of variance are summarized in Table 3(b). Insofar as the assumptions underlying the statistical technique are justified, the F value of 17.6 is statistically significant at the >99% level. This suggests that our simple donor-acceptor classification scheme accounts for a significant part of the variance of the $r(\text{H}\cdots\text{O})$ distances, *i.e.* a dependence of $r(\text{H}\cdots\text{O})$ on the nature of the donor and acceptor is indicated. The residual mean square (0.0167 Å²; see Table 3b) is an unbiased estimate of $\sigma^2(\text{residual})$ and the total mean square (0.0190 Å²) can be taken as an approximate estimate of $\sigma^2(\text{total})$. The quantity $\sigma^2(\text{model})$ can therefore be estimated from (1), giving: $\sigma^2(\text{total}) = 0.138$ Å²; $\sigma^2(\text{model}) = 0.048$ Å²; $\sigma^2(\text{residual}) = 0.129$ Å². This treatment is only approximate but serves to illustrate that $\sigma(\text{residual}) \gg \sigma(\text{model})$, even though $\sigma(\text{model})$ is significantly different from zero.

Other factors affecting $r(\text{H}\cdots\text{O})$

The $\sigma^2(\text{residual})$ term in (1) may be further partitioned as:

$$\sigma^2(\text{residual}) = \sigma^2(\text{exptl}) + \sigma^2(\text{physical}), \quad (3)$$

where $\sigma^2(\text{exptl})$ is the variance due to experimental errors in the observed $r(\text{H}\cdots\text{O})$ distances, and $\sigma^2(\text{physical})$ is the variance due to physical effects other than those accounted for by our simple donor-acceptor classification scheme. An example would be crystal packing forces. A recent study suggests that the average value of the random experimental errors in X-ray $r(\text{H}\cdots\text{O})$ distances is about 0.065 Å (Taylor & Kennard, 1983a). Taking this as an approximate estimate of $\sigma(\text{exptl})$, we get:

$$\begin{aligned} \sigma^2(\text{physical}) &= \sigma^2(\text{total}) - \sigma^2(\text{model}) - \sigma^2(\text{exptl}) \\ &\approx 0.138^2 - 0.048^2 - 0.065^2 = 0.112^2. \end{aligned} \quad (4)$$

Table 4. Mean $r(\text{H}\cdots\text{O})$ distances of bonds involving 'single' and 'multiple' acceptors (distances in Å)

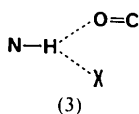
Donor	Type of bond Acceptor	'Single' acceptor*		'Multiple' acceptor*	
		N^\dagger	μ	N	μ
$\geq\text{N}-\text{H}$	Carboxyl	93	2.004 (14)	20	1.980 (28)
$\geq\text{N}-\text{H}$	Amide	417	1.923 (6)	155	1.962 (11)
$\geq\text{N}-\text{H}$	Carboxylate	16	1.877 (36)	46	1.960 (25)
RNH_3^+	Carboxyl	41	1.932 (20)	17	1.956 (28)
RNH_3^+	Carboxylate	38	1.774 (14)	165	1.859 (9)
R_2NH_2^+	Carboxylate	28	1.765 (13)	17	1.828 (26)

* Defined in text.

† N = number of observations in sample, μ = sample mean.

Thus, *most* of the variation in the $r(\text{H}\cdots\text{O})$ distances is due to crystal packing forces and subtle changes in donor and acceptor properties not accounted for by our simple classification scheme.

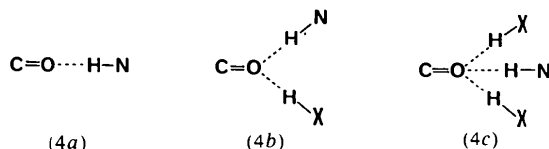
One factor that is known to influence $r(\text{H}\cdots\text{O})$ is the presence of a second hydrogen-bond acceptor atom (X in 3): it was shown (Taylor, Kennard & Versichel, 1984) that the $r(\text{H}\cdots\text{O})$ distances of three-centre (*i.e.* bifurcated) bonds are significantly longer than those of two-centre (*i.e.* linear) bonds. It is also known that charged donors (N^+-H) are more likely to form three-centre bonds than uncharged donors ($\text{N}-\text{H}$) (Taylor, Kennard & Versichel, 1984). Thus, the non-bonded repulsions between the donor and acceptor groups in (3), which tend to increase $r(\text{H}\cdots\text{O})$, appear to outweigh the inherent tendency of charged donors to form shorter hydrogen bonds than uncharged donors.



