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 (σ) 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 σ, 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 4σ 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 $\geq 117.5°$ 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.
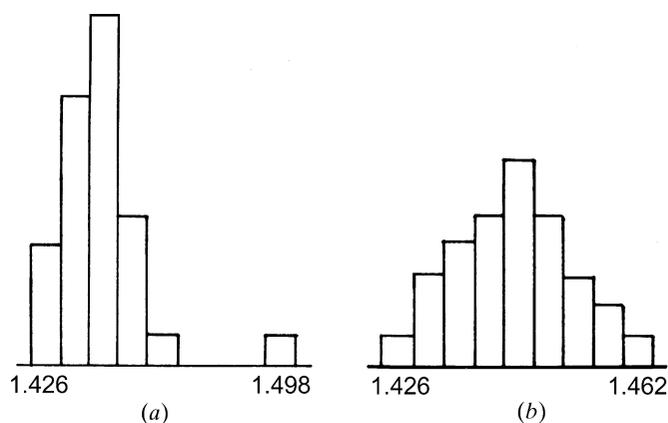

Fig. 9.5.2.1. Effect of the removal of outliers (contributors that are $> 4\sigma$ from the mean) for the C—C bond in C_{ar}—C≡N fragments. Relevant statistics (see text) are:

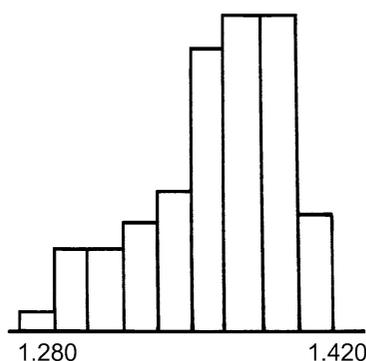|	d	m	σ	q_l	q_u	n
(*a*) before	1.445	1.444	0.012	1.436	1.448	32
(*b*) after	1.455	1.444	0.008	1.436	1.448	31.

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.

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 $Csp^3—Csp^3$, $Csp^3—Csp^2$, $Csp^3—C_{ar}$, $Csp^3—Csp^1$, $Csp^2—Csp^2$, $Csp^2—C_{ar}$, $Csp^2—Csp^1$, $C_{ar}—C_{ar}$, $C_{ar}—Csp^1$, $Csp^1—Csp^1$, $Csp^2{=}Csp^2$, $C_{ar}{\cdots}C_{ar}$, $Csp^1{\equiv}Csp^1$. The symbol C_{ar} represents

aryl carbon in six-membered rings, which is treated separately from Csp^2 throughout the table. The symbol $\cdots$ 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.* $C_{ar}—\mathbf{C{\equiv}N}$ (aryl cyanides); $C—\mathbf{CH_2}—\mathbf{O}—C_{ar}$ (primary alkyl aryl ethers); $(C—O)_2—\mathbf{P}({\cdots}\mathbf{O})_2$ (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.

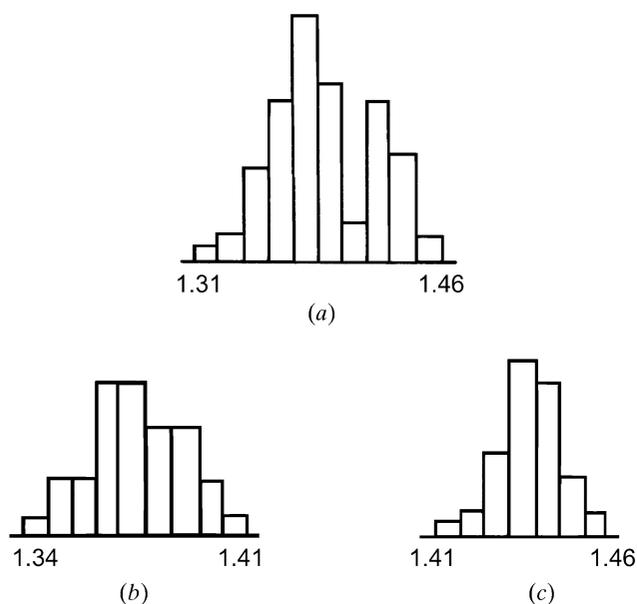|	As	B	Br	C	Cl	F	H	I	N	O	P	S	Se	Si	Te
As	1	*	*	6	1	1	0	*	2	2	+	2	+	+	+
B		2	2	5	2	2	0	*	4	2	1	4	0	0	0
Br			1	5	+	0	0	*	*	1	*	+	*	*	*
C				175	16	13	10	2	97	119	12	40	3	6	2
Cl					*	0	0	*	*	1	2	1	*	2	1
F						0	0	0	1	0	2	2	0	3	*
H							0	0	2	3	0	0	0	0	0
I								1	*	1	+	*	0	0	1
N									15	18	15	8	*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

