

## 9.5. Typical interatomic distances: organic compounds

BY F. H. ALLEN, D. G. WATSON, L. BRAMMER, A. G. ORPEN AND R. TAYLOR

### 9.5.1. Introduction

The determination of molecular geometry is of vital importance to our understanding of chemical structure and bonding. The majority of experimental data have come from X-ray and neutron diffraction, microwave spectroscopy, and electron diffraction. Over the years, compilations of results from these techniques have appeared sporadically. The first major compilation was Chemical Society Special Publication No. 11: *Tables of Interatomic Distances and Configuration in Molecules and Ions* (Sutton, 1958). This volume summarized results obtained by diffraction and spectroscopic methods prior to 1956; a supplementary volume (Sutton, 1965) extended this coverage to 1959. Summary tables of bond lengths between carbon and other elements were also published in Volume III of *International Tables for X-ray Crystallography* (Kennard, 1962). Some years later, the Cambridge Crystallographic Data Centre (Allen, Bellard, Brice, Cartwright, Doubleday, Higgs, Hummelink, Hummelink-Peters, Kennard, Motherwell, Rodgers & Watson, 1979) produced an atlas-style compendium of all organic, organometallic and metal-complex crystal structures published in the period 1960–1965 (Kennard, Watson, Allen, Isaacs, Motherwell, Pettersen & Town, 1972). More recently, a survey of geometries determined by spectroscopic methods (Harmony, Laurie, Kuczkowski, Schwendemann, Ramsay, Lovas, Lafferty & Maki, 1979) has extended coverage in this area to mid-1977.

The production of further comprehensive compendia of X-ray and neutron diffraction results has been precluded by the steep rise in the number of published crystal structures, as illustrated by Fig. 9.5.1.1. Printed compilations have been effectively superseded by computerized databases. In particular, the Cambridge Structural Database now (1991) contains bibliographic, chemical, and numerical results for some 86 000 organo-carbon crystal structures. This machine-readable file fulfils the function of a comprehensive structure-by-structure compendium of molecular geometries. However, the amount of data now held in the CSD is so large that there is also a need for concise, printed tabulations of average molecular dimensions.

The only tables of average geometry in general use are those contained in Sutton (1958, 1965) which list mean bond lengths for a variety of atom pairs and functional groups. Since these

early tables were based on data obtained before 1960, we have used the CSD to prepare a new table of average bond lengths in organic compounds. Table 9.5.1.1 given here specifically lists average lengths for bonds involving the elements H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, and I. Mean values are presented for 682 different bond types involving these elements.

### 9.5.2. Methodology

#### 9.5.2.1. Selection of crystallographic data

All results given in Table 9.5.1.1 are based on X-ray and neutron diffraction results retrieved from the September 1985 version of the CSD. Neutron diffraction data only were used to derive mean bond lengths involving H atoms. This version of the CSD contained results for 49 854 single-crystal diffraction studies of organo-carbon compounds; 10 324 of these satisfied the acceptance criteria listed below and were used in the averaging procedures:

(i) Structure is 'organic', *i.e.* belongs to the CSD chemical classes 1–65 or 70 (*Cambridge Crystallographic Data Centre User Manual*, 1978).

(ii) Atomic coordinates for the structure have been published and are available in the CSD.

(iii) Structure was determined from diffractometer data.

(iv) Structure does not contain unresolved numeric data errors from the original publication (such errors are usually typographical and are normally resolved by consultation with the authors).

(v) Structure was not reported to be disordered.

(vi) Only structures of higher precision were included on the basis of *either* (a) the crystallographic *R* factor was  $\leq 0.07$  and the reported mean estimated standard deviation (e.s.d.) of the C—C bond lengths was  $\leq 0.010 \text{ \AA}$  (corresponds to AS flag = 1 or 2 in the CSD), *or* (b) the crystallographic *R* factor  $\leq 0.05$  and the mean e.s.d. for C—C bonds was not available in the database (AS = 0 in the CSD).

(vii) Where the structure of a given compound had been determined more than once within the limits of (i)–(vi), then only the most precise determination was used.