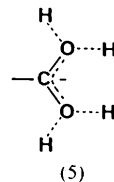
The crystallographic environment of the acceptor $\text{C}=\text{O}$ group is also relevant. This was shown by comparing the mean $r(\text{H}\cdots\text{O})$ distance of bonds involving 'single acceptors' (4a) with the corresponding value for bonds involving 'multiple acceptors' [4b, or, more rarely, 4c, where $X = \text{N}, \text{O}, \text{S}$ or halogen, $r(\text{H}\cdots\text{O}) < 2.4 \text{ \AA}$ and $\alpha(\text{C}=\text{O}\cdots\text{H}) \geq 90^\circ$].* A separate comparison was made for each of the donor-acceptor categories in Table 2, but only six categories

* Our dataset included 98 intermolecular bonds which caused difficulty at this stage of the analysis. In each of these bonds, the $\text{C}=\text{O}$ group was found to form a short contact to an electronegative atom X [$X = \text{O}, \text{N}, \text{S}$ or halogen: $\alpha(\text{C}=\text{O}\cdots\text{X}) \geq 90^\circ$; $r(\text{O}\cdots\text{X}) < \text{sum of O and X van der Waals radii}$] but not to any H atom bonded to X . This could represent: (a) a short $\text{O}\cdots\text{X}$ non-bonded interaction; (b) a $\text{C}=\text{O}\cdots\text{H}-\text{X}$ hydrogen bond. Possibility (b) exists because some H-atom coordinates are missing from the Cambridge Structural Database; these usually correspond to H atoms that were not located in X-ray analyses. A simple and safe policy was to exclude the 98 bonds from any part of the statistical analysis that focused on the number of hydrogen bonds accepted by the $\text{C}=\text{O}$ groups. This was done, both here and in part 1 of the series (Taylor, Kennard & Versichel, 1983).

contained sufficient data for the comparison to be meaningful. Results for these six are summarized in Table 4.



The mean $r(\text{H}\cdots\text{O})$ distance of bonds involving multiple acceptors is longer than that of bonds involving single acceptors for all but one of the donor-acceptor categories studied. In four cases ($\geq\text{N}-\text{H}\cdots\text{amide}$, $\geq\text{N}-\text{H}\cdots\text{carboxylate}$, $\text{RNH}_3^+\cdots\text{carboxylate}$, $\text{R}_2\text{NH}_2^+\cdots\text{carboxylate}$) the difference between the single- and multiple-acceptor mean values is statistically significant at or above the 90% level (Mann-Whitney tests). Steric repulsions between the two (or three) donor groups in (4b) (or 4c) are probably responsible for these observations. The difference between the single- and multiple-acceptor mean values is particularly large for hydrogen bonds involving carboxylate acceptors. Steric repulsions are likely to be especially important for these bonds because: (a) *both* O atoms in the carboxylate group normally accept two or more hydrogen bonds, (5); (b) the $r(\text{H}\cdots\text{O})$ distances of bonds involving carboxylates tend to be shorter than those of bonds involving amides, ketones and carboxyls (see above).



Theoretical calculations (Kollman, 1977; Schuster, 1976) suggest that the formation of the $\text{C}=\text{O}\cdots\text{H}-\text{X}$ bond in (4b) will polarize the acceptor molecule so as to increase the electron density at O. The $\text{C}=\text{O}$ group therefore becomes a better acceptor towards the $\text{N}-\text{H}$ group. This electronic effect would be expected to oppose the steric effect mentioned above. Table 4 suggests that the steric factors predominate, except, possibly, for bonds involving the weakest type of acceptor (*i.e.* carboxyls).