(*a*)

(*b*) (*c*)

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
$C_{ar}{\cdots}$						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.

A combination of chemical name and linear formulation is often employed to increase the precision of the definition, *e.g.* NH_2—C=O in acyclic amides; C=C—$C(=O)$—C=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: Csp^3—Csp^3: in cyclobutane (any substituent); X—C—F_3 (X = C, H, N, O); C_{ar}—NH—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 $C^\#$. 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, —C—CH_2—OH, —C_2—CH—OH and —C_3—C—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 $C^\#$ 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
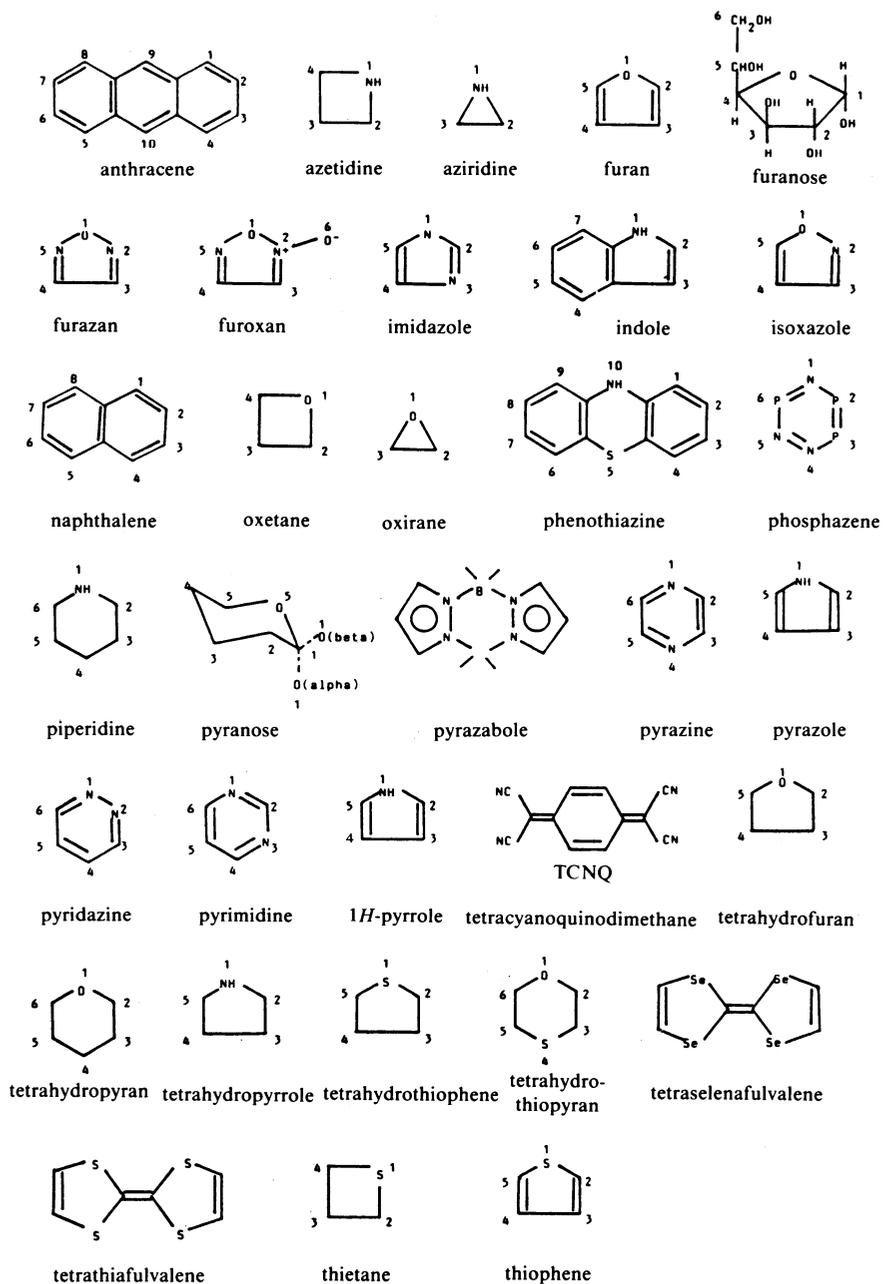

