

9. BASIC STRUCTURAL FEATURES

Fig. 9.5.3.1(a) contains the number of discrete average bond lengths given in the table for each element pair. 682 average values are cited for 65 element pairs, of which 511 (75%) involve carbon. Bond-length values from individual structures are given for a further 30 element pairs indicated by an asterisk in Fig. 9.5.3.1(a). Individual structures are identified by their CSD reference code (*e.g.* BOGSUL), and short-form literature references, ordered alphabetically by reference code, are given in Appendix 2. For eight element pairs, the acceptance criterion (vi) was relaxed to include all available structures, irrespective of precision. These entries are denoted by a dagger in the table. No bonds were found for 25 element pairs within the subset of CSD used in this study.

Each entry in Table 9.5.1.1 contains nine columns, of which six record the statistics of the bond-length distribution described above. The content of the remaining three columns: 'Bond', 'Structure', 'Note', are now described.

9.5.3.1. Ordering of entries: the 'Bond' column

For an element pair $X-Y$, the primary ordering is alphabetic by element symbols according to the rows of Fig. 9.5.3.1(a); *i.e.* X changes slowest, Y fastest. The complete sequence runs from As—As to Te—Te with bonds involving carbon in the natural position: As—C...C—C...C—Te. Within a given $X-Y$ pair, a secondary ordering is based on the coordination numbers (j) of X and Y , and on the nature of the bond between them. The bond definition is of the form $X(j)-Y(j)$, with j decreasing fastest for Y , slowest for X , and with all single bonds preceding any multiple bonds. For carbon, the formal hybridization state replaces (but is equivalent to) the coordination number and it is for this element that the ordering rules are most clearly required. The ordering of the most populous C—C, C—N, C—O sections is illustrated in Fig. 9.5.3.1(b). The 13 possible C—C combinations follow the sequence Csp^3-Csp^3 , Csp^3-Csp^2 , Csp^3-C_{ar} , Csp^3-Csp^1 , Csp^2-Csp^2 , Csp^2-C_{ar} , Csp^2-Csp^1 , $C_{ar}-C_{ar}$, $C_{ar}-Csp^1$, Csp^1-Csp^1 , $Csp^2=Csp^2$, $C_{ar}\equiv C_{ar}$, $Csp^1\equiv Csp^1$. The symbol C_{ar} represents

aryl carbon in six-membered rings, which is treated separately from Csp^2 throughout the table. The symbol \equiv is used to indicate a delocalized double or aromatic bond according to context.

9.5.3.2. Definition of 'Substructure'

The chemical environment of each bond is normally defined by a linear formulation of the substructure. The target bond is set in bold type, *e.g.* $C_{ar}-C\equiv N$ (aryl cyanides); $C-CH_2-O-C_{ar}$ (primary alkyl aryl ethers); $(C-O)_2-P(\equiv O)_2$ (phosphate diesters). Occasionally, the chemical name of a functional group or ring system is used to define bond environment, *e.g.* in naphthalene, C2—C3; in imidazole, N1—C2. To avoid any possible ambiguity in these cases, we include numbered chemical diagrams in Fig. 9.5.3.2.

	As	B	Br	C	Cl	F	H	I	N	O	P	S	Se	Si	Te
As	1	*	*	6	1	1	0	*	2	2	+	2	+	+	+
B	2	2	5	2	2	0	*	4	2	1	4	0	0	0	0
Br	1	5	+	0	0	*	*	1	*	+	*	*	*	*	*
C	175	16	13	10	2	97	119	12	40	3	6	2			
Cl	*	0	0	*	*	1	2	1	*	2	1				
F	0	0	0	1	0	2	2	0	3	*					
H	0	0	2	3	0	0	0	0	0	0					
I	1	*	1	+	*	0	0	1							
N	15	18	15	8	*	6	*								
O	4	20	9	*	5	2									
P	2	3	1	1	*										
S	5	1	1	2											
Se	1	0	+												
Si	1	0													
Te	*														

(a)