$\alpha(\text{N}-\text{H}\cdots\text{O})$

Preliminary observations

Fig. 4(a) shows the distribution of $\alpha(\text{N}-\text{H}\cdots\text{O})$ for the complete sample of 1509 hydrogen bonds. The mean and standard deviation are $158.3 (4)$ and 15.6° , respectively. The distribution is affected by a

geometrical factor, in that the number of possible configurations with $\varphi - \delta\varphi < \alpha(\text{N-H}\cdots\text{O}) < \varphi + \delta\varphi$ is proportional to $\sin \varphi$ (Fig. 4b) (Kroon, Kanters, van Duijneveldt-van Rijdt, van Duijneveldt & Vliegthart, 1975). Correction for this factor was achieved by multiplying each bar of the histogram by $N/\sin \bar{\varphi}$, where $\bar{\varphi}$ is the average of the upper and lower limits of the bar and N is a normalization constant. The corrected histogram (Fig. 4c) has a seminormal standard deviation of about 16° . The shape of Fig. 4(c) is obviously consistent with an overall energetic preference for the linear N-H \cdots O arrangement. However, it has been observed that small deviations from hydrogen-bond linearity may be energetically favourable in some circumstances (Umeyama & Morokuma, 1977; Kroon *et al.*, 1975).

Comparison of intermolecular and intramolecular bonds

Fig. 5(a, b) shows the distribution of $\alpha(\text{N-H}\cdots\text{O})$ for the 152 intramolecular bonds and the 1357 intermolecular bonds, respectively. The mean value of $\alpha(\text{N-H}\cdots\text{O})$ is appreciably smaller for the intramolecular bonds [$132.5 (15)^\circ$] than for the intermolecular bonds [$161.2 (3)^\circ$], as predicted from our study of the $r(\text{H}\cdots\text{O})$ and $r(\text{N}\cdots\text{O})$ distributions. A Mann-Whitney test showed that the difference between the means is highly ($>99.9\%$) significant. Intramolecular bonds were excluded from the remainder of our study of $\alpha(\text{N-H}\cdots\text{O})$.

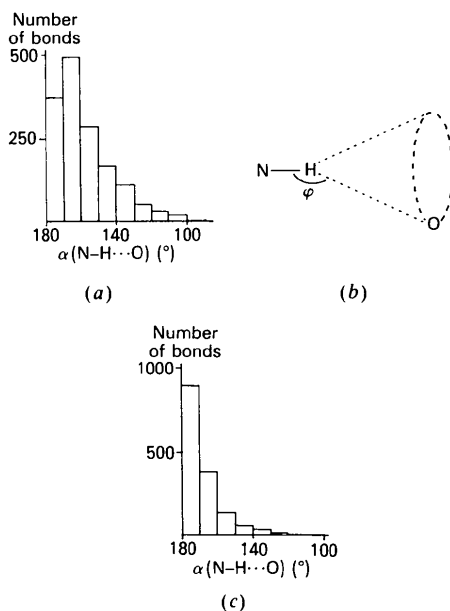


Fig. 4. (a) Distribution of $\alpha(\text{N-H}\cdots\text{O})$ for complete sample of 1509 bonds. (b) Geometrical factor affecting $\alpha(\text{N-H}\cdots\text{O})$ distribution. (c) Distribution of $\alpha(\text{N-H}\cdots\text{O})$, corrected for geometrical factor shown in (b).

Correlation of $\alpha(\text{N-H}\cdots\text{O})$ and $r(\text{H}\cdots\text{O})$

The $\alpha(\text{N-H}\cdots\text{O})$ angles and $r(\text{H}\cdots\text{O})$ distances of the 1357 intermolecular bonds have a Spearman rank correlation coefficient of -0.521 . This is significantly different from zero at the $>99.9\%$ confidence level. Thus, $\alpha(\text{N-H}\cdots\text{O})$ is inversely correlated with $r(\text{H}\cdots\text{O})$ (Olovsson & Jönsson, 1976). Fig. 6 shows the correlation graphically; the first (*i.e.* left-most) point represents the mean $\alpha(\text{N-H}\cdots\text{O})$ angle of intermolecular bonds with $1.6 < r(\text{H}\cdots\text{O}) < 1.7 \text{ \AA}$, the next point represents the mean angle for bonds with $1.7 < r(\text{H}\cdots\text{O}) < 1.8 \text{ \AA}$, and so on. The steady decrease of the mean $\alpha(\text{N-H}\cdots\text{O})$ angles with increasing $r(\text{H}\cdots\text{O})$ is apparent. The correlation may be ascribed to: (a) the unfavourable N \cdots O repulsion in short, non-linear N-H \cdots O bonds; (b) the loss of charge-transfer energy as $\alpha(\text{N-H}\cdots\text{O})$ is decreased – this may be more serious in strong hydrogen bonds than in weak ones. However, *ab initio* energy decomposition studies (Umeyama & Morokuma, 1977) show that charge transfer, exchange repulsion and electrostatic interaction energies all vary when a hydrogen bond is distur-