Fig. 9.5.3.2. Alphabetized index of ring systems referred to in the table; the numbering scheme used in assembling the bond-length data is given where necessary.

N, torsional constraints in conjugated systems; (c) any peculiarities of a particular bond-length distribution, e.g. sample dominated by C* = methyl; (d) references to previously published surveys of crystallographic results relevant to the substructure in question. We do not claim that these references are in any way comprehensive and we would be grateful to authors for notification (to FHA) of any omissions. This will serve to improve the content of any future version of the table.

9.5.4. Discussion

It should be remembered that this table has been derived from the organic section of CSD. We are aware that a number of organic bond types which occur very frequently in organometallics and metal complexes (e.g. $C \cdots C$ in cyclopentadienyl, $C—P$ in triphenylphosphine, etc.) are either absent or poorly represented in this work. These omissions are rectified in Chapter 9.6. We also note that certain bond types listed here (e.g. As—O, Si—O, Si—N, etc.) will occur with greater frequency in inorganic compounds. The interested reader is referred to the Inorganic Crystal Structure Database (Bergerhof, Hundt, Sievers & Brown, 1983) for a machine-readable compendium of more relevant structural data.

The tabulation given here represents the first stage in a major project designed to obtain the average geometries of function groups, rigid rings, and the low-energy conformations of flexible rings. Details of mean bond lengths, valence angles, and conformational preferences in a wide range of substructures will form the basis of a machine-readable 'fragment library' for use in molecular modelling and other areas of research. The systematic survey will be extended to derive information about distances, angles, directionality, and environmental dependence of hydrogen bonds and non-bonded interactions.

APPENDIX 1
Notes to Table 9.5.1.1

(1) Sample dominated by B—CH$_3$. For longer bonds in B$^-$—CH$_3$, see LITMEB10 [B(4)—CH$_3$ = 1.621–1.644 Å].

(2) $p(\pi)—p(\pi)$ bonding with Bsp^2 and Nsp^2 coplanar (τBN = $0 \pm 15°$) predominates. See G. Schmidt, R. Boese & D. Bläser [Z. Naturforsch. Teil B (1982), **37**, 1230–1233].

(3) 84 observations range from 1.38 to 1.62 Å and individual values depend on substituents on B and O. For a discussion of borinic acid adducts, see S. J. Rettig & J. Trotter [Can. J. Chem. (1982), **60**, 2957–2964].

(4) See M. Kaftory (1983). [In The chemistry of functional groups. Supplement D: The chemistry of halides, pseudo-halides and azides, Part 2, Chap. 24, edited by S. Patai & Z. Rappoport. New York: John Wiley.]

(5) Bonds that are endocyclic or exocyclic to any three- or four-membered rings have been omitted from all averages in this section.

(6) The overall average given here is for $Csp^3—Csp^3$ bonds which carry only C or H substituents. The value cited reflects the relative abundance of each 'substitution' group. The 'mean of means' for the nine subgroups is 1.538 ($\sigma = 0.022$) Å.