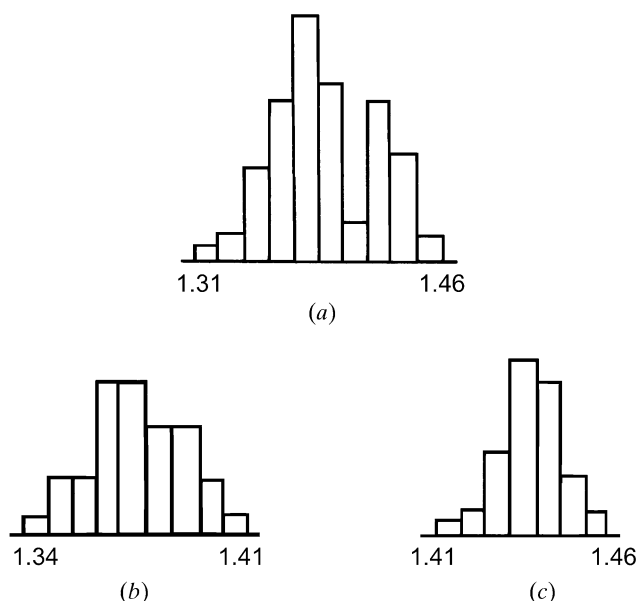


Fig. 9.5.2.3. Resolution of the bimodal distribution of C—N bond lengths in $C_{ar}-N(Csp^3)_2$ fragments: (a) complete distribution; (b) distribution for planar N, mean valence angle at N > 117.6°; (c) distribution for pyramidal N, mean valence angle at N in the range 108–114°.

	C—C						C—N				C—O		
	Csp^3	Csp^2	C_{ar}	Csp^1	Csp^2	C_{ar}	Csp^1	N(4)	N(3)	N(2)	N(1)	O(2)	O(1)
Csp^3	27	30	6	4	-	-	-	5	22	2	-	60	-
Csp^2	24	14	2	-	-	-	-	28	1	-	17	-	
C_{ar}	2	2	-	-	-	-	1	9	2	-	4	-	
Csp^1	1	-	-	-	-	-	-	-	-	-	-	-	
$Csp^2=$	34	-	-	-	1	8	-	-	-	-	38	-	
$C_{ar}\equiv$	24	-	-	3	9	-	-	-	-	-	-	-	
$Csp^1\equiv$	5	-	-	1	5	-	-	-	-	-	-	-	

(b)

Fig. 9.5.3.1. (a) Distribution of mean bond-length values reported in the table by element pair. An asterisk indicates a bonded pair represented by less than four contributors in the original data set. A '+' indicates bonded pairs located when restrictions on R factor and reported e.s.d. limits were lifted (see text). (b) Distribution of mean bond-length values reported in the table for C—C, C—O, C—N.

9.5. TYPICAL INTERATOMIC DISTANCES: ORGANIC COMPOUNDS

A combination of chemical name and linear formulation is often employed to increase the precision of the definition, *e.g.* $\text{NH}_2-\text{C}=\text{O}$ in acyclic amides; $\text{C}=\text{C}-\text{C}(=\text{O})-\text{C}=\text{O}$ in benzoquinone. Finally, for very simple ions, the accepted conventional representation is deemed to be sufficient, *e.g.* in NO_3^- , SO_4^{2-} , *etc.*

The chemical definition of substructure may be followed by brief qualifying information, concerning substitution, conformational restrictions, *etc.* For example: $\text{Csp}^3-\text{Csp}^3$: in cyclobutane (any substituent); $\text{X}-\text{C}-\text{F}_3$ ($\text{X} = \text{C}, \text{H}, \text{N}, \text{O}$); $\text{C}_{\text{ar}}-\text{NH}-\text{Csp}^3$ (Nsp^3 : pyramidal). Where the generic symbol X is unqualified, it denotes any element type, including hydrogen. If the qualifying information is too extensive, then it will be given as a table footnote (see below).

The 'Substructure' column is designed to convey as much unambiguous information as possible within a small space. For Csp^3 , we have employed the short forms C^* and $\text{C}^\#$. C^* indicates

Csp^3 whose bonds, additional to those specified in the linear formulation, are to C or H atoms only. C^*-OH would then represent the group of alcohols CH_3-OH , $-\text{C}-\text{CH}_2-\text{OH}$, $-\text{C}_2-\text{CH}-\text{OH}$ and $-\text{C}_3-\text{C}-\text{OH}$. C^* is frequently used to restrict the secondary environment of a given bond to avoid the perturbing influence of, *e.g.*, electronegative substituents. The symbol $\text{C}^\#$ is merely a space-saving device to indicate any Csp^3 atom and includes C^* as a subset.