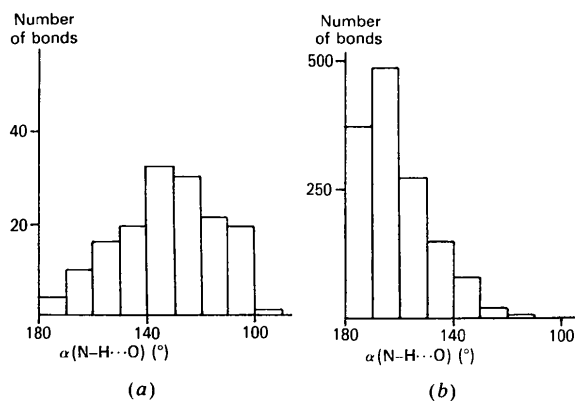


Fig. 5. Distribution of $\alpha(\text{N-H}\cdots\text{O})$ for (a) the 152 intramolecular bonds, and (b) the 1357 intermolecular bonds.

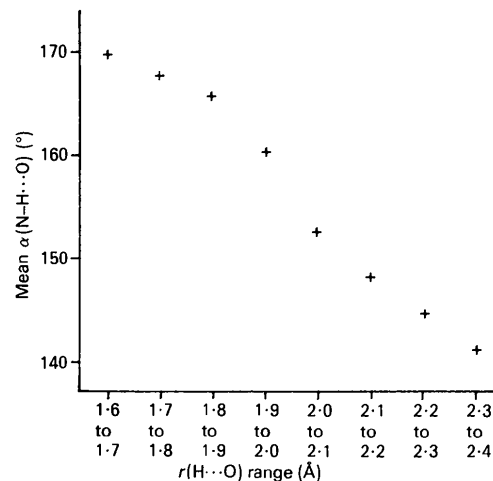


Fig. 6. Plot of mean $\alpha(\text{N-H}\cdots\text{O})$ against $r(\text{H}\cdots\text{O})$ range.

ted from linearity. Thus, the $\alpha(\text{N-H}\cdots\text{O})-r(\text{H}\cdots\text{O})$ correlation may be due to a complicated superposition of many effects.

Other factors affecting $\alpha(\text{N-H}\cdots\text{O})$

The sensitivity of $\alpha(\text{N-H}\cdots\text{O})$ to small changes in the chemical environment is so pronounced that it tends to obscure systematic correlations between $\alpha(\text{N-H}\cdots\text{O})$ and the nature of the donor and/or acceptor. For example, we divided a subset of chemically similar hydrogen bonds (intermolecular, two-centre bonds of the type $\geq\text{N-H}\cdots\text{amide}$) into two groups, one consisting of bonds involving 'single acceptors', the other consisting of bonds involving 'multiple acceptors' (defined as above). The mean $\alpha(\text{N-H}\cdots\text{O})$ angles of the two groups were *not* significantly different [single acceptors: $164.1(6)^\circ$; multiple acceptors: $163.5(10)^\circ$].