#### 9.5.2.2. Program system

All calculations were performed on the University of Cambridge IBM 3081 D using the programs *BIBSER*, *CONNSE*, *RETRIEVE*, *GEOM78*, and *PLUTO78* (Allen *et al.*, 1979). A stand-alone program was written to implement the selection criteria, whilst a new program (*STATS*) was written to perform the statistical calculations described below. It was also necessary to modify *CONNSE* to improve the precision with which it locates chemical substructures. In particular, the program was altered to permit the location of atoms with specified coordination numbers. This was essential in the case of carbon so that atoms with coordination numbers 2, 3 and 4 (equivalent to formal hybridization states  $sp^1$ ,  $sp^2$ ,  $sp^3$ ) could be distinguished easily and reliably. Considerable care was taken to ensure that the correct molecular fragment was located by *GEOM78* in the generation of geometrical tabulations. This often involved the explicit specification of H atoms in fragments, and the extensive use of geometrical tests on valence and torsion angles. Considerable use was also made of chemical structural diagrams, which are available in the Cambridge in-house version

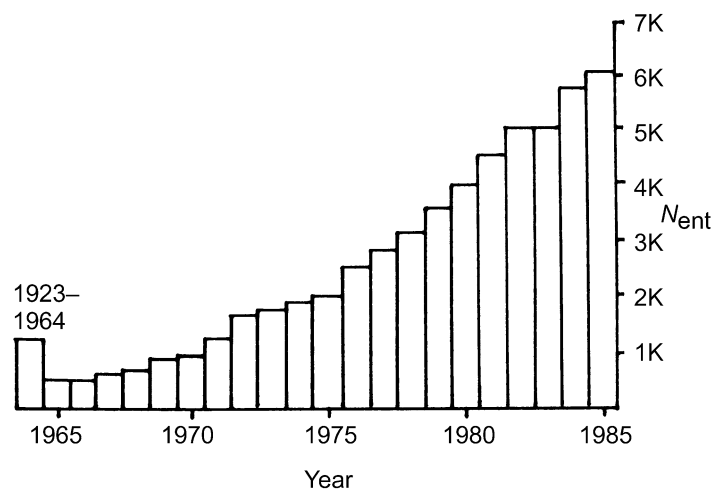


Fig. 9.5.1.1. Growth of the Cambridge Structural Database 1965–1985 as number of entries ( $N_{ent}$ ) published in a given year.

## 9.5. TYPICAL INTERATOMIC DISTANCES: ORGANIC COMPOUNDS

of the CSD for some 81% of all entries. Chemical diagrams proved useful, for example, in identifying the various coordination environments commonly adopted by atoms such as As, B, P, *etc.*

### 9.5.2.3. Classification of bonds

The classification of bonds used in Table 9.5.1.1 is based on common functional groups, rings and ring systems, coordination spheres, *etc.* It is designed: (i) to appear logical, useful and reasonably self-explanatory to chemists, crystallographers, and others who may use the table; (ii) to permit a meaningful average value to be cited for each bond length. With reference to (ii), it was considered that a sample of bond lengths could be averaged meaningfully if: (a) the sample was unimodally distributed; (b) the sample standard deviation ( $\sigma$ ) was reasonably small, ideally less than *ca* 0.02 Å; (c) there were no conspicuous outlying observations – those that occurred at  $> 4\sigma$  from the mean were automatically eliminated from the sample by *STATS*, other outliers were inspected carefully; (d) there were no compelling chemical reasons for further subdivision of the sample.

### 9.5.2.4. Statistics

Where there are less than four independent observations of a given bond length, then each individual observation is given explicitly in the table. In all other cases, the following statistics were generated by the program *STATS*.

(i) The unweighted sample mean,  $d$ , where

$$d = \sum_{i=1}^n d_i/n$$

and  $d_i$  is the  $i$ th observation of the bond length in a total sample of  $n$  observations. Recent work (Taylor & Kennard, 1983, 1985, 1986) has shown that the unweighted mean is an acceptable (even preferable) alternative to the weighted mean, where the  $i$ th observation is assigned a weight equal to  $1/\sigma^2(d_i)$ . This is especially true (Taylor & Kennard, 1985) where structures have been pre-screened on the basis of precision.

(ii) The sample median,  $m$ . This has the property that half of the observations in the sample exceed  $m$ , and half fall short of it.

(iii) The sample standard deviation, denoted here as  $\sigma$ , where:

$$\sigma = \sum_{i=1}^n [(d_i - d)^2/(n - 1)]^{1/2}$$

(iv) The lower quartile for the sample,  $q_l$ . This has the property that 25% of the observations are less than  $q_l$  and 75% exceed it.

(v) The upper quartile for the sample,  $q_u$ . This has the property that 25% of the observations exceed  $q_u$ , and 75% fall short of it.