(7) See (a) F. H. Allen [Acta Cryst. (1980), **B36**, 81—96] and (b) F. H. Allen [Acta Cryst. (1981), **B37**, 890—900].

(8) See F. H. Allen [Acta Cryst. (1984), **B40**, 64—72].

(9) See F. H. Allen [Tetrahedron (1982), **38**, 2843–2853].

(10) See F. H. Allen [Tetrahedron (1982), **38**, 645–655].

(11) Cyclopropanones and cyclobutanones excluded.

(12) See W. B. Schweizer & J. D. Dunitz [Helv. Chim. Acta (1982), **65**, 1547–1554].

(13) See L. Norskov-Lauritsen, H.-B. Bürgi, P. Hoffmann & H. R. Schmidt [Helv. Chim. Acta (1985), **68**, 76–82].

(14) See P. Chakrabarti & J. D. Dunitz [Helv. Chim. Acta (1982), **65**, 1555–1562].

(15) See J. L. Hencher (1978). [In The chemistry of the $C \equiv C$ triple bond, Chap. 2, edited by S. Patai. New York: John Wiley.]

(16) Conjugated: torsion angles about central C—C single bond is $0 \pm 20°$ (cis) or $180 \pm 20°$ (trans).

(17) Unconjugated: torsion angle about central C—C single bond is 20–160°.

(18) Other conjugative substituents excluded.

(19) TCNQ is tetracyanoquinodimethane (see diagrams).

(20) No difference detected between $C2 \cdots C3$ and $C3 \cdots C4$ bonds.

(21) Derived from neutron diffraction results only.

(22) Nsp^3: pyramidal; mean valence angle at N is in the range 108–114°.

(23) Nsp^2: planar; mean valence angle at N is $\geq 117.5°$.

(24) Cyclic and acyclic peptides.

(25) See R. H. Blessing [J. Am. Chem. Soc. (1983), **105**, 2776–2783].

(26) See L. Lebioda [Acta Cryst. (1980), **B36**, 271–275].

(27) $n = 3$ or 4; i.e. tri- or tetrasubstituted ureas.

(28) Overall value also includes structures with mean valence angle at N in the range 115–118°.

(29) See F. H. Allen & A. J. Kirby [J. Am. Chem. Soc. (1984), **106**, 6197–6200].

(30) See A. J. Kirby (1983). [The anomeric effect and related stereoelectronic effects at oxygen. Berlin: Springer.]

(31) See B. Fuchs, L. Schleifer & E. Tartakovsky [Nouv. J. Chim. (1984), **8**, 275–278].

(32) See S. C. Nyburg & C. H. Faerman [J. Mol. Struct. (1986), **140**, 347–349].

(33) Sample dominated by P—CH$_3$ and P—CH$_2$—C.

(34) Sample dominated by C* = methyl.

(35) See A. Kálmán, M. Czugler & G. Argay [Acta Cryst. (1981), **B37**, 868–877].

(36) Bimodal distribution resolved into 22 'short' bonds and 5 longer outliers.

(37) All 24 observations come from BUDTEZ.

(38) 'Long' O—H bonds in centrosymmetric $O \cdots H \cdots O$ H-bonded dimers are excluded.

(39) N—N bond length also dependent on torsion angle about N—N bond and on nature of substituent C atoms – these effects are ignored here.

(40) N pyramidal has average angle at N in the range 100–113.5°; N planar has average angle $\geq 117.5°$.

(41) See R. R. Holmes & J. A. Deiters [J. Am. Chem. Soc. (1977), **99**, 3318–3326].

(42) No detectable variation in S=O bond length with type of C substituent.

APPENDIX 2
Short-form references to individual CSD entries cited by reference code in Table 9.5.1.1

REFCODE	Journal	Vol.	Page	Year
ACBZPO01	J. Am. Chem. Soc.	97	6729	1975
ACLTEP	J. Organomet. Chem.	184	417	1980
ASAZOC	Dokl. Akad. Nauk SSSR	249	120	1979
BALXOB	J. Am. Chem. Soc.	103	4587	1981