One factor that we examined particularly closely was the variation of $\alpha(\text{N-H}\cdots\text{O})$ with donor type. It has been suggested that this may be of some importance (Olovsson & Jönsson, 1976). The results of our study are summarized in Table 5. At least two factors are likely to be relevant. Firstly, the tendency for donor groups to form three- and four-centre bonds (*i.e.* 'bifurcated' and 'trifurcated' bonds) decreases in the approximate order: $\text{R}_3\text{NH}^+ > \text{R}_2\text{NH}_2^+ > \text{RNH}_3^+ > \text{NH}_4^+ > \geq\text{N}^+-\text{H} = \geq\text{N}-\text{H}$. Since three- and four-centre bonds are usually less linear than two-centre bonds (Taylor, Kennard & Versichel, 1984), this should produce a steady decrease in mean $\alpha(\text{N-H}\cdots\text{O})$ along the series: $\geq\text{N}-\text{H} = \geq\text{N}^+-\text{H} > \text{NH}_4^+ > \text{RNH}_3^+ > \text{R}_2\text{NH}_2^+ > \text{R}_3\text{NH}^+$. However, the variation of $r(\text{H}\cdots\text{O})$ with donor type (Tables 2, 3a), together with the inverse correlation between $\alpha(\text{N-H}\cdots\text{O})$ and $r(\text{H}\cdots\text{O})$, will favour approximately the opposite order. The first column of Table 5 gives results for the complete sample of 1357 intermolecular bonds. Differences between the means are rather small compared with their standard errors, but there seems to be a slight tendency for $\alpha(\text{N-H}\cdots\text{O})$ to decrease down the column. This implies that the first of the above factors is the more important. Results in the second column of Table 5 are for intermolecular, *two-centre* bonds only. They do not show any obvious trends, which suggests that physical effects other than those discussed above are important.

$r(\text{N-H})$

Our study of $r(\text{N-H})$ distances was necessarily confined to the 83 bonds determined by neutron diffraction (since all X-ray data were normalized – see above). Fifteen of these bonds were determined at low temperature. All results given here are based on the complete sample of 83 bonds; results based on the 68 room-temperature bonds were not significantly

Table 5. Dependence of $\alpha(\text{N-H}\cdots\text{O})$ on nature of donor (angles in deg)

Donor	All intermolecular bonds		Two-centre intermolecular bonds	
	<i>N</i> *	μ	<i>N</i>	μ
$\geq\text{N}-\text{H}$	830	162.1 (4)	731	163.4 (4)
$\geq\text{N}^+-\text{H}$	61	159.9 (18)	54	162.6 (17)
NH_4^+	75	160.9 (17)	59	166.1 (14)
RNH_3^+	317	160.0 (7)	229	163.6 (7)
R_2NH_2^+	59	157.5 (14)	33	162.1 (16)
R_3NH^+	15	158.1 (27)	6	164.0 (31)

* *N* = number of observations in sample, μ = sample mean.

different. From a statistical point of view, the justification for excluding X-ray data from our study of $r(\text{N-H})$, but including the data in all other parts of the analysis, is as follows. Parameters that do not involve H-atom positions [*e.g.* $r(\text{N}\cdots\text{O})$, $r(\text{C}=\text{O})$] are determined with comparable precision by X-ray and neutron diffraction. Thus, results from both methods should be used. Parameters that do involve H-atom positions are determined much more precisely by neutron diffraction than by X-ray diffraction. In the case of $r(\text{N-H})$, this means that the variance of a sample of X-ray data is likely to be much larger than that of a sample of neutron data; thus, X-ray data should not be used. However, parameters such as $r(\text{H}\cdots\text{O})$, $\alpha(\text{N-H}\cdots\text{O})$, *etc.* are so sensitive to changes in the chemical environment that samples of X-ray and neutron data are likely to have similar variances, despite the difference in experimental precisions (Taylor & Kennard, 1983*a, b*). Thus, both X-ray and neutron results should be used for these parameters.

The observed distribution of $r(\text{N-H})$ is shown in Fig. 7(a); the mean and standard deviation are 1.030 (2) and 0.016 Å respectively. Fig. 7(b) shows the scatterplot of $r(\text{N-H})$ against $r(\text{H}\cdots\text{O})$. These parameters have a Spearman rank correlation coefficient of -0.720 , which is significantly different from zero at the $>99.9\%$ confidence level. Thus, $r(\text{N-H})$ is inversely correlated with $r(\text{H}\cdots\text{O})$ (Olovsson & Jönsson, 1976; Koetzle & Lehmann, 1976). Two factors may be relevant here. Firstly, there may be a genuine physical correlation between the N-H bond lengths and the hydrogen-bond distances, as suggested by spectroscopic and theoretical results (Kollman, 1977; Schuster, 1976). Secondly, the correlation may be an artifact of systematic thermal-motion effects. The observed $r(\text{N-H})$ distances are not corrected for libration, and it is possible that the thermal motion of the H atom – and hence the librational shortening of $r(\text{N-H})$ – becomes progressively smaller as the strength of the hydrogen bond is increased. [Hydrogen bonding is known to produce a blue shift in intramolecular bending modes involving the donor-proton bond (Pimentel & McClellan, 1960).] Librational bond-length corrections are notoriously dependent on the model used for the thermal motion (Koetzle, Lehmann, Verbist & Hamilton, 1972), and