(vi) The number ( $n$ ) of observations in the sample.

The statistics given in the final table correspond to distributions for which the automatic  $4\sigma$  cut-off (see above) had been applied, and any manual removal of additional outliers (an infrequent operation) has been performed. In practice, a very small percentage of observations was excluded by these methods. The major effect of removing outliers is to improve the sample standard deviation, as shown in Fig. 9.5.2.1 in which a single observation is deleted.

The statistics chosen for tabulation effectively describe the distribution of bond lengths in each case. For a symmetrical, normal distribution: the mean ( $d$ ) will be approximately equal to the median ( $m$ ); the lower and upper quartiles ( $q_l, q_u$ ) will be approximately symmetric about the median:  $m - q_l \simeq q_u - m$ ,

and 95% of the observations may be expected to lie within  $\pm 2\sigma$  of the mean value. For a skewed distribution,  $d$  and  $m$  may differ appreciably and  $q_l$  and  $q_u$  will be asymmetric with respect to  $m$ . When a bond-length distribution is negatively skewed as in Fig. 9.5.2.2, *i.e.* very short values are more common than very long values, then it may be due to thermal-motion effects; the distances used to prepare the table were not corrected for thermal libration.

In a number of cases, the initial bond-length distribution was clearly bimodal, as in Fig. 9.5.2.3(a). All cases of bimodality were resolved on chemical grounds before inclusion in the table, on the basis of hybridization, conformation-dependent conjugation interactions, *etc.* For example, the histogram of Fig. 9.5.2.3(a) was resolved into the two discrete unimodal distributions of Figs. 9.5.2.3(b), (c), which correspond to planar N( $sp^2$ ), pyramidal N( $sp^3$ ), respectively. The mean valence angle at N was used as the discriminator, with a range of 108–114° for N( $sp^3$ ) and  $\geq 117.5^\circ$  for N( $sp^2$ ).

### 9.5.3. Content and arrangement of the table

The upper triangular matrix of Fig. 9.5.3.1(a) shows the 120 possible element-pair combinations that can be formed from the 15 elements As, B, Br, C, Cl, F, H, I, N, O, P, S, Se, Si, Te.

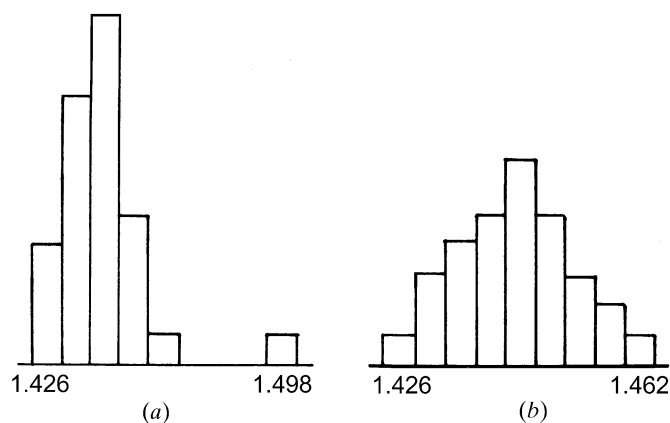


Fig. 9.5.2.1. Effect of the removal of outliers (contributors that are  $> 4\sigma$  from the mean) for the C—C bond in  $C_{ar}-C\equiv N$  fragments. Relevant statistics (see text) are:

	$d$	$m$	$\sigma$	$q_l$	$q_u$	$n$
(a) before	1.445	1.444	0.012	1.436	1.448	32
(b) after	1.455	1.444	0.008	1.436	1.448	31.

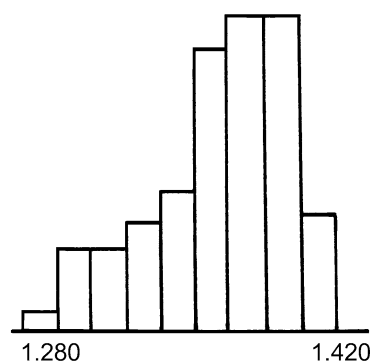


Fig. 9.5.2.2. Skewed distribution of B—F bond lengths in  $BF_4^-$  ions:  $d = 1.365$ ,  $m = 1.372$ ,  $\sigma = 0.029$ ,  $q_l = 1.352$ ,  $q_u = 1.390$  for 84 observations. Note that  $d \neq m$  and that  $q_l, q_u$  are asymmetrically disposed about the mean  $d$ .