9.5.3.3. Use of the 'Note' column

The 'Note' column refers to the footnotes collected in Appendix 1. These record additional information as follows: (a) additional details concerning the chemical definition of substructures, *e.g.* the omission of three- and four-membered rings; (b) statements of geometrical constraints used in obtaining the cited average, *e.g.* definition of planarity or pyramidality at

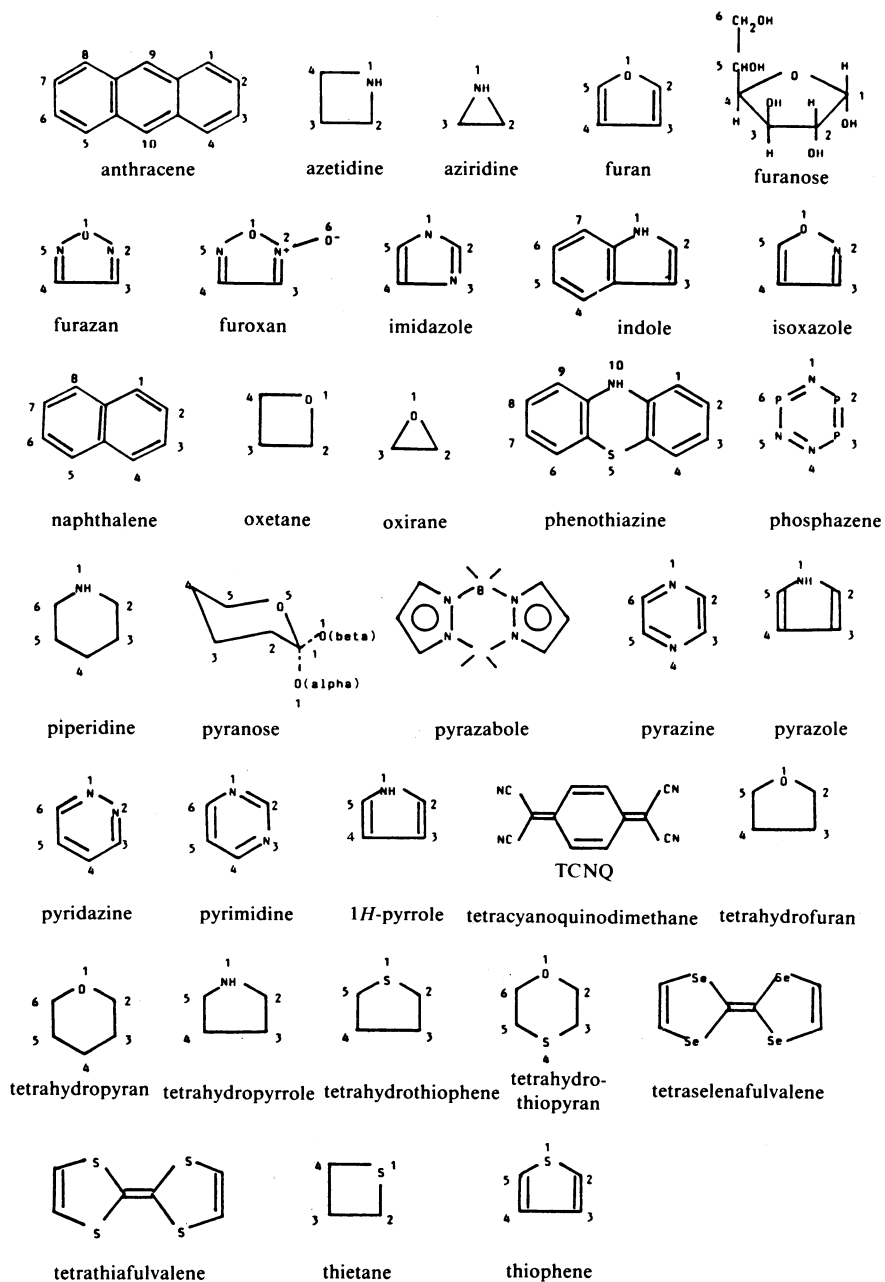


Fig. 9.5.3.2. Alphabetized index of ring systems referred to in the table; the numbering scheme used in assembling the bond-length data is given where necessary.