the contribution of libration to the observed $r(\text{N-H})-r(\text{H}\cdots\text{O})$ correlation cannot be assessed with any certainty. However, in a statistical analysis of $\text{O-H}\cdots\text{O}$ hydrogen bonds, it was calculated that approximately half of the observed $r(\text{O-H})-r(\text{H}\cdots\text{O})$ correlation was due to systematic librational effects (Ceccarelli, Jeffrey & Taylor, 1981).

Fig. 7(c) shows the scatterplot of $r(\text{N-H})$ against $r(\text{N}\cdots\text{O})$; the Spearman rank correlation coefficient is -0.474 (significant at the $>99.9\%$ level). It has been suggested that this correlation might be useful for predicting $r(\text{N-H})$ distances in cases where the proton positions have not been determined experimentally (Olovsson & Jönsson, 1976). We therefore calculated the least-squares regression line relating $r(\text{N-H})$ and $r(\text{N}\cdots\text{O})$; this is shown in Fig. 7(c). The regression analysis-of-variance (Table 6) shows that the regression model accounts for a small but statistically significant part of the $r(\text{N-H})$ variance. The deviations of the observed points from the least-squares line show some systematic trends: points at short $r(\text{N}\cdots\text{O})$ distances tend to fall above the line, and *vice versa*. A better fit might be obtained by omitting some of the outlying points on Fig. 7(c), or by using a polynomial expression as suggested by earlier authors (Olovsson & Jönsson, 1976; Hamilton & Ibers, 1968). Additionally, it might be advantageous to modify the regression model so as to take account of the formal charge on the N atom [the mean $r(\text{N-H})$ distance of N^+-H bonds is $1.033(2)$ Å, compared with $1.021(3)$ Å for N-H bonds]. However, all of these regression models would only provide estimates of the *librationally shortened* $r(\text{N-H})$ values.

$r(\text{C=O})$

The $r(\text{C=O})$ and $r(\text{H}\cdots\text{O})$ distances of the bonds involving carboxylate acceptors have a Spearman rank correlation coefficient of -0.021 ; this is not significantly different from zero. However, the corresponding figures for bonds involving other types of acceptors are -0.104 (amides), -0.470 (ketones) and -0.364 (carboxyls). All of these coefficients are sig-

Table 6. Analysis of variance for regression of $r(\text{N-H})$ on $r(\text{N}\cdots\text{O})$

Regression equation: $r(\text{N-H}) = 1.2293 - 0.0700 r(\text{N}\cdots\text{O})$.				
Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Regression	1	0.002424	0.002424	10.9
Residual	81	0.018081	0.000223	
Total	82	0.020505	0.000250	

nificantly different from zero at the $>99.5\%$ level, suggesting that there is a correlation between the $r(\text{C=O})$ distance of the acceptor group and the strength of the hydrogen bond(s) in which the group is involved.

If the observed correlation is physically meaningful (and, again, we cannot exclude the possibility that it is partly due to librational effects), it is reasonable to suppose that $r(\text{C=O})$ will also be dependent on the number of hydrogen bonds accepted by the C=O group (Donohue, 1950). In order to confirm this, we divided the acceptor groups into 'single' (4a), 'double' (4b) and 'triple acceptors' (4c). These were defined as above, the only difference being that the 'multiple acceptors' referred to previously were subdivided into double and triple acceptors. The mean $r(\text{C=O})$ distances of the various types of acceptors are summarized in Table 7(a). As expected, the $r(\text{C=O})$ distance tends to lengthen as the number of hydrogen bonds accepted by the C=O group increases. This trend seems to be less pronounced for carboxylate acceptors than for amides and carboxyls. Mann-Whitney tests showed that the difference between the single- and double-acceptor mean values is statistically ($>97.5\%$ level) significant for carboxylates, amides, ketones and carboxyls. Table 7(b) gives the Spearman rank correlation coefficients of $r(\text{C=O})$ and $r(\text{H}\cdots\text{O})$ for the various types of hydrogen bonds. All of the coefficients are negative, and most are significantly different from zero. Thus, our results confirm that the observed $r(\text{C=O})$ distances are positively correlated with both the number and strength of the hydrogen bonds accepted by the C=O groups.

Table 7(a) also shows that approximately 67% of carboxylate C=O groups accept two or more hydro-

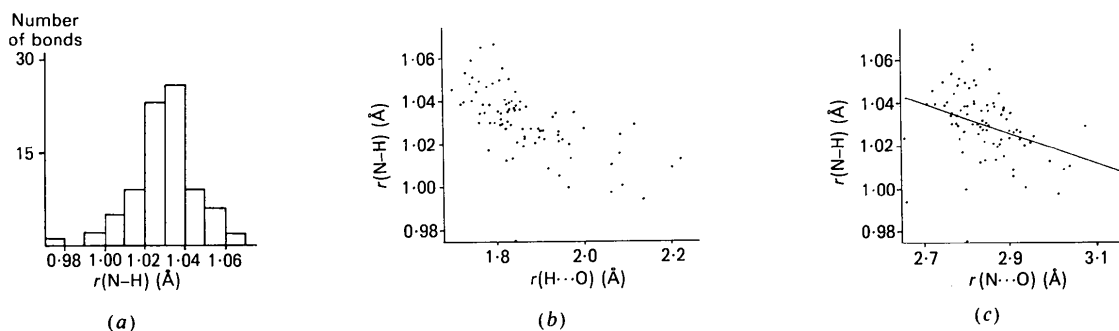


Fig. 7. (a) Distribution of $r(\text{N-H})$ for the 83 bonds determined by neutron diffraction. (b) Scatterplot of $r(\text{N-H})$ against $r(\text{H}\cdots\text{O})$. (c) Scatterplot of $r(\text{N-H})$ against $r(\text{N}\cdots\text{O})$; least-squares regression line is also shown.

Table 7. *Dependence of $r(\text{C}=\text{O})$ on environment of acceptor group*

Acceptor type	'Single' acceptor†		'Double' acceptor†		'Triple' acceptor†	
	N ‡	μ	N	μ	N	μ
Carboxyl	168	1.209 (1)	45	1.221 (3)	1	1.233 (-)
Ketone	54	1.234 (2)	16	1.241 (3)	0	—
Amide	479	1.229 (1)	113	1.238 (1)	4	1.252 (4)
Carboxylate	98	1.242 (1)	176	1.248 (1)	19	1.253 (1)

(b) Spearman rank correlation coefficients of $r(\text{C}=\text{O})$ and $r(\text{H}\cdots\text{O})$

Acceptor type	'Single' acceptor	'Double' acceptor	'Triple' acceptor
Carboxyl	-0.344§	-0.476§	—
Ketone	-0.639§	-0.432§	—
Amide	-0.141§	-0.335§	-0.470
Carboxylate	-0.086	-0.119§	-0.304§

* If a C=O group is involved in two crystallographically independent N-H...O=C hydrogen bonds, both will normally be in the data set on which our survey is based. In constructing Table 7(a), duplicate occurrences of the same C=O group were eliminated, i.e. each crystallographically independent C=O group was counted once, and only once.

† Defined in text.

‡ N = number of observations in sample, μ = sample mean.

§ Significantly different from zero at or above 95% confidence level (one-tailed test).

gen bonds. Corresponding figures for amides, ketones and carboxyls are 20, 23 and 21%, respectively. This is consistent with our conclusion that carboxylates form stronger hydrogen bonds than amides, ketones and carboxyls.

Olga Kennard is a member of the external staff of the Medical Research Council.

References

ALLEN, F. H., BELLARD, S., BRICE, M. D., CARTWRIGHT, B. A., DOUBLEDAY, A., HIGGS, H., HUMMELINK, T., HUMMELINK-PETERS, B. G., KENNARD, O., MOTHERWELL, W. D. S., RODGERS, J. R. & WATSON, D. G. (1979). *Acta Cryst.* **B35**, 2331–2339.

BERKOVITCH-YELLIN, Z., ARIEL, S. & LEISEROWITZ, L. (1983). *J. Am. Chem. Soc.* **105**, 765–767.

CECCARELLI, C., JEFFREY, G. A. & TAYLOR, R. (1981). *J. Mol. Struct.* **70**, 255–271.

DONOHUE, J. (1950). *J. Am. Chem. Soc.* **72**, 949–953.

HAMILTON, W. C. (1964). *Statistics in Physical Science*, pp. 173–174. New York: Ronald Press.

HAMILTON, W. C. & IBERS, J. A. (1968). *Hydrogen Bonding in Solids*, pp. 52–53. New York: Benjamin.

JEFFREY, G. A. & LEWIS, L. (1978). *Carbohydr. Res.* **60**, 179–182.

KOETZLE, T. F. & LEHMANN, M. S. (1976). *The Hydrogen Bond*, edited by P. SCHUSTER, G. ZUNDEL & C. SANDORFY, Vol. II, ch. 9. Amsterdam: North-Holland.

KOETZLE, T. F., LEHMANN, M. S., VERBIST, J. J. & HAMILTON, W. C. (1972). *Acta Cryst.* **B28**, 3207–3214.

KOLLMAN, P. A. (1977). *Modern Theoretical Chemistry*, edited by H. F. SCHAEFER, Vol. 4, ch. 3. New York: Plenum.

KROON, J., KANTERS, J. A., VAN DUJNEVELDT-VAN DE RIJDT, J. G. C. M., VAN DUJNEVELDT, F. B. & VLIAGENTHART, J. A. (1975). *J. Mol. Struct.* **24**, 109–129.

OLOVSSON, I. & JÖNSSON, P.-G. (1976). *The Hydrogen Bond*, edited by P. SCHUSTER, G. ZUNDEL & C. SANDORFY, Vol. II, ch. 8. Amsterdam: North-Holland.

PIMENTEL, G. C. & MCCLELLAN, A. L. (1960). *The Hydrogen Bond*, p. 70. San Francisco: Freeman.

SCHUSTER, P. (1976). *The Hydrogen Bond*, edited by P. SCHUSTER, G. ZUNDEL & C. SANDORFY, Vol. I, ch. 2. Amsterdam: North-Holland.

SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. International Student Edition. Tokyo: McGraw-Hill Kogakusha.

SNEDECOR, G. W. & COCHRAN, W. G. (1980). *Statistical Methods*, 7th ed. Ames: Iowa State Univ. Press.

TAYLOR, R. & KENNARD, O. (1983a). *Acta Cryst.* **B39**, 133–138.

TAYLOR, R. & KENNARD, O. (1983b). *Acta Cryst.* **B39**, 517–525.

TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1983). *J. Am. Chem. Soc.* **105**, 5761–5766.

TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1984). *J. Am. Chem. Soc.* **106**, 244–248.

UMEYAMA, H. & MOROKUMA, K. (1977). *J. Am. Chem. Soc.* **99**, 1316–1332.

WETHERILL, G. B. (1981). *Intermediate Statistical Methods*, pp. 325–329. London: Chapman & Hall.

Acta Cryst. (1984). **B40**, 288–294

Studies of Bonding at S in the Solid State. The Structure of 1-(Methylsulfonyl)-2-(phenylthio)ethane at 86 K

BY F. MO, B. C. HAUBACK AND S. WINTHER

Institutt for Røntgenteknikk, Universitetet i Trondheim–NTH, N-7034 Trondheim, Norway

(Received 15 September 1983; accepted 15 December 1983)

Abstract

Structure data for a variety of crystalline molecules containing S in different oxidation states show that sulfides represent a structural discontinuity relative to the corresponding sulfoxides and sulfones. The

apparent anomalous behavior of sulfides seems to be general, but is different in aromatic and aliphatic compounds. Simple models for bonding cannot predict the observed anomalies. The trend in C–S–C angles, with a minimum for the sulfoxides, was observed previously in gaseous molecules. The observa-